Measuring Economic Integration

Volume 59 Issue 2 June 2013
For checked items

view abstracts
download to citation manager

Articles

CESifo Economic Studies, Vol. 59, 2/2013, 195–198 doi:10.1093/cesifo/ift008

Introduction to Issue on Measuring Economic Integration

Volker Nitsch* and Nikolaus Wolf†

*Technische Universität Darmstadt, Fachbereich Rechts- und Wirtschaftswissenschaften, Residenzschloss, Marktplatz 15, 64283 Darmstadt, Germany. e-mail: nitsch@vwl.tu-darmstadt.de

†Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, Spandauer Straße 1, 10178 Berlin, Germany. e-mail: nikolaus.wolf@wiwi.hu-berlin.de

Among a wider public, global market integration is often taken for granted. Products offered in local stores typically originate from all over the world. Goods may also have crossed borders numerous times during the production process. Online marketplaces, such as eBay, facilitate effortless business with (basically anonymous) partners in remote places. After all, goods and services are traded over thousands of kilometers every day.

Still, integration is not an irreversible process. Institutional, financial and political factors, among others, provide fault lines of fragmentation that may limit further integration or even promote disintegration. As the collapse of world trade in the wake of the financial crisis in 2008/2009 has shown, for instance, trade may dry up quickly. Also, institutional processes (such as WTO negotiations) could become stagnant, and even countries can be dissolved. In view of these developments, the mechanisms of integration have again taken center stage in scholarly discussions of international economics.

A key challenge in this debate is the empirical measurement of economic integration and its underlying determinants. How can we compare levels of integration over time and across regions? What assumptions are necessary to identify barriers to the integration of goods and factor markets? How can we estimate the effect of policy changes on integration if the policies themselves are endogenous? The articles collected in this issue, presented and discussed at a conference at CESifo Munich in February 2011, deal with these and related questions.

Jeffrey Bergstrand, in his keynote address to the conference, takes the title of the meeting, 'Measuring Economic Integration', seriously. Starting from the observation that there is little precise and consistent evidence on the quantitative effect of a change in trade costs on trade, Bergstrand reviews two methodological issues in the measurement of trade costs. Specifically, he critically reviews methods for estimating the elasticity of trade with respect to trade costs when the true values of bilateral trade costs are unobservable (as they typically are in reality) and the (imperfect) estimates of trade costs suffer from endogeneity.

© The Author 2013. Published by Oxford University Press on behalf of Ifo Institute, Munich. All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

The article by Costas Arkolakis and Marc-Andreas Muendler builds on the rapidly growing literature on the empirics of firm-level trade. A key difficulty of work in this area is the generally limited availability of relevant data sets. To the extent that data are available at all, the data are often only accessible to national researchers under strict rules. Also, firm-level data sets frequently differ across countries in the list and definition of variables, the level of disaggregation, the period that is covered and other features of design. As a result, many empirical findings are derived from national data only, which are then sought to be replicated, under different settings, for other countries. Arkolakis and Muendler provide one of the rare attempts to produce comparable results using data from multiple sources, thereby addressing the associated robustness concerns directly. Analyzing data for four countries, Brazil, Chile, Denmark and Norway, they present a number of interesting stylized facts on patterns of market entry and sales by destination.

Cletus Coughlin and Dennis Novy are concerned with the famous border effect in trade, the empirical finding that trade within countries sizably exceeds the countries' cross border trade, after holding constant for other determinants of trade such as the economic size of the trading partners and the distance between them. Specifically, Coughlin and Novy are interested in the relative magnitude of the estimated effect. To analyze this issue, they combine data sets that allow them to jointly analyze trade within US states, trade between US states and the international trade of US states. Interestingly, they find that the international border effect (that adds to the effect of crossing a state border) is smaller than the state border effect itself, reflecting the strong local concentration of economic activity.

Angela Cheptea's article, in somewhat related fashion, extends conventional analyses of international trade by adding another dimension of trade transactions, domestic trade. More importantly, she argues that the resulting estimates of the border effect provide a reasonable benchmark for assessing the level of integration. Analyzing trade patterns within Europe, Cheptea finds that countries in Central and Eastern Europe have a much larger trade potential than previous estimates suggest.

Although the articles by Coughlin & Novy and Cheptea apply a well-established and widely used tool to analyze empirical patterns of trade, the gravity model, Tibor Besedeš has introduced another interesting technique to describe trade activities, survival analysis. Moving beyond the documentation of the notable finding that the majority of trade relationships are remarkably short-lived, Besedeš demonstrates how survival analysis can be used to explore the effects of trade liberalization. For the North American Free Trade Agreement (NAFTA), he presents evidence

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

that the trade agreement has actually increased the hazard of exports ceasing, along with some other interesting findings.

In recent years, there has been a growing interest in reliable cost-benefit analyses of policy measures to promote integration. For instance, many countries operate export promotion agencies or provide financial assistance to exporting firms, even though relatively little is known about the effectiveness of such measures. Torfinn Harding and Beata Javorcik contribute to this literature by examining to what extent government-sponsored investment promotion intermediaries actually help in overcoming information barriers. In a large cross-country analysis, they find that countries with more professional agencies (as measured, for instance, by the quality of their internet web site) tend to attract a greater volume of foreign direct investment.

For a proper assessment of the evolution of integration over time, it seems particularly useful to take a long-term perspective. Any extension of the time dimension, however, typically comes at the cost of limited data availability. For instance, few economic variables are actually reasonably comparable over long periods. Also, country coverage may be low for century-long comparisons. Still, Vadym Volosovych's article forcefully highlights the sizable benefits of such an approach. Using principal components analysis, he examines capital market integration of 15 industrialized economies for the period from 1875 to 2009. Volosovych not only documents an increase in financial integration in recent years but also identifies factors that help to explain the observed variation in integration over time.

Jarko Fidrmuc, Iika Korhonen and Ivana Bátorová examine another measure of economic integration, the correlation (or synchronization) of national business cycles. They focus particularly on an issue of current interest, the integration of China in the world economy. In their analysis of dynamic correlations of business cycles at different frequencies, Fidrmuc, Korhonen and Bátorová find that the business cycle in China is substantially different from many OECD countries. More notably, their results indicate that countries actively engaged in trade with China tend to display a lower degree of business-cycle synchronization with other OECD countries.

Martin Uebele rounds off the collection of analytical designs and methodological approaches to analyzing market integration in this volume. Uebele's analysis deviates from the other articles along two key dimensions. First, he focuses on prices, implicitly assuming that price differentials (or changes in differentials) reflect market frictions. Second, he performs an outright historical analysis, analyzing the period from 1806 to 1907 (i.e. the first wave of globalization). Applying dynamic factor analysis on a panel of annual wheat prices for up to 67 cities, Uebele

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

finds a particularly strong push towards market integration in the first half of the 19th century. This finding seems remarkable, as major innovations in transportation technology were only realized later.

Overall, the contributions to this special issue of CESifo Economic Studies provide, in our view, an excellent overview of current state-of-the-art work on economic integration, covering a broad, but necessarily limited, range of topics. Hopefully, the articles provide motivation and inspiration for further research in the field.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

CESifo Economic Studies, Vol. 59, 2/2013, 199–222 doi:10.1093/cesifo/ift004

Measuring the Effects of Endogenous Policies on Economic Integration

Jeffrey H. Bergstrand

Department of Finance, Mendoza College of Business, and Kellogg Institute for International Studies, University of Notre Dame, Notre Dame, IN 46556, USA, and CESifo Munich. e-mail: bergstrand.1@nd.edu.

Abstract

Despite widespread anecdotal evidence that lower trade barriers increase international trade, there is little firm quantitative evidence of the 'trade-cost elasticity' of trade flows, one of the two key aggregate statistics that have recently been identified as sufficient to quantify the economic welfare effects of trade-policy liberalizations and/or trade-cost reductions (the other statistic being the import-penetration ratio). In other words, most estimates of the trade-cost elasticity are imprecise and lack consistency. In this article, we discuss two issues that are critical in better assessing empirically the trade-flow and welfare effects of trade liberalizations (or trade-cost changes). The first issue is how to quantify the trade-cost elasticity when trade costs themselves are approximated imperfectly. The second issue is that typical empirical evaluations to estimate the impact of trade-policy liberalizations on trade flows use the 'gravity equation'. However, the self-selection of country pairs into such agreements introduces endogeneity bias in the estimation of the trade-cost elasticity in gravity equations, requiring better identification techniques.(JEL codes: F10; F12 and F13)

Keywords: International trade, economic integration agreements, gravity equations

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

1 Introduction

One of the most prominent aspects of the world's transformation over the last six decades has been increased globalization. Globalization is broader than just increased economic interactions; it also embraces increased exchanges of cultures, attitudes, and mores. In economic terms, increased globalization typically refers to increased international trade flows, investment flows, and migration flows. Of course, such flows are endogenous to changes in—what I term—'natural' and 'unnatural' costs (among other factors to be discussed later).

For tractability, let me define briefly what I mean by the terms 'natural' and 'unnatural' costs. I will refer to those costs associated with geography and technology as 'natural' costs. For instance, an obvious natural cost to shipping a good internationally is distance; larger distances raise the cost of transport. But such costs are also influenced by technology (which is, of course, man-made). However, in my discussion below, I will treat technology as exogenous and treat technological innovations that have

© The Author 2013. Published by Oxford University Press
on behalf of Ifo Institute, Munich. All rights reserved.
For permissions, please email: journals.permissions@oup.com

increased globalization—such as containerization for the shipment of goods internationally—as reductions in 'natural' trade costs.

By contrast, I will refer to 'policy-based' (or 'man-made') costs as 'unnatural' costs. In most cases, these costs are created by government policies—such as tariffs on internationally traded goods—and these can be raised or lowered by policymakers. Despite enormous progress by policymakers worldwide in lowering these 'unnatural' trade, investment, and migration costs since World War II, policy-based impediments to world trade, investment, and migration still exist and remain substantive. As one piece of evidence that the world is not yet—as Friedman (2005) would characterize it—'flat', Eaton and Kortum (2002) in a seminal article show using a calibration exercise that the world is much closer to one of 'autarky'—that is, one where world trade flows are prohibitively expensive—rather than one of frictionless trade.

Surprisingly, despite the fact that international trade economists spend considerable time and effort explaining that reductions in (natural and unnatural) trade costs increase trade—and such augmented trade improves consumers' welfare—we actually know quite little *quantitatively* of the impact of international economic costs on international economic flows. Specifically, we lack precise and consistent quantitative knowledge of the impact of (bilateral) trade costs on (bilateral) trade flows, of investment costs on foreign direct and portfolio investment flows, and of migration costs on migration flows—much less the cross-impact of each of these costs on the other flows. In fact, we also lack firm systematic data on trade costs themselves!

In a recent influential article on international trade flows' determinants, Arkolakis et al. (2012) show that there is a wide class of 'Quantitative Trade Models' in the international trade literature for which the welfare effect of trade-cost reductions can be summarized with two aggregate statistics. One of these statistics is the 'import-penetration' ratio, which is simply one minus the share of aggregate national expenditures on domestically produced goods. The second is the 'trade-cost elasticity', which is the percentage change in trade flows in response to a 1% change in an *ad valorem* measure of 'trade costs' (natural or unnatural). While the first statistic can be estimated fairly precisely using national income accounts data, the second statistic has remained elusive empirically. Arkolakis et al. (2012) actually devote the last substantive section of their article articulating some of the issues related to estimating this trade-cost elasticity (which they denote ϵ). They argue that the principal remaining issues for estimating ϵ are econometric ones. In particular, they note—citing empirical estimates of ϵ in Baier and Bergstrand (2001)—that one of the key econometric issues is the 'standard orthogonality

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

condition', that is, whether measures of trade costs are *exogenous* variables in typical gravity equations.

This article takes up where Arkolakis et al. (2012) left off in discussing econometric (or estimation) issues in measuring the trade-cost elasticity—as well as measuring trade costs themselves—not addressed in their article. In section 2, we address first methods to estimate the trade-cost elasticity in the presence of only imperfect measures of true *ad valorem* trade costs (denoted τ). The world is not so generous as to provide observable measures of true trade costs. In this section, however, we assume such imperfect trade-cost measures are orthogonal to the gravity equation's error term (i.e., no endogeneity bias). In section 3, we address methods to estimate the trade-cost elasticity when such (imperfect) trade-cost measures are not orthogonal to the error term, that is, when there is endogeneity bias. Section 4 concludes.

2 Estimating the Trade-Cost Elasticity with Imperfect Exogenous Trade-Cost Measures

2.1 Background

Arkolakis et al. (2012), henceforth ACRC, recently re-examined theoretically the key elements in measuring the economic welfare 'gains from trade'. Looking back over developments in the international trade literature over the past 30 years, ACRC found that the 'gains from trade' were basically measurable in a wide class of—what they termed—'Quantitative Trade Models' (QTMs). This broad class of models includes endowment-economies with Armington preferences (Anderson 1979; Anderson and van Wincoop 2003), monopolistic competition models with increasing returns to scale (Krugman 1980), Ricardian models with perfectly competitive firms with heterogeneous productivities (Eaton and Kortum 2002), and Melitz-type models with heterogeneous firms, monopolistic competition, and variable and fixed exporting costs (Melitz 2003; Chaney 2008; Redding 2011).

The principal conclusion from the ACRC article is that the welfare gains from trade in this broad class of QTMs could be quantified in terms of two key statistics. The first statistic is the share of aggregate national expenditures on domestically produced goods, which they denote λ_{jt} for country j in year t $(0 < \lambda_{jt} < 1)$. The second statistic is the 'trade-cost elasticity', denoted ϵ (which they actually refer to as the 'trade elasticity'). The trade-cost elasticity is defined as $\epsilon = d \ln X_{ijt} / d \ln \tau_{ijt} < 0$, where X_{ijt} is the aggregate bilateral trade flow from country i to country j in year t and τ_{ijt} is the gross *ad valorem* bilateral

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

trade cost associated with trade flow X_{ijt} (either natural or unnatural cost; $\tau_{ijt} > 1$). The main insight in ACRC is that—for the wide class of QTMs listed above—the welfare gain from a reduction in trade costs simplifies to:

$$d\ln W_{jt} = (1/\epsilon)d\ln \lambda_{jt} \tag{1}$$

where W_{jt} is country j's welfare (or real income of the representative consumer) in year t.

The economic intuition for this simple and common result for all these QTMs is that a reduction in trade costs for importing country j improves the country's terms of trade. This improvement in the terms of trade can be inferred from the changes in its relative import demands from various countries. The reason that this simple and 'general' result surfaces across a broad class of QTMs is that it relies on a very small set of feasible assumptions, common to this broad class of models. Each of the models listed above shares only four primitive assumptions: Dixit–Stiglitz preferences; one factor of production (typically, labor); linear cost functions; and perfect or monopolistic competition. Also, all the QTMs share three 'macro-level' restrictions: multilateral trade balance; aggregate profits are a constant share of aggregate revenue; and import demand systems are constant-elasticity-of-substitution (CES).

It turns out that all of the QTMs listed above share the fact that bilateral import flows can be described in equilibrium by a 'gravity equation'. For instance, in the Armington endowment-economy model in Anderson and van Wincoop (2003), the implied gravity equation is:

$$X_{ijt} = Y_{it}Y_{jt}\left(\frac{(\beta_{it}p_{it})^{-(\sigma-1)}\tau_{ijt}^{-(\sigma-1)}}{\sum_{k=1}^{K} Y_{kt}(\beta_{kt}p_{kt})^{-(\sigma-1)}\left(\tau_{kjt}\right)^{-(\sigma-1)}}\right) \tag{2}$$

where X_{ijt} is the trade flow from i to j in year t, Y_{it} (Y_{jt}) is gross domestic product (GDP) of i in t, β is an unobservable preference parameter for i's good, p_{it} is the price of i's good, and σ is the elasticity of substitution in consumption.[1] So here, $\epsilon = -(\sigma - 1)$.

In the context of CES preferences, increasing returns to scale, and monopolistic competition, the Krugman (1980) model (summarized in

[1] See also Anderson (1979) and Bergstrand (1985).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Baier and Bergstrand (2001) and Feenstra (2004)) yields a gravity equation:

$$X_{ijt} = N_{it}Y_{jt}\left(\frac{w_{it}^{-(\sigma-1)}\tau_{ijt}^{-(\sigma-1)}}{\sum_{k=1}^{K} N_{kt}(w_{kt})^{-(\sigma-1)}\left(\tau_{kjt}\right)^{-(\sigma-1)}}\right) \tag{3}$$

where N_{it} is the number of products/producers in i, and w_{it} is the wage rate for labor in i. Here also, $\epsilon = -(\sigma - 1)$.

In the context of a Ricardian model of international trade with perfect competition and heterogeneous firms/productivities, Eaton and Kortum (2002) generate a gravity equation:

$$X_{ijt} = T_{it}Y_{jt}\left(\frac{w_{it}^{-\gamma}\tau_{ijt}^{-\gamma}}{\sum_{k=1}^{K} T_{kt}(w_{kt})^{-\gamma}\left(\tau_{kjt}\right)^{-\gamma}}\right) \tag{4}$$

where T_{it} is the 'state of technology' in i, and γ is an index of the heterogeneity of firms' productivities (or comparative advantages). In the Eaton–Kortum model, $\epsilon = -\gamma$.

Finally, allowing for heterogeneous firms/productivities, monopolistic competition, increasing returns to scale, and fixed exporting costs, Melitz (2003) yields a gravity equation:

$$X_{ijt} = N_{it}Y_{jt}\left(\frac{w_{it}^{-\gamma}\tau_{ijt}^{-\gamma}f_{ijt}^{-[\gamma/(\sigma-1)-1]}}{\sum_{k=1}^{K} N_{kt}(w_{kt})^{-\gamma}\left(\tau_{kjt}\right)^{-\gamma}\left(f_{kjt}\right)^{-[\gamma/(\sigma-1)-1]}}\right) \tag{5}$$

where f_{ijt} is the fixed costs of exporting from i to j in t and, as noted in Chaney (2008), $\gamma = (\sigma - 1) + [\gamma - (\sigma - 1)]$. In this model's context, $(\sigma - 1)$ is the 'intensive margin' elasticity of trade with respect to variable trade costs (τ_{ij}), and $\gamma - (\sigma - 1)$ is the 'extensive margin' elasticity of trade with respect to variable trade costs. Here also, $\epsilon = -\gamma$.

Thus, while the primary goal of ACRC is to show that the welfare gains from trade liberalization across a broad range of QTMs are basically summarized by a simple function of two common key 'aggregate statistics', λ_{jt} and ϵ, a second major insight of the article is the importance for measuring $d\ln W_j$ of estimating the variable trade-cost elasticity (ϵ) with both *consistency* as well as *precision*.

ACRC argue in the final substantive section of their article (titled 'Estimating the Trade Elasticity') that the gravity equation provides a

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

very useful approach to estimating ϵ. For all the models above (in the absence of zero trade flows), one can express the implied gravity equation as:

$$\ln X_{ijt} = A_{it} + B_{jt} + \epsilon \ln \tau_{ijt} + \nu_{ijt} \tag{6}$$

where A_{it} is an exporter-time (fixed) effect, B_{jt} is an importer-time (fixed) effect, and ν_{ijt} is an error term. However, a key issue they note is that proper and consistent estimation of ϵ requires the 'standard orthogonality condition' holding: independence of τ_{ijt} with the gravity equation's error term (ν_{ijt}). Yet, ACRC stop here, noting that this is principally an *econometric* issue for which they have 'little to contribute'.

Where ACRC end, we begin. In this article, we consider two issues relevant to estimating ϵ. First, we do not observe true values of τ_{ijt}; typically, we explain trade flows using proxies for τ such as bilateral distance, which obscures identification of ϵ because of the unknown relationship between bilateral distance ($dist_{ij}$) and unobservable true bilateral trade cost τ_{ijt}. In the remainder of this section, we address methods to infer ϵ when τ is unobserved. A second issue is selection bias. Suppose country pairs self-select into economic integration agreements (EIAs). For instance, trade between a pair of countries may be below its 'natural' level because of barriers to trade unobservable to the econometrician. If trade is low where such barriers exist, inducing a country pair to form an EIA to reduce this impediment, coefficient estimates on right hand side (RHS) dummy variables for these EIAs in gravity equations may be biased downward (referred to as 'negative selection'). We address this issue in section 3.

2.2 Estimation issues

In a recent article, Bergstrand et al. (2013) address a method for estimating the trade-cost elasticity in the presence of imperfect measures of true *ad valorem* trade costs. For instance, consider the Krugman (1980) model of international trade. As discussed in ACRC, this model satisfies the four primitive assumptions in the broad class of QTMs discussed above, as well as the three 'macro-level' restrictions. This model yields a gravity equation of the form in equation (3):

$$X_{ij} = \frac{Y_i Y_j}{Y^W} \left(\frac{(Y_i/L_i)^{-\sigma} \left(\tau_{ij}\right)^{-(\sigma-1)}}{\sum_{k=1}^{N} (Y_k/Y^W)(Y_k/L_k)^{-\sigma} \left(\tau_{kj}\right)^{-(\sigma-1)}} \right) u_{ij} \tag{7}$$

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

where we have replaced w_i with Y_i/L_i because with one factor, labor, $Y_i = w_i L_i$. Under the assumption of market clearing, this gravity equation can be estimated structurally by assuming K market-clearing equations:

$$Y_i = \sum_{j=1}^{K} X_{ij} \quad i = 1,\ldots,K, \tag{8}$$

Assuming no zeros in trade flows, one can take the logarithm of equation (7) and obtain:

$$\begin{aligned} \ln X_{ij} = &-\ln Y^W + \ln Y_i + \ln Y_j - \sigma \ln(Y_i/L_i) - (\sigma - 1) \ln \tau_{ij} \\ &- \ln\left(\sum_{k=1}^{N} (Y_k/Y^W)(Y_k/L_k)^{-\sigma} \left(\tau_{kj}\right)^{-(\sigma-1)} \right) + \nu_{ij} \end{aligned} \tag{9}$$

Bergstrand et al. (2013) suggest three possible methods to estimate σ. One possible method is to estimate σ based on the coefficient for $\ln(Y_i/L_i)$. However, $\ln(Y_i/L_i) = \ln w_i$. Since w_i is endogenous, it is likely correlated with the error term ν_{ij}; coefficient estimates would likely be biased. A second possibility is to estimate σ from the coefficient estimate for $\ln \tau_{ij}$. However, in reality, τ_{ij} is not observable. Typically, one uses proxies for trade costs such as bilateral distance and dummy variables for the presence or absence of an EIA between country pairs using:

$$\ln X_{ij} = A_i + B_j + \rho(1-\sigma) \ln dist_{ij} + \psi(1-\sigma) EIA_{ij} + \nu_{ij} \tag{10}$$

where $dist_{ij}$ is the bilateral distance between the economic centers of regions i and j, and EIA_{ij} is a dummy variable assuming a value of 1 (0) if the two regions share (do not share) an EIA. Bergstrand et al. (2013) suggest a third way to estimate σ. Estimating equation (10) using fixed-effects A_i and B_j and exponentiating yields a set of estimates $\widehat{dist_{ij}^{\rho(1-\sigma)}}$ and $\widehat{e^{\psi(1-\sigma)EIA_{ij}}}$. The product of these two estimates is an estimate of $\widehat{\tau_{ij}^{1-\sigma}}$. Bergstrand et al. (2013) show that one can substitute $\widehat{\tau_{ij}^{1-\sigma}}$ into equation (7) (which eliminates the u_{ij}), $\widehat{\tau_{kj}^{1-\sigma}}$ into the same equation's analogue for X_{kj}, take the ratio of the two equations for X_{ij} and X_{kj}, and solve for an estimate of σ ($\hat{\sigma}$).

Using an extensive set of Monte Carlo simulations, Bergstrand et al. (2013) show that this technique provides *consistent* and *precise* estimates of gravity equation parameters, the elasticity of substitution in consumption (σ)—which is the 'trade elasticity' in this context—*and* the welfare changes from increases or decreases in trade costs.

To illustrate the relevance of this technique to empirical applications, Bergstrand et al. (2013) apply the framework to the well-known

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

McCallum 'border-puzzle' issue. The McCallum border puzzle refers to McCallum (1995). McCallum (1995) applied a traditional gravity equation—ignoring the role of relative prices discussed earlier—to the trade flows among Canadian provinces, among US states, and between Canadian provinces and US states, including a dummy variable for the presence of the 'national border', i.e., the US–Canadian border. McCallum found an enormous quantitative effect of this national border. He found that two typical Canadian provinces had *22 times* more goods trade than a typical pairing of a Canadian province with a US state.

This 'mind-boggling' empirical finding spurred a cottage industry of papers. Of the more important subsequent articles investigating this finding further, Anderson and van Wincoop (2003) argued that there were two essential elements missing from the analysis of the border in McCallum (1995). First, they derived gravity equation (2) above, which emphasizes the importance of relative prices in influencing trade flows, and demonstrated that estimating a traditional gravity equation *ignoring* relative prices would lead to omitted variables bias in the gravity-equation parameter estimates. Second, they emphasized that a better estimate of the border effect would be a *general equilibrium* comparative static estimate, rather than simply the partial effect. Anderson and van Wincoop themselves chose to use a nonlinear least squares estimation procedure for estimating the parameters and then used a nonlinear program to estimate the general equilibrium comparative static effects.

However, one of the shortcomings of the Anderson and van Wincoop approach is that it cannot *estimate* the actual elasticity of substitution in consumption. That is a problem when trying to estimate welfare effects; to do any welfare calculations, they would have to assume some value of σ. This elasticity is essential for welfare calculations.

Bergstrand et al. (2013) replicated the Anderson and van Wincoop (2003) study and its empirical results, assuming as there symmetric effects of the national border on trade (i.e., a single dummy coefficient estimate, regardless of the direction of the trade flow between the two countries). Bergstrand et al. (2013) showed that—owing to correlated errors (say, owing to omitted variables bias)—estimation of gravity-equation parameters would likely yield consistent parameter estimates if exporter and importer fixed effects were used instead. However, for the same gravity-equation parameter estimates, the Bergstrand et al. (2013) approach generated an *estimated* value of the trade elasticity of (approximately) 7, whereas Anderson and van Wincoop (2003) considered the effects of various alternative elasticities ranging from 2 to 20 (choosing 5 as a representative elasticity of substitution). Bergstrand et al. (2013) showed that the welfare effects of trade-cost changes—here, the elimination of the national

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

border—were not only quantitatively different between the two approaches, but also *qualitatively* different. To see the importance of consistent and precise estimation of the 'trade elasticity' as emphasized in Arkolakis et al. (2012), the approach in Anderson and van Wincoop (2003) using their preferred assumed elasticity of 5 implied a welfare gain for Canada but a welfare *loss* for the United States from eliminating the US–Canadian national border. By contrast, Bergstrand et al. (2013) found using the same *parameter* estimates—but using their *estimated* trade elasticity—a welfare gain for Canada *and* a welfare gain for the United States of eliminating the two countries' common border.

3 Estimating the Trade-Cost Elasticity with Imperfect Endogenous Trade-Policy Measures

The last substantive section of Arkolakis et al. (2012) argued that a useful methodology empirically to estimate the trade-cost elasticity was the gravity equation. However, the authors reminded the reader that estimates using the gravity equation must satisfy the 'standard orthogonality condition' to generate consistent parameter estimates. In the previous section of the article, we discussed the usefulness of recent gravity-equation estimates under a strong assumption: that the explanatory (or RHS) variables satisfied the standard orthogonality condition, that is, that they were 'exogenous'. In the case of bilateral distance and other natural trade costs, such an assumption is quite feasible. However, when the policy maker is interested in evaluating *ex post* the effects of measures of 'policy' on trade flows, a concern about the 'exogeneity' of the policy measures arises. In this section, we address recent methodologies to estimate with consistency and precision the effects of endogenous policies on trade flows or—in more general terms—the 'effects of endogenous policies on economic integration'.

Nearly 20 years have passed since Trefler (1993) showed that *ex post* empirical estimates of the effects of U.S. trade policies on U.S. imports were underestimated *considerably* by not accounting for the endogeneity of trade-policy measures.[2] The downward bias of such estimates was confirmed later in Lee and Swagel (1997) for a broad cross-section of country pairs' bilateral trade flows and trade-policy measures. More recently, Baier and Bergstrand (2007) argued that—in the context of EIAs and gravity equations—the endogeneity bias may well be attributed to *self-selection* bias. Using panel techniques to account for endogeneity bias,

[2] Trefler (1993) showed that—after accounting for endogeneity—the effect of trade liberalizations was 10 times that estimated otherwise.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Baier and Bergstrand (2007) showed—in the spirit of Trefler (1993) and Lee and Swagel (1997)—that previous estimates of the effects of EIAs on bilateral trade flows were underestimated considerably; the authors showed that after accounting for endogeneity, the effects of EIAs were five times that estimated otherwise. Several articles have shown—using instrumental variables or panel techniques—the downward bias of EIAs' effects when not accounting for endogeneity bias, including most recently Baier et al. (2011).

3.1 Sources of trade-policy endogeneity in gravity equations

In determining the potential correlation between the gravity equation's error term with the EIA dummy, one first needs to consider what determines the likelihood of a free trade agreement (FTA) between a pair of countries. Although trade economists have examined empirically for many years the determinants of tariff rates and nontariff barrier levels across industries and across countries, little empirical work has examined the determinants of EIAs until Mansfield and Reinhardt (2003) and Baier and Bergstrand (2004a). The former study examined political determinants of EIAs, whereas the latter presented a theoretical and empirical model of economic determinants of FTAs. Baier and Bergstrand (2004a) found strong empirical evidence that pairs of countries that have FTAs tend to share economic characteristics that their theory suggests should enhance economic welfare of the pairs' representative consumers. For instance, two countries tend to have an FTA the larger and more similar their GDPs, the closer they are to each other but the more remote the pair is from the rest-of-the-world (*ROW*), and the wider (narrower) the difference in their relative factor endowments with respect to each other (the *ROW*). But these include the same factors that tend to explain large trade flows. Thus, in terms of observable economic characteristics, countries with FTAs have 'chosen well', in the sense that most country pairs with FTAs tend to have the economic characteristics associated with considerable trade and with (in theory) welfare-enhancing net trade creation. Yet, the estimated probit functions in Baier and Bergstrand (2004a) have pseudo-R2 values of only 70%, still leaving considerable unobserved heterogeneity.

3.1.1 *Selection bias*

How does the unobserved heterogeneity in trade flow determinants matter? For instance, error term ν_{ij} in equation (10) may be representing unobservable (to the empirical researcher) policy-related barriers—tending to reduce trade between two countries—that are not accounted for by standard gravity equation RHS variables. As an example, suppose two countries have extensive unmeasurable domestic regulations (e.g., internal

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

shipping regulations) that inhibit trade (causing X_{ij} to be low). The likelihood of the two countries' governments selecting into an FTA may be high if there is a large expected welfare gain from potential bilateral trade creation if the EIA deepens liberalization beyond tariff barriers into domestic regulations (and other nontariff barriers). Thus, EIA_{ij} and the intensity of domestic regulations may be positively correlated in a cross-section of data, but the gravity equation error term v_{ij} and the intensity of domestic regulations may be negatively correlated. This reason suggests that EIA_{ij} and v_{ij} are negatively correlated, and the EIA coefficient will tend to be underestimated.

In support of this argument, numerous authors have noted that one of the major benefits of regionalism is the potential for 'deeper integration'. Lawrence (1996, p. xvii) distinguishes between 'international policies' that deal with border barriers, such as tariffs, and 'domestic policies' that are concerned with everything 'behind the nation's borders, such as competition and antitrust rules, corporate governance, product standards, worker safety, regulation and supervision of financial institutions, environmental protection, tax codes . . .' and other national issues. The General Agreement on Tariffs and Trade (GATT) and World Trade Organization (WTO) have been remarkably effective in the post-WWII era reducing border barriers. However, these institutions have been much less effective in liberalizing the domestic policies just named. As Lawrence states it, 'Once tariffs are removed, complex problems remain because of differing regulatory policies among nations (p. 7).' He argues that in many cases, EIA 'agreements are also meant to achieve deeper integration of international competition and investment' (p. 7). Gilpin (2000) echos this argument: 'Yet, the inability to agree on international rules or to increase international cooperation in this area has contributed to the development of both managed trade and regional arrangements' (p. 108; italics added). Preeg (1998) concludes:

> [Free] trade agreements over time, however, have tended to include a broader and broader scope of other trade-related policies. This trend is a reflection, in part, of the fact that as border restrictions [tariffs] are reduced or eliminated, other policies become relatively more important in influencing trade flows and thus need to be assimilated in the trade relationship (p. 50).

We believe this omitted variable (selection) bias is the major source of endogeneity facing estimation of FTA effects in gravity equations using cross-section data. Moreover, the arguments above suggest that policy makers' decisions to select into an EIA are likely related to the level of trade (relative to its potential level), and not to recent changes in trade

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

levels. Thus, the determinants of FTA are likely to be cross-sectional in nature.

3.1.2 *Simultaneity bias*

Consider the potential endogeneity bias created by simultaneity. As discussed earlier, there exists a large empirical literature in international trade on the effects of multilateral tariff and nontariff barriers on multilateral trade volumes, and the simultaneous effects of these trade volumes on multilateral barriers using cross-industry and cross-country data for particular years, cf., Trefler (1993) and Lee and Swagel (1997). Simultaneity may be an issue for EIA_{ij} in cross-section gravity equations, motivated as in these two studies. For example, holding constant typical gravity equation RHS variables, two countries (say, the United States and China) that possibly trade more than their 'natural' level, as predicted by a typical gravity equation, may create political pressures to avoid trade liberalization or possibly raise trade barriers. This would cause a negative simultaneity bias in the EIA coefficient estimate. On the other hand, the governments of two countries that trade more than their gravity-equation-suggested 'natural' level might be induced to form an EIA because there might potentially be less 'trade diversion' due to their extensive trading relationship, suggesting a positive simultaneity bias. However, as just noted since the decisions to select into EIAs are likely influenced by the levels of trade relative to natural levels, recent changes in trade levels are not likely to influence EIA formations.

3.1.3 *Measurement error*

Measurement error in an explanatory variable, such as an EIA dummy, is generally associated with negative bias (in absolute terms) in the variable's coefficient. For instance, with the classical 'errors-in-variables' assumption, the 0-1 EIA dummy variable would be correlated positively with the measurement error if the true trade-policy variable (say, the tariff rate) was assumed uncorrelated with the measurement error. In equation (10)'s context, the correlation between EIA_{ij} and the error term ν_{ij} would be negative, leading to the classical 'attenuation bias' of *EIA*'s coefficient estimate toward zero. This may be part of the reason—but neither the entire, nor even the most important, reason—*EIA* coefficient estimates have been underestimated.

Of course, the best method for eliminating this bias is construction of a continuous variable that would more accurately measure the degree of trade liberalization from various EIAs. If EIAs only eliminated bilateral tariff rates, one would ideally measure this liberalization with a change in the *ad valorem* tariff rate (for which data are poor). However, EIAs liberalize trade well beyond the elimination of tariffs. Calculation of such

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

measures is beyond the scope of this particular study, but is a useful direction for future research. Our goal rather is to discuss reliable estimates of the treatment effect of an EIA, similar to the 0-1 variable representing program participation in empirical labor economics. Thus, I constrain the discussion here to more accurate estimation of the *ex post* partial 'average treatment effect' (ATE) of an EIA dummy on trade flows, as has been used in the gravity equation literature for five decades.

3.2 Estimation using cross-section data for a single year

With cross-section data, standard econometric techniques to address omitted variables (and selection) bias include estimation using instrumental variables (IV) and Heckman control functions. Alternatively, with panel data, fixed effects and first differencing can be used to treat endogeneity bias; we discuss panel approaches in the next section.

Baier and Bergstrand (2002) was the first article to follow in the spirit of Trefler (1993) to apply IV to account for the endogeneity of EIAs in estimating their effect on trade flows. The first stage of the approach is to estimate the likelihood of a pair of countries having an EIA; this can be done using probit, logit, or linear probability functions. Several studies have used probit functions based upon Baier and Bergstrand (2004a) to estimate the probability of an EIA. Although most studies have found that the probit estimates provide 'good' predictions for the first stage of the estimation, the problem lies in 'identification' for the second stage (i.e., the 'exclusion restriction'). As with any IV application, there needs to be at least one variable explaining the probability of a pair having an EIA that does *not* also influence trade flows. Herein lies the practical problem of using IV to alleviate the endogeneity bias.

Various authors have had alternative success in identification for the second-stage regression. Baier and Bergstrand (2002, 2004b) tried various IV and Heckman control function approaches to account for endogenous EIAs. Before trying IV, Baier and Bergstrand using ordinary least squares (OLS) found an ATE for EIAs (defined there as free trade agreements and deeper EIAs only) of 34% using trade flows for year 2000 (with 7302 observations). The authors then tried several two-stage procedures. One way to achieve identification in the second stage using probit in the first stage derives from the nonlinearity of the probit function. Baier and Bergstrand (2004b) found that with such identification an EIA had a negative and statistically significant effect on trade flows. Another IV estimate used a first-stage probit including the standard gravity-equation covariates and also a measure of remoteness, a measure of bilateral absolute differences in capital–labor ratios, and a measure of the difference of the pair's capital–labor ratio relative to that of the *ROW*. The second-stage EIA

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

coefficient estimate was economically and statistically insignificant. Moreover, the 'identifying' variables had often been included in gravity equations in the past and may be correlated with the gravity-equation error term. Baier and Bergstrand (2004b) also obtained first-stage probit estimates using political variables. The resulting second-stage EIA coefficient estimate was also economically and statistically insignificant.

Baier and Bergstrand (2004b) also tried linear probability models for the first-stage estimates. For the first-stage regression using the gravity covariates and economic identification variables (remoteness and two capital–labor variables), the second-stage EIA coefficient estimate of 2.51 implied that the partial ATE of an EIA was to increase bilateral trade by over *1100%*. Using political variables in the first-stage instead yielded in the second stage an ATE of 733%. One of the benefits of using a linear probability model is the econometrically feasible inclusion of fixed effects in the first stage. Using the previously described economic variables in the first stage with fixed effects led to a statistically insignificant second-stage EIA coefficient estimate of 0.41; however, this implied an economically plausible 51% EIA partial effect. By contrast, using the political variables in the first stage including fixed effects, the second-stage EIA coefficient estimate was a statistically significant −3.97, implying the EIA lowered trade by 92%. The vast differences in the second-stage EIA coefficients led Baier and Bergstrand (2004b, 2007) to conclude that IV approaches yielded unstable results, likely owing predominantly to the inability to find economic or political variables that satisfied the 'exclusion restriction' with confidence. Another author that has tried to use IV similarly with mixed results is Magee (2003).[3]

More recently, Egger et al. (2011) accounted for the endogeneity of EIAs in a cross-section analysis of EIAs while also accounting for zeros in trade flows and also estimating general equilibrium—alongside partial equilibrium (ATE)—effects, combining in one analysis the insights of Baier and Bergstrand (2002, 2007), Helpman et al. (2008), and Anderson and van Wincoop (2003). Using trade flows from year 2000 applying IV as above, they found that not accounting for self-selection of country pairs into trading and into EIAs (by using a first-stage bivariate probit model) led to a downward bias of 75% in the EIA's effect, with 45 percentage points of the bias attributed to the endogeneity of EIAs. The

[3] An alternative method for estimating the ATE of EIAs uses Heckman's control-function approach. Baier and Bergstrand (2002) estimated similar specifications using this alternative approach with qualitatively similar findings; the control-function approach does not solve the endogeneity bias issue either. The likely problem is this: the vast number of variables that are correlated cross-sectionally with the probability of having an FTA are also correlated with trade flows, preventing elimination of the endogeneity bias using cross-section data.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

bias attributed to ignoring general equilibrium effects was minor, explaining only 4 percentage points of the downward bias.

3.3 Estimation using panel data

Given the problems associated with accounting for endogeneity of EIAs using instrumental variables and cross-section data, Baier and Bergstrand (2007), or BB, argued that a better approach to eliminate endogeneity bias of EIAs is to use panel techniques. In the context of the theory and endogenous self-selection of country pairs into EIAs, BB argue that one method to obtain consistent estimates of the ATE of EIAs is by fixed effects estimation of:

$$\ln X_{ijt} = \beta_0 + \beta_1 EIA_{ijt} + \eta_{ij} + \delta_{it} + \psi_{jt} + \epsilon_{ijt} \tag{11}$$

where η_{ij} is a country-pair fixed effect to capture all time-invariant unobservable bilateral factors influencing nominal trade flows, and δ_{it} and ψ_{jt} are exporter-time and importer-time fixed effects, respectively, to capture time-varying exporter and importer GDPs as well as all other time-varying country-specific unobservables in i and j influencing trade, *including* the exporter's and importers' 'multilateral price/resistance' terms (cf., Anderson and Wincoop 2003). We refer to this as the Fixed-Effects (FE) specification. It is important to note that, in most gravity-equation applications using a comprehensive set of RHS variables, the vast bulk of 'bilateral' trade-cost variables are time invariant, such as bilateral distance, common border, common language, etc. For instance, in Helpman et al. (2008), the only time-varying bilateral trade-cost variables are their EIA dummy and a dummy for the presence or absence of a currency union. As explained in BB, using panel data ATEs of EIAs (represented by β_1) estimated using equation (11) are likely to be consistent and precise. BB showed that the ATE of the typical EIA on nominal trade flows was 0.76, implying that the typical EIA increased bilateral trade by 114% after 10 years.

BB also used an alternative specification using first-differencing:

$$\Delta \ln X_{ij,t-(t-5)} = \beta_0 + \beta_1 \Delta EIA_{ij,t-(t-5)} + \delta_{i,t-(t-5)} + \psi_{j,t-(t-5)} + \upsilon_{ij,t-(t-5)} \tag{12}$$

We refer to this as the First-Difference (FD) specification. Note that the bilateral country-pair fixed effects are eliminated; however, the exporter-time and importer-time fixed effects are retained to capture time-varying exporter and importer GDPs and multilateral price terms. The latter effects were ignored in Foster et al. (2011), creating potential omitted variables bias.

Standard econometric discussions of treating endogeneity bias using panel data focus on a choice between estimation using the FE and FD

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

specifications, cf., Wooldridge (2002, Ch. 10). As Wooldridge notes, when the number of time periods (T) exceeds two, the FE estimator is more efficient under the assumption of serially uncorrelated error terms ϵ_{ijt}. The FD estimator is more efficient (when $T > 2$) under the assumption that the error term ϵ_{ijt} follows a random walk (i.e., that the error term $\upsilon_{ij,t-(t-5)} = \epsilon_{ijt} - \epsilon_{ij,t-5}$ is white noise).[4]

First-differencing the panel data yields some potential advantages over fixed effects. First, it is quite plausible that the unobserved factors influencing the likelihood of an EIA (say, trade below its 'natural' level) are likely slow moving and hence serially correlated. If the ϵ_{ijt} are highly serially correlated, the inefficiency of FE is exacerbated as T gets large. This suggests that differencing the data will increase estimation efficiency for our large T panel. Second, aggregate trade flow data and real GDP data are likely 'close to' unit-root processes. Using FE is equivalent to differencing data around the mean (in BB's sample, year 1980); this may create a problem, as T is large in our panel. As Wooldridge (2000, p. 447) notes, if the data follow unit-root processes and T is large, the 'spurious regression problem' can arise in a panel using FE. FD yields data that deviates from the previous period of our panel, and thus is closer to a unit-root process. Consequently, the preferred estimation technique in BB and Baier et al. (2011) is the FD approach.

One of the other potential contributions of BB's panel methodology was to show that the full impact of EIAs on trade flows took 10–15 years. One reason is that most EIAs are 'phased-in' over 5–10 years. The second reason is the lagged effect of the trade-cost changes (such as terms-of-trade changes) on trade flows. As in BB, using a panel allows for differentiating the shorter-term effects (5 years) from the longer-term effects (5–15 years). Using the FD specification, BB found that the ATE of an EIA (FTA or higher degree of economic integration) was 0.61, imply an 84% increase in trade after 15 years.

While such positive ATE estimates for EIA dummy variables were interpreted in the context of either Armington or Krugman models as EIAs increasing trade volumes of existing homogeneous firms (the 'intensive margin'), consideration of zeros in bilateral trade, fixed export costs, and firm heterogeneity have led researchers more recently to examine various 'extensive margins' of trade. Such extensive margins fall under three general categories: country, goods, and firm. The existence of zeros in aggregate bilateral trade flows among many country pairs has

[4] When the number of time periods is exactly two ($T = 2$), estimation with FE and FD produce identical estimates and inferences; then, FD is easier to estimate. When $T > 2$, the choice depends upon the assumption one wants to make about the distribution of the error term ϵ_{ijt}.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

led some researchers to explore the probability that a pair of countries trades at all; to the extent that an EIA affects this probability, this changes the *country* extensive margin of trade and potentially economic welfare.

However, the few empirical studies to date using gravity equations for a large number of country pairs and EIAs to examine extensive- and intensive-margin effects of EIAs have led to two puzzling results. First, two studies have used cross sections for a single year. Helpman et al. (2008), or HMR, found evidence using a cross-section and a two-stage estimator that EIAs influenced the country extensive margin, but had *no* significant effect on the intensive margin of trade (for existing firms). Egger et al. (2011) explored the country extensive and intensive margins also using a cross-section and a two-stage estimator and found in contrast a significant positive effect of EIAs on the intensive margin in their preferred specification, but no extensive margin effect.[5] The absence of country extensive-margin effects from an EIA suggests that trade liberalization does *not* lead to increases in varieties of goods from new trade partners, a potentially important source of welfare gains. The opposite EIA effect finding of the two articles is a puzzle.

A second margin is known as the 'goods' margin of trade. Hummels and Klenow (2005), or HK, introduced this notion by examining zeros in bilateral trade flows at *highly disaggregated product-category levels*.[6] The motivation for HK was to explore in a cross-section a fundamental question: Do large economies export more because they export larger quantities of a given good (intensive margin) *or* a wider set of goods (extensive margin)?[7] They found in their cross-section that 60% of larger exports of large economies was attributable to the extensive margin of 'goods' trade; specifically, as the exporter country's economic size grew, it exported a larger number of product categories (or 'goods') to more markets. The finding also that larger economies import more goods from more partners is important because it suggests an *improvement in welfare* owing to the consumption of more varieties; yet, 40% of the increased trade was still explained by the intensive margin in this cross-section. However, HK did not explore the relationship between trade liberalizations and the intensive and extensive goods margins of trade.

Only three empirical studies have explored the effect of trade liberalizations—and, in particular, EIAs—on the intensive and extensive goods

5 They also evaluated whether their results were biased by omitting firm-heterogeneity, but concluded that firm-heterogeneity had no significant effect (also in contrast with HMR). Their preferred specification accounted for endogeneity of EIAs.

6 By contrast, both HMR and Egger et al. (2011) used only bilateral *aggregate* trade flows to determine zeros in trade.

7 Each 'good' was a 6-digit SITC category. They also explored the effects of country size and per capita GDP on the quality of goods exported, as well as the two margins.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

margins of trade using the HK methodology. The earliest study using the HK decomposition to explore this issue is Hillberry and McDaniel (2002), focusing solely on the North American Free Trade Agreement (NAFTA). Although they do not attempt to establish causal effects from NAFTA to trade increases, they provide a decomposition of post-NAFTA trade among the three partners into goods intensive and extensive margins using 4-digit Standard International Trade Classification (SITC) data. They find evidence of both margins changing between 1993 and 2001. Kehoe and Ruhl (2009) examined NAFTA, the earlier Canada–US FTA trade liberalization, and some structural transformations using a modified version of the HK decomposition methodology and applied to a series of cross-sections. Similar to Hillberry and McDaniel (2002), they do not conduct an econometric analysis trying to explain the effect of NAFTA (or the Canada–US FTA) on trade flows conditional on other variables. They decompose actual goods extensive- and intensive-margin changes post-agreement also using 4-digit SITC data for goods categories from Feenstra et al. (2005). They find significant evidence of both extensive and intensive margin changes using their modified HK decomposition methodology. Both studies' evidence of goods intensive and extensive margins of trade expanding following the signing of NAFTA suggests the *need* for a comprehensive *econometric* analysis (conditional on other covariates) of the effects of EIAs in general on the goods intensive and extensive margins of trade, in the spirit of HK's original analysis of the effect of country size and per capita GDP on the two goods' margins.[8] However, the one panel study that did such a comprehensive analysis—Foster et al. (2011)—examined only short-run (5-year window) EIA effects motivated by a traditional gravity equation (ignoring multilateral price/resistance terms); they found economically small extensive margin effects and no intensive margin effects.[9]

In the context of the recent developments in the trade literature emphasizing intensive versus extensive margin effects, the panel approach in BB allows for differential timing of these trade-margin effects. In reality, one would expect that the intensive margin would be affected by a trade-cost change sooner than the extensive margin, because intensive margin changes in volumes do not require any start-up costs. Such costs—critical to the extensive margin—may delay the entry of new firms into exporting,

[8] Using a methodology similar to HK for estimating the goods extensive margin, Feenstra and Kee (2007) provided an econometric analysis of the effect of NAFTA on the extensive margin of Mexico's exports to the United States; they found a significant effect of NAFTA's reduction in tariff rates on this margin.

[9] With the short window, the authors could not address short-run versus long-run effects, likely missed phase-in and lagged terms-of-trade effects, and did not distinguish between alternative types of EIAs (in terms of depth of economic integration).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

and thus we should expect the intensive margin to be influenced in the shorter term and the extensive margin in the longer term, as the results in Bernard et al. (2009) show. The panel data approach allows for evaluating this issue.[10]

BB did not estimate differential effects of various types of EIAs (in terms of depth of integration) on trade flows. Magee (2008) and Roy (2010) using the methodology of BB found that trade flows were impacted by larger amounts for customs unions relative to FTAs. However, no empirical study has until recently examined the differential impact of FTAs relative to deeper EIAs on the extensive *versus* intensive margins—much less the differential *timing* of such effects.[11]

Baier et al. (2011) recently addressed the effects of EIAs on the 'goods' extensive and intensive margins of trade. First, they extended Baier and Bergstrand (2007)'s panel econometric methodology for the (partial) effects of EIAs on aggregate trade flows using a gravity equation to examine in a large country-pair setting the effects of virtually all EIAs on the extensive and intensive goods margins, using the HK trade-margin-decomposition methodology. In the context of an econometric analysis, they are the first to find economically and statistically significant EIA effects on both the intensive *and* extensive (goods) margins in the context of a large number of country-pairs, EIAs, and years, in contrast to HMR and Egger et al. (2011).

Second, Baier et al. (2011) allowed for various *types* of EIAs—one-way preferential trade agreements (OWPTAs), two-way preferential trade

[10] These differential timing effects were ignored in Foster et al. (2011). As discussed earlier, two recent theoretical articles suggest a reason for the low trade-cost elasticity of trade flows in macroeconomic analyses using time-series data and the relatively higher trade-cost elasticities of trade in cross-sectional trade analyses. Ruhl (2008) explains this puzzle by noting that the macroeconomic time-series approach is estimating the intensive-margin effect of trade, whereas the trade literature's cross-sectional approach is capturing the extensive-margin effect, due to export fixed costs for new producers delaying trade effects and entry. In a complementary approach, Arkolakis et al. (2011) present a demand-oriented staggered-adjustment "Calvo-pricing" approach to explain the lower time-series elasticity in terms of solely an intensive-margin effect, and the higher long-run cross-section trade-cost elasticity capturing the longer-term extensive-margin elasticity as well.

[11] It is useful to note here a parallel literature examining the effect of GATT and/or WTO membership on trade flows. For brevity, we note that there now appears little convincing evidence of substantive GATT/WTO effects on trade, once one accounts for EIA dummies, multilateral resistance, and unobserved country-pair fixed effects (as we do here). This is the conclusion of Eicher and Henn (2011) (though they found a nontrivial WTO 'terms-of-trade' effect) and of Felbermayr and Kohler (2010) who examined possible extensive-margin effects; Eicher and Henn (2011) ignored extensive- versus intensive-margin effects. We also note an issue raised in Martin and Ng (2004), which is the role of multilateral tariff reductions under the GATT/WTO. Most-Favored-Nation (MFN) tariff cuts could also be affecting results. However, such MFN tariff cuts by country would be accounted for by the exporter-time and importer-time fixed effects.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

agreements (TWPTAs), FTAs, and a variable for customs unions, common markets, and economic unions (CUCMECU)—and they decomposed trade flows into extensive and intensive margins using the HK methodology.[12] While two recent studies have adapted the Baier–Bergstrand methodology for estimating the effect of differing 'types' of EIAs on bilateral aggregate trade flows, no econometric studies had examined the effect of various types of EIAs on the (goods) extensive and intensive margins of trade using a large number of country pairs and EIAs.[13] Neither HMR nor Egger et al. (2011) distinguished among various types of EIAs in their analyses of country intensive and extensive margins; also, Foster et al. (2011) used only a single EIA dummy. Baier et al. (2011) find not only that deeper EIAs have larger trade effects than FTAs, and the latter have larger effects than (partial) two-way and one-way PTAs, but they distinguished between these various trade effects at the extensive and intensive margins using a panel of (disaggregate) bilateral trade flows from 1962 to 2000 covering 98% of world exports.

Third, Bernard et al. (2009) is likely the only empirical study to date to explore the 'timing' of extensive and intensive margin responses to shocks. Using cross-sectional variation to examine long-run aspects, Bernard et al. (2009) find that variation in trade flows across country pairs is explained largely by the extensive margin, using firm-level data (the 'firm' margin); this result is consistent with HK using their 'goods' margin. But using *time-series* variation, Bernard et al. (2009) find that a larger proportion of trade variation can be explained by the intensive margin at short (5-year) time intervals. They show that, following the Asian financial crisis of 1997, virtually all of the variation in trade flows within 2–3 years could be explained by the *intensive* margin. This finding is consistent with two recent theoretical studies arguing that the low trade-cost elasticity found in macroeconomic analyses of business cycles should be associated with the intensive margin of trade compared with the relatively higher trade-cost elasticity found in international trade, which reflects extensive margin effects.[14] Allowing for differential 'timing' of EIA effects using panel data, Baier et al. (2011) find the first comprehensive empirical

12 Owing to few observations on common markets and economic unions, they combined these two types of 'deeper' EIAs with customs unions to form the variable CUCMECU, representing 'deep' EIAs.

13 The two studies that extended the Baier–Bergstrand framework to differing types of EIAs are Magee (2008) and Roy (2010); both found that customs unions had larger trade effects than FTAs. However, neither study examined extensive versus intensive margin issues.

14 Ruhl (2008) explains the delayed effect of the extensive-margin effects to fixed export costs on the supply side, while Arkolakis et al. (2011) explain the delayed effect of the extensive-margin effects to "Calvo-pricing" by consumers on the demand side.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

evidence that the shorter-term effects of EIAs on trade flows are more at the (goods) intensive margin and longer-term effects are more at the extensive margin (the latter entailing either fixed export costs or staggered 'Calvo' pricing by consumers), consistent with recent theoretical studies and empirical results in Bernard et al. (2009).

While the articles discussed in this section have addressed how to better estimate with consistency and precision EIA coefficient estimates—such as $\psi(1-\sigma)$ in equation (10)—in the presence of endogeneity, they have not addressed how to identify σ specifically. However, future work may want to pursue a combination of the issues raised in sections 2 and 3 to better identify consistently and precisely σ in the presence of imperfect endogenous trade-policy measures.

4 Conclusion

Despite widespread anecdotal evidence that lower trade barriers increase international trade, there is little firm quantitative evidence of the 'trade-cost elasticity' of trade flows, one of two key aggregate statistics that have recently been identified as sufficient to quantify the economic welfare effects of trade-policy liberalizations and/or trade-cost reductions (the other statistic being the import-penetration ratio). In other words, most estimates of the trade-cost elasticity are imprecise and lack consistency. In this article, we discussed two issues that are critical in better assessing empirically the trade-flow and welfare effects of trade liberalizations (or trade-cost changes). The first issue was how to quantify the trade-cost elasticity when trade costs themselves are approximated imperfectly. Various articles have suggested methods to better estimate with consistency and precision the trade-cost elasticity when the RHS variables can be treated as 'exogenous'. The second issue was that typical empirical evaluations to estimate the impact of trade-policy liberalizations on trade flows use the 'gravity equation'. However, the self-selection of country pairs into such agreements introduces endogeneity bias in the estimation of the trade-cost elasticity in gravity equations, requiring better identification techniques. Various articles have suggested methods to better identify the effects of EIAs on trade flows, adjusting for self-selection bias. Future work may want to pursue a combination of these two methodological issues to better identify consistently and precisely σ in the presence of imperfect endogenous trade-policy measures.

Acknowledgments

I am grateful to Volker Nitsch and Nikolaus Wolf for the invitation to write this article and to CESIfo for their financial support. Much of the

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

work cited in here is related to research with Scott Baier, Peter Egger, Michael Feng, and Mario Larch, and I am grateful for their collaborations.

References

Anderson, J. E. (1979), "A Theoretical Foundation for the Gravity Equation", *American Economic Review* **69**, 106–116.

Anderson, J. E. and E. van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", *American Economic Review* **93**, 170–192.

Arkolakis, C., A. Costinot and A. Rodriguez-Clare (2012), "New Trade Models, Same Old Gains?", *American Economic Review* **102**, 94–130.

Arkolakis, C., J. Eaton and S. Kortum (2011), "Staggered Adjustment and Trade Dynamics", Working paper.

Baier, S. L. and J. H. Bergstrand (2001), "The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity", *Journal of International Economics* **53**, 1–27.

Baier, S. L. and J. H. Bergstrand (2002), "On the Endogeneity of International Trade Flows and Free Trade Agreements, Manuscript (August)", University of Notre Dame.

Baier, S. L. and J. H. Bergstrand (2004a), "Economic Determinants of Free Trade Agreements", *Journal of International Economics* **64**, 29–63.

Baier, S. L. and J. H. Bergstrand (2004b), "Do Free Trade Agreements Actually Increase Members' International Trade? Manuscript (October)", University of Notre Dame.

Baier, S. L. and J. H. Bergstrand (2007), "Do Free Trade Agreements Actually Increase Members' International Trade?", *Journal of International Economics* **71**, 72–95.

Baier, S. L., J. H. Bergstrand and M. Feng (2011), "Economic Integration Agreements and the Margins of International Trade", Working paper.

Bergstrand, J. H. (1985), "The Gravity Equation in International Trade: Some Micro- Economic Foundations and Empirical Evidence", *Review of Economics and Statistics* **67**, 474–481.

Bergstrand, J. H., P. Egger and M. Larch (2013), "Gravity Redux: Estimation of Gravity-Equation Coefficients, Elasticities of Substitution, and General Equilibrium Comparative Statics Under Asymmetric Bilateral Trade Costs", *Journal of International Economics* **89**, 110–121.

Bernard, A. B., J. B. Jensen, S. J. Redding and P. K. Schott (2009), "The Margins of U.S. Trade", *American Economic Review* **99**, 487–493.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Chaney, T. (2008), “Distorted Gravity: The Intensive and Extensive Margins of International Trade”, *American Economic Review* **98**, 1707–1721.

Eaton, J. and S. Kortum (2002), “Technology, Geography, and Trade”, *Econometrica* **70**, 1741–1779.

Egger, P., M. Larch, K. E. Staub and R. Winkelmann (2011), “The Trade Effects of Endogenous Preferential Trade Agreements”, *American Economic Journal: Economic Policy* **3**, 113–143.

Eicher, T. and C. Henn (2011), “In Search of WTO Trade Effects: Preferential Trade Agreements Promote Trade Strongly, but Unevenly”, *Journal of International Economics* **83**, 137–153.

Feenstra, R. (2004), *Advanced International Trade: Theory and Evidence*, Princeton University Press, Princeton, NJ.

Feenstra, R. C. and H. L. Kee (2007), “Trade Liberalization and Export Variety: A Comparison of Mexico and China”, *The World Economy* **30**, 5–21.

Feenstra, R. C., R. E. Lipsey, H. Deng, A. C. Ma and H. Mo (2005), “World Trade Flows: 1962-2000, National Bureau of Economic Research”, Working Paper No. 11040, January.

Felbermayr, G. and W. Kohler (2010), “Modelling the Extensive Margin of World Trade: New Evidence on GATT and WTO Membership”, *The World Economy* **33**, 1430–1469.

Foster, N., J. Poeschl and R. Stehrer (2011), “The Impact of Preferential Trade Agreements on the Margins of International Trade”, *Economic Systems* **35**, 84–97.

Friedman, T. (2005), *The World is Flat*, Farrar, Straus, and Giroux, New York.

Gilpin, R. (2000), *The Challenge of Global Capitalism*, Princeton University Press, Princeton, NJ.

Helpman, E., M. Melitz and Y. Rubinstein (2008), “Estimating Trade Flows: Trading Partners and Trading Volumes”, *Quarterly Journal of Economics* **123**, 441–487.

Hillberry, R. and C. McDaniel (2002), “A Decomposition of North American Trade Growth since NAFTA, United States International Trade Commission”, Working Paper 2002-12-A, July.

Hummels, D. and P. J. Klenow (2005), “The Variety and Quality of a Nation’s Exports”, *American Economic Review* **95**, 704–723.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Kehoe, T. J. and K. J. Ruhl (2009), "How Important is The New Goods Margin in International Trade?", Federal Reserve Bank of Minneapolis Staff Report 324 (June).

Krugman, P. R. (1980), "Scale Economies, Product Differentiation, and the Pattern of Trade", *American Economic Review* **70**, 950–959.

Lawrence, R. Z. (1996), *Regionalism, Multilateralism, and Deeper Integration*, Brookings Institution, Washington, DC.

Lee, J.-W. and P. Swagel (1997), "Trade Barriers and Trade Flows Across Countries and Industries", *Review of Economics and Statistics* **79**, 372–382.

Magee, C. S. P. (2008), "New Measures of Trade Creation and Trade Diversion", *Journal of International Economics* **75**, 349–362.

Mansfield, E. D. and E. Reinhardt (2003), "Multilateral Determinants of Regionalism: The Effects of GATT/WTO the Formation of Preferential Trading Arrangements", *International Organization* **57**, 829–862.

Martin, W. and F. Ng (2004), *A Note on Sources of Tariff Reductions in Developing Countries, 1983–2003*, World Bank, Washington, DC.

McCallum, J. (1995), "National Borders Matter: Canada-US Regional Trade Patterns", *American Economic Review* **85**, 615–623.

Melitz, M. J. (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Profitability", *Econometrica* **71**, 1695–1725.

Preeg, E. H. (1998), *From Here to Free Trade*, University of Chicago Press, Chicago, IL.

Redding, S. J. (2011), "Theories of Heterogeneous Firms and Trade", *Annual Review of Economics* **3**, 77–105.

Roy, J. (2010), "Do Customs Union Members Engage in More Bilateral Trade than Free-Trade Agreement Members?", *Review of International Economics* **18**, 663–681.

Ruhl, K. (2008), "The International Elasticity Puzzle", Working paper.

Trefler, D. (1993), "Trade Liberalization and the Theory of Endogenous Protection: An Econometric Study of U.S. Import Policy", *Journal of Political Economy* **101**, 138–160.

Wooldridge, J. M. (2000), *Introductory Econometrics*, South-Western Publishing, Chicago, IL.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross-Section and Panel Data*, MIT Press, Cambridge, MA.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

CESifo Economic Studies, Vol. 59, 2/2013, 223–248 doi:10.1093/cesifo/ifs028
Advance Access publication 11 August 2012

Exporters and Their Products: A Collection of Empirical Regularities

Costas Arkolakis* and Marc-Andreas Muendler†

*Yale University, CESifo and NBER. e-mail: costas.arkolakis@yale.edu
†UC San Diego, CAGE, CESifo and NBER. e-mail: muendler@ucsd.edu

Abstract

We present a set of empirical regularities that characterize the export activity of firms. We decompose firm-level exports by product category across destination markets in a consistent manner for four data sets from Brazil, Chile, Denmark, and Norway. We relate the empirical regularities to new trade theories that connect microeconomic activity to aggregate outcomes. Our findings corroborate main motivating facts and may help discipline future theoretical work. (JEL codes: F12, L11, F14).

Keywords: International trade, heterogeneous firms, multi-product firms, firm and product panel data, Brazil, Chile, Denmark, Norway

1 Introduction

The recent surge of empirical research in international trade that uses firm-level data has opened new avenues for theory but also raises challenges. Micro data on exporters, their products, and their destinations offer a number of rich statistics that are useful in disciplining models of international trade and in sharpening our information on the costs of export market access. An important empirical concern with micro-level statistics is their robustness under alternative levels of disaggregation and across countries at different stages of development. The purpose of this article is to establish key features of trade data that are robust across developing and industrialized countries and across levels of aggregation.

We collect a set of empirical regularities that characterize the export activity of firms and their products across foreign destinations. To establish robustness, we apply the same statistical methodology to data from a group of four export countries—Brazil, Chile, Denmark, and Norway—with comprehensive data on export participation, destination markets, and export products among manufacturing firms. We conduct our statistical analysis at varying levels of product aggregation, and compare statistics to earlier findings for France and the USA. Given the success of recent trade theories in explaining export activity at the firm-level, we use our cross-country statistics to validate main insights from international trade models, but also to motivate and potentially discipline future theoretical work.

© The Authors 2012. Published by Oxford University Press
on behalf of Ifo Institute, Munich. All rights reserved.
For permissions, please email: journals.permissions@oup.com

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Our focus lies on patterns of entry and sales at individual export destinations across firms and firm–products from different source countries. For each export country we use three-dimensional data for firms, their destination markets, and their export products. By design, such three-dimensional data cover two extensive margins of export activity, and one remaining intensive margin. The first extensive margin is that of firm entry into a foreign market with the firm's first export product at the destination. The second extensive margin is that of product entry by the same firm at a given foreign market with additional products beyond the first exported good. Related to this second extensive margin of export activity, we call the number of products that an exporter ships to a destination the firm's *exporter scope* at the destination. The remaining intensive margin covers the individual sales per product at the destination.

We use firm–product–destination data, motivated by the widely documented regularity that multi-product firms dominate export–market participation. Bernard et al. (2009) show for US trade data in the year 2000, for instance, that firms that export more than five products at the Harmonized System (HS) 10-digit level make up 30% of exporting firms but account for 97% of all exports. Iacovone and Javorcik (2008) and Arkolakis and Muendler (2010) document related concentration patterns for Mexico and Brazil, for example.

We present two sets of basic statistics: statistics related to the entry and sales of firms, and statistics related to the entry and sales of products by firm. For each of the two sets we document concentration patterns and their relation to country characteristics. We consider these basic statistics as benchmark regularities that any successful model of trade and market structure might want to confront.

Entry and sales statistics at the firm level suggest that exporting is strongly fragmented by national markets. Only a fraction of firms overcomes the barriers to export–market access. Our analysis of these statistics is similar to Eaton et al. (2004) but our contribution is to establish that these regularities persist across source countries with different characteristics. The evidence is consistent with the idea that fixed entry costs per firm as well as per-unit shipping costs keep national markets separate, and it supports conventional assumptions in recent models of international trade.

Entry and sales statistics at the firm-product level show for a destination market such as the USA that there are only a few exporters with wide exporter scope and large sales, but there are many narrow-scope and small-sales firms. This evidence is consistent with the idea that there are also fixed entry costs to a firm's expansion of its product scope in addition to per-unit shipping costs that separate markets. Our analysis of these statistics is similar to Arkolakis and Muendler (2010) and this article

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

establishes that the regularities persist across source countries with different characteristics. To match the regularities, models need to explain the high frequency of exporters that have small sales and ship only a few products as well as the simultaneous dominance of a few wide-scope and large-sales firms in total exports.

For the extensive margin of product entry within firms, gravity-type regressions similar to Bernard et al. (2011) suggest that the average exporter scope across firms in a market is not significantly related to destination–market size, as measured by GDP, but exporter scope is related to distance. The reverse is the case for the remaining intensive margin of sales per firm–product. Sales per firm-product are unrelated to distance but significantly related to destination–market size as measured by GDP. The evidence is consistent with the idea that firms face repeated and similar market-entry costs for their products destination by destination so that average exporter scope is not responsive to local market size.

This article has five more sections. Section 2 presents our data sources for Brazil, Chile, Denmark, and Norway, and our data preparation. Section 3 reports statistics on export–market presence by source country and destination characteristics. Section 4 explores the distributions of exports and exporter scope from our four source countries in a leading export market, the USA. Section 5 relates the exporter scope of a source country's firms, and the complementary margins of bilateral exporting, to destination–market characteristics. We offer concluding remarks in Section 6.

2 Data

We apply consistent methods to the preparation of exporter–product–destination data for Brazil, Chile, Denmark, and Norway, and to the computation of statistics. We also compare our evidence to published statistics for France and the USA.[1] In product space, we restrict the sample to manufactured products. On the firm side, we restrict the sample to manufacturing firms and their direct exports of manufactures. This restriction makes our findings closely comparable to statistics previously published by Eaton et al. (2004) on France and Bernard et al. (2011) on the USA, for example.

The Brazilian exporter data derive from the universe of customs declarations for merchandize exports during the year 2000 at SECEX (*Secretaria de Comércio Exterior*). Transactions of any value and weight are included

[1] Results for exporter–destination–product data from Greece (courtesy of Dinopoulos et al. 2012) also exhibit closely related patterns to the ones presented in this article.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

in these declarations data. From these customs records, we construct a three-dimensional data set of Brazilian manufacturing exporters, their destination countries, and their export products at the HS 6-digit level. At this disaggregation level, customs codes are identical across countries. In the raw exports data from SECEX, product codes are 8-digit numbers under the common Mercosur nomenclature (NCM), of which the first 6 digits coincide with the first 6 HS digits. We aggregate the original monthly exports data to the HS 6-digit product, firm, and year level for most of our data work, but stay at the NCM 8-digit level for comparisons to the US evidence in gravity equations. We use the formal-sector employer–employee records RAIS (Relação Anual de Informações Sociais from the Brazilian labor ministry) to link the manufacturing exporter data to the universe of Brazilian manufacturing firms for the total firm count.

A similar three-dimensional data set of Chilean exporters derives from the universe of annual customs declarations by Chilean manufacturing firms in 2000. As is the case for the Brazilian data, transactions of any value and weight are included in the Chilean declarations data. Roberto Álvarez kindly shared the Chilean data for 2000 (for a data description see Álvarez et al. 2007). We aggregate the annual data from the HS 8-digit to the HS 6-digit level for cross-country comparison in most of our data work but, similar to Brazil, we stay at the HS 8-digit level for a US comparison in gravity estimation. For Chile, we do not have the total manufacturing firm count in the data. As an estimate, we use export participation among Chilean plants in 2000 from reported statistics in Bergoeing et al. (2011, Table 5).

Evidence on the Danish exports is courtesy of Ina C. Jäkel (Jäkel 2012). The Danish data derive from the Globid data base at the Department of Economics and Business, Aarhus University, and Statistics Denmark. The reporting thresholds are 3000 Danish Krones (approximately 400 US dollars in 2000) and a weight of one ton per monthly transaction total for shipments to destinations inside the European Union, and 7500 Danish Krones (approximately 1000 US dollars in 2000) and a weight of one ton per monthly transaction total for shipments to destinations outside the European Union. Below the threshold, firms may voluntarily report. The final three-dimensional data set of Danish exporters, their respective destination countries, and their export products is at the HS 6-digit level after mapping the Danish product codes to the HS 6-digit level. The Danish data include both domestic and foreign activity of Danish manufacturing firms so that no additional data treatments are required.

Evidence on the Norwegian exporters is courtesy of Andreas Moxnes (Irarrazabal et al. 2010). The Norwegian data are based on customs

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

declarations for an exhaustive sample of Norwegian non-oil exporters in 2000, which are then further restricted to manufacturing firms (NACE sectors 15 through 37). The reporting threshold for inclusion of an exporter in the data is 1000 Norwegian Krones (approximately 125 US dollars in 2000) for the total annual transactions value per firm. The resulting three-dimensional data set of Norwegian exporters, their respective destination countries, and their export products is also at the HS 6-digit level. For further details on the customs data, see Irarrazabal et al. (2010) and (2011). To obtain data on Norwegian manufacturing firms including non-exporters, Moxnes has linked the Norwegian customs data with Norwegian manufacturing firm data. The combined data set excludes a small number of manufacturing exporters from the customs data but statistics on exports to the USA are largely unaffected.

For Brazil and Chile, we also present a set of additional statistics with product–market information by destination country and sector. For this purpose, we map the HS 6-digit codes to ISIC revision 2 at the 2-digit level and link our data to World Trade Flow (WTF) data for the year 2000 (Feenstra et al. 2005) and to *Unido* Industrial Statistics (UNIDO 2005).[2] In gravity regressions, we use CEPII bilateral geographic distance data (the mean distance between Brasília or Santiago de Chile on the one hand and foreign capital cities in Kilometers on the other hand) and the IMF's International Financial Statistics for GDP (in current US$).

3 Export Market Presence

We start our investigation with an assessment of the destinations that exporting firms reach, and the characteristics of the destinations that attract many exporters.

3.1 Frequency of export market presence

We first take the perspective of the exporting country and its firms. For each of our four source countries we plot the number of firms against the number of destinations to which these firms ship in 2000. The destination count includes the home country so that nonexporters appear as having one market. Figure 1 depicts the plots in log–log graphs, replicating Eaton et al. (2004). The number of firms that reach a given number of destinations declines relatively smoothly and monotonically, from a large number of firms that serve only a single market (their home market) to the point where a handful of firms serves a large number of markets.

[2] Our extended SITC-to-ISIC concordance is available at econ.ucsd.edu/muendler/resource.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

To measure the decline in destination reach, we fit a linear regression line to the graphs in Figure 1 by regressing the log number of firms with a given number of destinations on their log number of destinations. For France, Eaton et al. (2004) report the coefficient estimate of this elasticity to be −2.5.[3] Our estimated elasticities are −2.48 for Brazil (standard error 0.065) and −2.35 (0.079) for Chile. In contrast, the economies of Denmark and Norway exhibit less pronounced declines in destination reach, with elasticities of −1.98 (0.054) in Denmark and −1.94 (0.056) in Norway.[4]

The smaller the total number of manufacturing firms in a source country, the smoother the elasticity of the number of firms with respect to the number of markets. Norway exhibits the smallest dropoff between single-destination and two-destination firms, and has only 8688 manufacturing companies in 2000. Denmark has 20 470 manufacturing firms in 2000 and exhibits a somewhat more pronounced dropoff. (Our imputed total number of manufacturing firms for Chile in 2000 is 31 322.) Brazil, in contrast, hosts 697 259 manufacturing firms in 2000—multiple times the manufacturing firm counts even for France and the USA of 234 300 and 191 648 in 1986 and 1987 (Eaton et al. 2004). Only a relatively small fraction of the many Brazilian manufacturing firms exports, contributing to the strong dropoff in the number of Brazilian firms between the first and second destination market.

To summarize the evidence for all these countries (Brazil, Chile, Denmark, Norway, and France), the modal manufacturing firm is a nonexporter. The modal exporter ships to only one foreign destination. And only a small fraction of firms ships to a wide number of destinations. This evidence provides a sense of the difficulty of exporting. Only select firms

3 Our descriptive regression imposes a log-linear relationship between the firm count and the destination count for simplicity. Chaney (2011), in contrast, emphasizes a detectably concave curvature in the relationship.

4 The largely smooth relationship between the number of firms and the number of destinations that they reach exhibits a stark dropoff between one and two destination markets, however. This dropoff is especially pronounced in Brazil and Denmark but is also clearly observable in Chile and Norway. For Brazil, Chile, and Norway, this dropoff is entirely driven by the transition from nonexporters to exporters because we lack information on home-market sales for those countries. For Denmark, the count of firms with sales to a single market could in principle also reflect exporters with a single export market but no domestic Danish sales. In 2000 in Denmark, all exporters in our sample also have Danish domestic sales. One concern with this stark dropoff in the number of firms between the first and second destination market is that the above reported regression coefficients are upward biased compared to estimates for only export destinations (destination counts of two or more). In none of our sample countries, however, the coefficient changes strongly. In Brazil, the country with the strongest dropoff, the elasticity of the number of firms with respect to the number of markets changes from −2.48 including single-destination firms to −2.33 excluding single-destination firms. In Denmark, the country with the second strongest dropoff, the elasticity changes from −1.98 just to −1.96.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

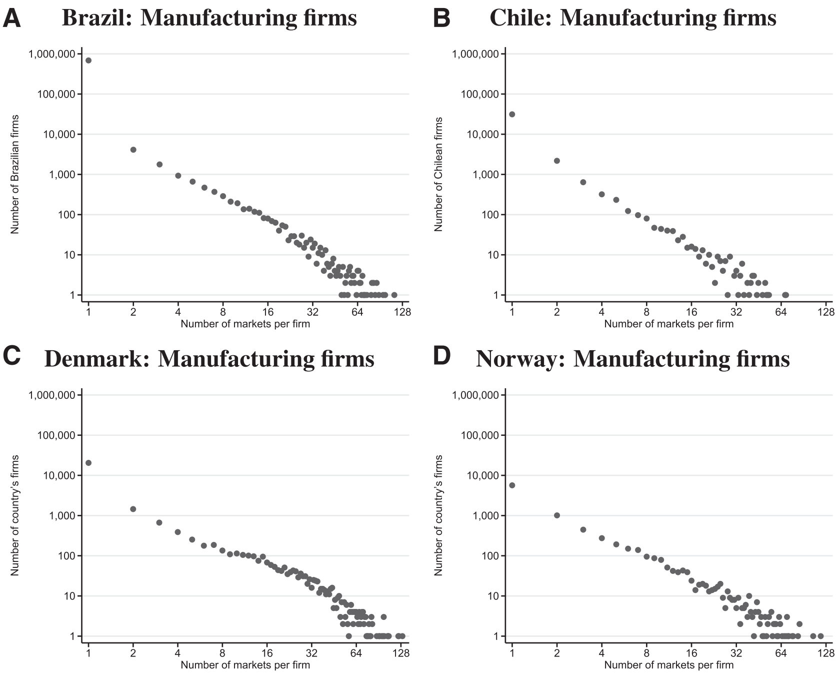

Figure 1 Export market presence. *Sources*: Brazilian *SECEX* 2000, Chilean customs data 2000 (Álvarez et al. 2007), Danish Globid data 2000 (see Jäkel 2012), Norwegian combined customs and manufacturing firm data 2000 (compare to 2004 data by Irarrazabal et al. 2010); manufacturing firms and their manufactured products. *Note*: Graphs for Brazil, Chile, and Norway under the assumption that every manufacturer has sales in the domestic market. For Chile, nonexporters imputed from nonexporting Chilean plants in 2000 (Bergoeing et al. 2011, Table 5).

are able to overcome obstacles to exporting, reflected in a substantive elasticity with which the number of firms declines as additional export destinations are reached.

Trade models can generate selection of firms into exporting based on variable trade costs alone (Bernard et al. 2003; Melitz and Ottaviano 2008). However, models where variable trade costs are the only barrier to trade are typically not able to generate both the observed relative size of exporters, compared to nonexporters, and the strong selection into exporting (Bernard et al. 2003). To come to terms with both these regularities, international trade models typically require additional exporting costs either in the form of fixed costs (Roberts and Tybout 1997; Melitz 2003; Chaney 2008) or in the form of increasing marketing costs to penetrate foreign markets (Arkolakis 2010; Eaton et al. 2011).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

3.2 Destination market size

We now relate export market entry to characteristics of the destination country. Among the potentially relevant destination country attributes is market size and the attraction that market size exerts on foreign firms and firm–products. For this exercise, we measure a destination country d's market size X_d as its absorption, defined as gross manufacturing production plus imports minus exports (in billions of US dollars).[5]

Firm entry and market size: a common framework to interpret bilateral trade volumes is the gravity equation, which relates exports T_{sd} between a source country s and a destination d to the market sizes of s and d and geographic distance d_{sd} between the two countries:

$$T_{sd} = \kappa\, X_s X_d / d_{sd}$$

for some constant κ. Following the approach in Eaton et al. (2004) for each of our source countries s, we define a source country's market share in a destination country's absorption simply as $\lambda_{sd} \equiv T_{sd}/X_d$ so that the exports T_{sd} from s to d can be understood as

$$T_{sd} = \lambda_{sd} X_d.$$

The market share λ_{sd} is commonly thought to be partly driven by the distance between s and d.

Using our firm-level data, we can also decompose total exports T_{sd} from source country s to destination d into

$$T_{sd} = M_{sd}\, \bar{t}_{sd}, \tag{1}$$

where M_{sd} is the number of exporters in s with shipments to d, and $\bar{t}_{sd} \equiv T_{sd}/M_{sd}$ are these exporter's mean sales to d (see e.g. Eaton et al. 2004). This decomposition accounts for the (first) extensive margin of market presence by firms. The remaining intensive margin of average export sales per firm subsumes both the (second) extensive margin of product entry and the sales per firm–product into a broad intensive-margin term. We will turn to product entry in the next two sections. For now, we combine the definition of market share λ_{sd} with decomposition (1).

The left panel of Figure 2 depicts the relationship among three of the four elements in the definition and the decomposition: $T_{sd} = \lambda_{sd} X_d = M_{sd}\, \bar{t}_{sd}$. On the horizontal axis is the market size measure

[5] Gross manufacturing production is from *Unido* Industrial Statistics (UNIDO 2005), and exports and imports are from World Trade Flow (WTF) data for the year 2000 (Feenstra et al. 2005). For Brazil's exporters in 2000, we cover 171 destination countries with this absorption measure, for Chile in 2000 we cover 140 countries.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

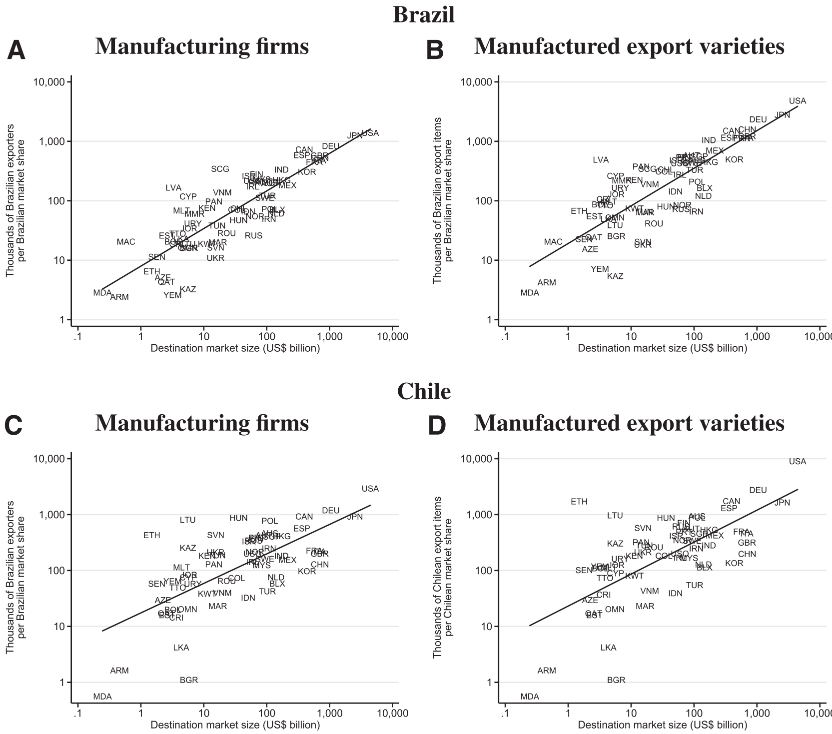

Figure 2 Market size and exporter presence. *Sources*: Brazilian *SECEX* 2000 (A and B), Chilean customs data 2000 (C and D), manufacturing firms (A and C), and manufactured firm-products (B and D) at the HS 6-digit level, linked to *WTF* (Feenstra et al. 2005) and *Unido* Industrial Statistics (UNIDO 2005). *Note*: Market size is absorption by a country's manufacturing sector. Each manufacturing firm's export product is one variety. The slopes of the fitted lines are 0.632 (standard error 0.049) for manufacturing firms from Brazil (A), 0.632 (0.053) for manufactured firm-products from Brazil (B), 0.527 (0.075) for manufacturing firms from Chile (C), and 0.571 (0.075) for manufactured firm-products from Chile (D).

X_d. On the vertical axis is the number of source country s's exporters divided by the source country's market share at d: M_{sd}/λ_{sd}. This division is meant to partly control for the effect of distance between s and d. When normalized by market share λ_{sd}, the number of Brazilian firms (Figure 2A) and Chilean firms (Figure 2C) selling to a destination increases systematically with market size, but with an elasticity less than unity.

Both Figures 1 and 2 illustrate the rarity of prolific exporters. Figure 1 documented that a few firms reach many markets, and Figure 2 shows that

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

only large markets sustain many exporters from a given source country. Taken together, this evidence is consistent with the explanation that only a few firms reach small markets whereas most firms ship to a limited number of large markets. These patterns indicate that exporters face substantial entry costs and support the emphasis on the importance of extensive margins in explaining overall trade. The robustness across countries is consistent with the idea that the nature of entry costs (though not necessarily their levels) may be similar across countries.

Several explanations are consistent with the robust positive association between destination market size and exporter presence from any source country. As a market's size increases, it becomes more likely that firms from any source country can expect earnings that exceed the entry costs of accessing the destination market (see e.g. Arkolakis 2010; Eaton et al. 2011). Another consistent explanation is that richer countries demand a broader set of vertically differentiated varieties under non–homothetic demand, attracting more entrants with quality-differentiated products from any source country (see e.g. Fajgelbaum et al. 2009; Simonovska 2010). Yet another theory consistent with this evidence is that larger markets promote the formation of trading networks, which in turn facilitate the entry of firms from any source country (see e.g. Rauch 1999; Chaney 2011). Note, however, a flip side of the positive association between market size and market entry with an elasticity of less than unity is a positive association between the average size of exporters and market size with an elasticity of more than unity. Matching the latter association quantitatively essentially requires the explicit modeling of an entry cost as in Eaton et al. (2011).

Variety entry and market size: prior to the availability of individual firm–product data, much empirical research has considered export goods as classified by product category. Under this perspective, firms can be viewed as providing their brand, and the brand in turn provides the platform for specific products to be launched. A decomposition of total exports T_{sd} related to this view of the product space is:

$$T_{sd} = V_{sd}\,\bar{a}_{sd}, \tag{2}$$

where $V_{sd} \equiv \sum_{\omega\in\Omega_{sd}} G_d(\omega) = M_{sd}\,\bar{G}_{sd}$ is the number of branded products (or ''varieties'') shipped to d, and ω denotes the individual firm or brand within the set $\mathbf{\Omega}_{sd}$ of firms that ship from s to d. The average scale of the branded products is $\bar{a}_{sd} = [\sum_{\omega\in\Omega_{sd}} t_d(\omega)]/[\sum_{\omega\in\Omega_{sd}} G_d(\omega)] = \bar{t}_{sd}/\bar{G}_{sd}$ (similar to Broda and Weinstein 2006, identical under the convention that every source country is a single exporter $M_{sd} = 1$). For empirical implementation, we define a branded export variety as a manufacturing firm's export product at the HS 6-digit level.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Figure 2B and D depicts firm–product entry using the relationship among three of the four elements in the definition of exports and decomposition (2): $T_{sd} = \lambda_{sd} X_d = V_{sd} \bar{a}_{sd}$. On the horizontal axis is the market size measure X_d. On the vertical axis is the number of source country s's firm–products divided by the source country's market share at d: V_{sd}/λ_{sd}. When normalized by market share λ_{sd}, the number of Brazilian firm–products (Figure 2B) and Chilean firm-products (Figure 2D) selling to a destination increases systematically with market size, but with an elasticity less than unity.

Overall, elasticities of firm–product entry with respect to market size are similar to the earlier elasticities of just firm entry. For Brazil, the slopes of the regression lines in Figure 2 are statistically indistinguishable. For Chile, the elasticity of firm–product entry with respect to market size is somewhat larger than the elasticity of just firm entry. We will return to a discussion of this elasticity after investigating the distribution of exporter scope and the response of mean exporter scope to foreign market size in the following two sections.

A comparison between the right and left panels of Figure 2 suggests that there is a potentially separate role for within-firm product differentiation. Put differently, there appears to be a (second) extensive margin of product entry by the firm. The similarity of firm–product entry with firm entry suggests that the nature of firm–product entry costs is comparable to that of firm entry costs discussed earlier. Firm–product entry costs have been modeled by Bernard et al. (2011) and Arkolakis and Muendler (2010). Allowing for both increasing marginal cost by product as in Eckel and Neary (2010) and for local fixed entry costs that depend on the number of products, Arkolakis and Muendler (2010) analyze the quantitative nature of firm–product entry costs.

4 Sales and Product Distributions Across Firms

The exporter scope and sales decisions conditional on entry are the main points of interest in this and the next section. To look underneath the surface of firms' entry and sales decisions, in this section we analyze the size and product scope distributions across firms. We focus on a single large destination market to emphasize the evidence in the cross section of firms from a common source country. The destination country at the top right extreme of the graphs in Figure 2 is the USA, so we use the USA as the destination market for our data exploration in this section.[6]

[6] For Brazil and Chile as source countries we report sales and product distributions also for other destination markets beyond the USA in a comprehensive online Data Appendix to Arkolakis and Muendler (2010) at econ.ucsd.edu/muendler/papers/abs/braxpmkt.html.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

4.1 Sales distribution

We first investigate the variation of exports among exporters that ship to the USA. For each source country—Brazil, Chile, Denmark and Norway—we rank the exporters according to their total sales in the USA in 2000. For each percentile of the firm's total sales distribution, we then compute the total sales at that percentile and plot the sales against the percentile using a log scale on the vertical axis. Figure 3 displays the graphs. Especially in the more advanced countries Denmark and Norway, the total sales distribution exhibits an approximate power law behavior. The distributions deviate from power law behavior in the lower tail, however. Especially in Brazil and Chile, but also to some degree in Norway, sales at small firms decline more than proportionally with the percentile. In summary, there are a few large-sales firms but many small-sales firms.

For an explanation of deviant small-firm behavior in the lower tail see Arkolakis (2010). That paper explains the existence of many exporters with minor sales in the low tail by introducing increasing marketing costs, which firms incur when they reach additional consumers within a destination market. Additional heterogeneity of sales of firms can be attributed to random variation across markets as discussed in detail by Eaton et al. (2011).

The distributions are similarly concentrated in the high tails between source countries, and the plots for our four source countries look similar to the one for France reported in Eaton et al. (2011).

The robustness of the sales distribution across our source countries, but also its robustness across destinations for any given source country (Arkolakis and Muendler 2010), presents a regularity that theory needs to come to terms with. Trade volumes and the gains from trade depend on the concentration of the sales distribution. In frameworks with heterogeneous firms, imposing a Pareto distribution on productivity turns out to be a sufficient distributional assumption to generate stable sales distributions across source countries for a wide range of demand functions (see Arkolakis et al. 2012). The Pareto distribution is closed under truncation so that, conditional on entry, the productivity distribution remains Pareto. A CES demand system with symmetric elasticities can generate export sales with a Pareto distribution in the upper tail that is robust across source–destination country pairs (see e.g. Chaney 2008; Arkolakis and Muendler 2010; Eaton et al. 2011).

4.2 Exporter scope distribution

We now turn to a main new variable that can be computed from firm–product–destination data following Arkolakis and Muendler (2010): a firm's exporter scope at a given destination, which we define as the

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

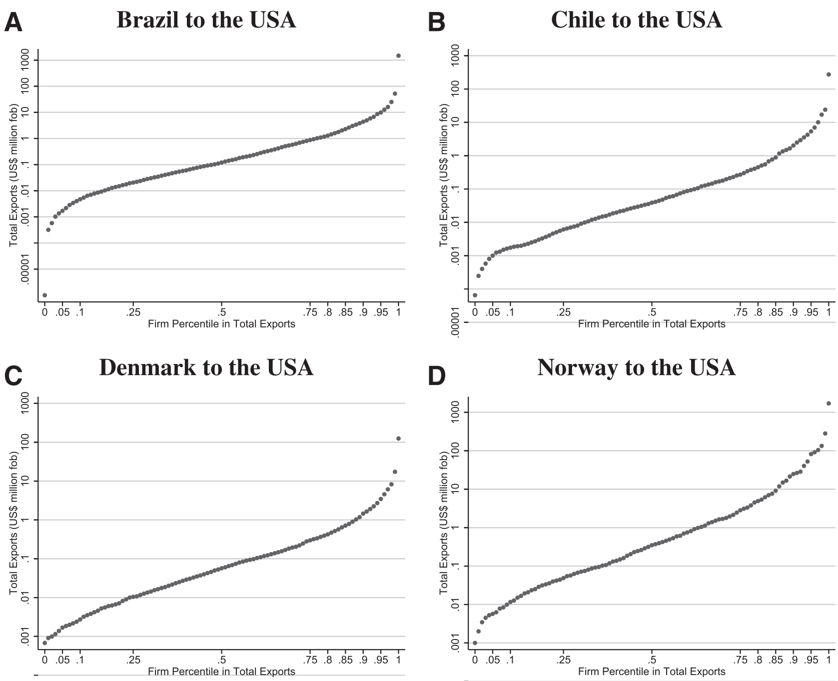

Figure 3 Total sales distributions in the USA. *Sources*: Brazilian *SECEX* 2000, Chilean customs data 2000, Danish Globid data 2000, Norwegian combined customs and manufacturing firm data 2000; manufacturing firms and their manufactured products.

number of products at the HS 6-digit level shipped by an exporter to a destination.[7] For each source country we now rank the country's exporters according to their exporter scope in the USA in 2000. For each percentile of the firm's exporter scope distribution, we then compute the exporter scope at that percentile and plot the scope against the percentile using a log scale on the vertical axis. The graphs are shown in Figure 4.

Exporter scope is a discrete variable but the overall shapes of the distributions approximately resemble those of power-law distributed variables. The median exporter from Brazil and Chile in 2000 ships just one product to the USA. The median exporter from Denmark and Norway, in contrast, ships two products to the US market. Even in the largest export market, the United States, the exporter scope of the typical (median) firm is just one or two products. Interestingly, the breadth of exporter scope is

[7] We turn to robustness checks with finer levels of product disaggregation in the following section and in a comprehensive online Data Appendix to Arkolakis and Muendler (2010) at econ.ucsd.edu/muendler/papers/abs/braxpmkt.html.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

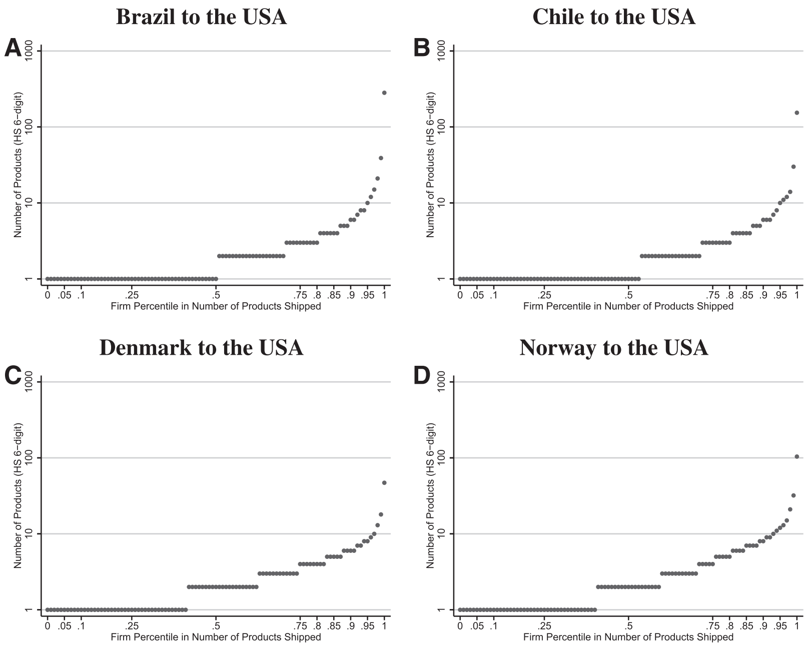

Figure 4 Exporter scope distributions in the USA. *Source*: Brazilian *SECEX* 2000, Chilean customs data 2000, Danish Globid data 2000, Norwegian combined customs and manufacturing firm data 2000; manufacturing firms and their manufactured products. Products at HS 6-digit level.

reversed between the two country groups, Brazil–Chile on the one hand and Denmark–Norway on the other hand, at the high end of exporter scope. The Brazilian manufacturing exporter with the widest exporter scope in the USA ships 273 products at the HS 6-digit level, and the widest-scope Chilean exporter sells more than 100 products. In contrast, the widest-scope Norwegian exporter to the USA ships just a little more than 100 products, and the top Danish exporter even fewer products.

In summary, there are only a few wide-scope and large-sales firms, but there are many narrow-scope and small-sales firms at a given destination. Models that strive to explain the role of multi-product exporters therefore need to explain the high frequency of exporters that have small sales and ship only a few products and the simultaneous dominance of a few wide-scope and large-sales firms in total exports. Conceptually, a combination of the models by Arkolakis (2010) and Arkolakis and Muendler (2010) could simultaneously generate these relationships: the least productive firms would pay a low marketing cost to reach only a few consumers at a destination but would also typically choose to sell only a few products to those consumers.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

5 Product Exports and Market Characteristics

We now look behind the distribution of exporter scope and investigate more closely product entry, its relation to the other two margins and its relation to destination–market characteristics. Beyond the (first) extensive margin of firm presence, in this section we decompose destination by destination an exporter's sales into the (second) extensive margin of the firm's number of products at a destination—the *exporter scope*—and the remaining intensive margin of the exporter's average sales per product at the destination, which we call *exporter scale*.

5.1 Export margins and gravity

As a start, we relate back to the common framework of the gravity equation to describe the multilateral export data. There is a natural extension of the two earlier decompositions (1) and (2) in Section 4 to the case of three jointly known export dimensions in firm, product, and destination data. Departing from decompositions (1) and (2), an extended margin decomposition can account for the newly observable extensive margin of firms' exporting products and also consider the average number of products per firm, or mean exporter scope:

$$T_{sd} = M_{sd}\,\bar{G}_{sd}\,\bar{a}_{sd}, \tag{3}$$

where $\bar{G}_{sd} \equiv \sum_{\omega\in\Omega_{sd}} G_d(\omega)/M_{sd}$ is the exporter's mean scope, and $\bar{a}_{sd} \equiv \bar{t}_{sd}/\bar{G}_{sd}$ is these exporter's mean scale. This decomposition generalizes both decompositions (1) and (2) and naturally accounts for the firm's average exporter scope $G_d(\omega)$.[8] Exporter scope is a central variable in recent theories of multi-product exporters, including Feenstra and Ma (2008), Eckel and Neary (2010), Bernard et al. (2011), Nocke and Yeaple (2006), Dhingra (2010), Mayer et al. (2011), Arkolakis and Muendler (2010).

While (3) is one natural generalization of the earlier decompositions from Section 4, it is not the only possible extension. Total exports T_{sd} can also be decomposed into:

$$T_{sd} = M_{sd}\,\hat{G}_{sd}\,\hat{a}_{sd}, \tag{4}$$

where $\hat{G}_{sd} \equiv \sum_{\omega\in\Omega_{sd}} G_d(\omega)$ now is the total number of products exported from s to d by any firm (the HS-6 digit categories filled by anyone), and $\hat{a}_{sd} \equiv \bar{t}_{sd}/\hat{G}_{sd}$ is the 'average value of exports per product per firm' (Bernard et al. 2007; p. 121). This decomposition

8 Note that $\bar{a}_{sd}$ is the weighted arithmetic mean of $a_d(\omega)$ over all firms ω, with weights $G_d(\omega)$: $\bar{a}_{sd} = \sum_{\omega\in\Omega_{sd}} G_d(\omega)\,a_d(\omega)/(\sum_{\omega\in\Omega_{sd}} G_d(\omega)) = \bar{t}_{sd}/\bar{G}_{sd}$.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

generalizes decomposition 1 but does not naturally generalize decomposition 2 because $\hat{a}_{sd} \equiv (\bar{G}_{sd}/\hat{G}_{sd})\,\bar{a}_{sd}$. Moreover, the total number of products $\hat{G}_{sd}$ exported by any firm from s to d is not directly related to the (second) extensive margin of average product entry within firms.

So as to accommodate both possible extensions (3) and (4), Bernard et al. (2011) propose an all-encompassing quadruple decomposition of total exports T_{sd} into

$$T_{sd} = M_{sd}\,\hat{G}_{sd}\,\mu_{sd}\,\bar{a}_{sd}, \tag{5}$$

where $\hat{G}_{sd} \equiv \sum_{\omega\in\Omega_{sd}} G_d(\omega)$ is the total number of products exported from s to d, μ_{sd} is the share of firm–product combinations with positive product exports, which Bernard et al. (2011) call the 'density of trade', and $\bar{a}_{sd} \equiv \bar{t}_{sd}/\bar{G}_{sd}$ is the average exporter scale of the firm-products at the destination. This quadruple decomposition can be transformed back into our triple decomposition (3) by setting $\bar{G}_{sd} = \hat{G}_{sd}\,\mu_{sd}$ (as in Arkolakis and Muendler 2010). Once transformed back, the number of exporters M_{sd} reflects the (first) extensive margin of firm entry, the average exporter scope $\bar{G}_{sd} = \hat{G}_{sd}\,\mu_{sd}$ reflects the (second) extensive margin of product entry by a firm at the destination, and $\bar{a}_{sd}$ covers the remaining intensive margin of the exporter's mean exporter scale. Alternatively, one can set $\hat{a}_{sd} = \mu_{sd}\,\bar{a}_{sd}$ and get back to the triple decomposition by Bernard et al. (2007).

Table 1 presents the results from relating the quadruple margin decomposition of (5) to two foremost gravity equation variables: market size at the destination and distance between source and destination country. For the USA, Bernard et al. (2011) present gravity evidence for the quadruple decomposition using GDP and distance between capital cities, and we follow their specification for comparability.[9] Table 1 reports the coefficients from an OLS regression of the log of each of the four variables in (5) on both log GDP and log distance. By the properties of OLS, we can add all four coefficients in a row to arrive at the total gravity coefficient from a regression of T_{sd} on log GDP and log distance. Rounding error aside, the sum of the coefficients in Columns (2)–(5) is equal to the coefficient in Column (1). We can also add pairs of coefficients in the quadruple regression behind Table 1 to recover results for the alternative triple

[9] Log GDP differs from the absorption-based market size measure of Section 3 in two main regards: absorption is based on gross manufacturing production (for a derivation see Eaton et al. 2011), whereas GDP is a value-added measure for the whole economy, and absorption corrects for trade imbalances.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

decompositions (3) and (4). Importantly, the sum of the coefficients in Columns (3) and (4) is the estimate for the contribution of the (second) extensive margin of average product entry by the firms at a destination.

Several striking facts emerge from a comparison of the USA to Brazil and Chile in Table 1. Results for the USA in Bernard et al. (2011) are reported at the HS 10-digit level only (a unique level of disaggregation). For Brazil and Chile, we report results both at the finest possible level for these two countries (NCM-8 for Brazil and HS-8 for Chile), and for the HS 6-digit product classification. The HS is identical across countries around the world at the 6-digit level. First and perhaps most importantly, signs of all coefficients are identical between all three countries. This is true both for the fine product classification at the 8-digit level and for the HS 6-digit product classification for Brazil and Chile, which makes results most widely comparable across countries. Moreover, the pattern of statistical significance is identical between all three countries and both levels of aggregation, with only the distance coefficient in the mean exporter scale regression (column (5)) lacking statistical significance at the 1% level.

The magnitude of coefficients is quite similar across countries and levels of product classification, too. The coefficients on log GDP are so close that their equality cannot be rejected between the USA and Brazil for total exports T_{sd}. Neither can the equality of coefficients on the total number of products $\hat{G}_{sd}$ be rejected between the USA on the one hand and Brazil at the HS 6-digit level or Chile at either level of aggregation on the other hand. As emphasized before, this striking robustness across countries is consistent with the idea that the nature of entry costs (though not necessarily their levels) may be similar across countries. For the regressor log distance, however, magnitudes of coefficients vary more strongly across countries. A reason is perhaps that other gravity-related measures of trade barriers—such as language, lacking contiguity, customs-related trade costs, and other policy barriers—covary in important ways with distance but in different ways for different source countries (Anderson and Van Wincoop 2004).

The decomposition of bilateral trade flow components in Table 1 allows us to focus on the (second) extensive margin of product entry by firms more closely. The sum of coefficients in Columns (3) and (4) returns the gravity coefficients for the (second) extensive margin of product entry by firm ($\ln \bar{G}_{sd} = \ln \hat{G}_{sd} + \ln \mu_{sd}$). For both log GDP and log distance, the coefficients in columns (3) and (4) have the same sign across all three countries. The coefficients in Columns (3) and (4) also have the same sign for any level of product classification in Brazil and Chile, respectively. Most strikingly, for the log GDP regressor the total coefficient sum across all columns is close to zero in all five specifications. Under a triple decomposition following (3), the regression of $\bar{G}_{sd}$ on log GDP and log distance

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 1 Gravity and the quadruple exports decomposition

	Log Total exports $\ln T_{sd}$ (1)	Log # Firms $\ln M_{sd}$ (2)	Log # Total products $\ln \hat{G}_{sd}$ (3)	Log share Pos. Prod. exp. $\ln \mu_{sd}$ (4)	Log sales/# prod./firm $\bar{a}_{sd}$ (5)
US exports 2002 (HS 10-digit level)					
Log GDP	1.01	0.71	0.55	−0.48	0.23
	(0.04)*	(0.03)*	(0.03)*	(0.03)*	(0.02)*
Log distance	−1.37	−1.17	−1.10	0.84	0.05
	(0.17)*	(0.15)*	(0.15)*	(0.13)*	(0.10)
Obs.	175	175	175	175	175
R^2	0.82	0.76	0.68	0.66	0.37
Brazilian exports 2000 (HS 6-digit level)					
Log GDP	0.97	0.56	0.59	−0.56	0.38
	(0.05)*	(0.04)*	(0.04)*	(0.04)*	(0.03)*
Log distance	−2.03	−1.95	−2.37	1.95	0.34
	(0.26)*	(0.18)*	(0.20)*	(0.18)*	(0.16)
Obs.	174	174	174	174	174
R^2	0.67	0.63	0.64	0.63	0.48
Brazilian exports 2000 (NCM 8-digit level)					
Log GDP	0.97	0.56	0.60	−0.56	0.38
	(0.05)*	(0.04)*	(0.04)*	(0.04)*	(0.03)*
Log distance	−2.03	−1.95	−2.40	1.95	0.37
	(0.26)*	(0.18)*	(0.20)*	(0.18)*	(0.16)
Obs.	174	174	174	174	174
R^2	0.67	0.63	0.65	0.63	0.47
Chilean exports 2000 (HS 6-digit level)					
Log GDP	0.85	0.51	0.54	−0.51	0.30
	(0.09)*	(0.05)*	(0.05)*	(0.05)*	(0.05)*
Log distance	−1.05	−1.22	−1.59	1.22	0.54
	(0.41)	(0.23)*	(0.26)*	(0.23)*	(0.25)
Obs.	160	160	160	160	160
R^2	0.39	0.47	0.45	0.47	0.21
Chilean exports 2000 (HS 8-digit level)					
Log GDP	0.85	0.51	0.54	−0.51	0.30
	(0.09)*	(0.05)*	(0.05)*	(0.05)*	(0.05)*
Log distance	−1.05	−1.22	−1.60	1.22	0.55
	(0.41)	(0.23)*	(0.26)*	(0.23)*	(0.25)
Obs.	160	160	160	160	160
R^2	0.39	0.47	0.45	0.47	0.21

Sources: Bernard et al. (2011) for US 2002 manufacturing firms, Brazilian *SECEX* 2000, Chilean customs data 2000; manufacturing firms and their manufactured products. Products at the HS 10-digit level for the USA; at HS 6-digit and NCM 8-digit levels for Brazil; at HS 6-digit and 8-digit levels for Chile.

Note: Total exports T_{sd} are decomposed into $T_{sd} = M_{sd}\,\hat{G}_{sd}\,\mu_{sd}\,\bar{a}_{sd}$, where M_{sd} is the number of exporters in s with shipments to destination d, $\hat{G}_{sd} \equiv \sum_{\omega\in\Omega_{sd}} G_d(\omega)$ is the total number of products exported from s to d by any firm, μ_{sd} is the fraction of firm–product combinations with positive exports which Bernard et al. (2011) call the 'density of trade', and $\bar{a}_{sd} = [\sum_{\omega\in\Omega_{sd}} t_d(\omega)]/[\sum_{\omega\in\Omega_{sd}} G_d(\omega)]$ is the mean exporter scale. Results from country-level ordinary least squares regressions for the dependent variable noted at the top of each column projected on the covariates listed in the first column. Estimates of the constant suppressed. Standard errors in parentheses: * marks statistically significant difference from zero at the 1% level.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

results in a GDP coefficient between 0.03 and 0.04 for Brazil and Chile at either level of product aggregation (see Table A1) and that coefficient is not statistically different from zero at the 1% significance level, contrary to the individual coefficients in Table 1. We conclude that the partial correlation between log average exporter scope $G_d(\omega)$ and log GDP is close to zero, conditional on log distance. In contrast, log distance is significantly negatively related to log exporter scope $G_d(\omega)$, conditional on log GDP.

In summary, gravity-style regressions suggest that the (second) extensive margin of product entry by firms is not significantly related to destination–market size, as measured by GDP, but product entry is related to distance. The reverse is the case for the remaining intensive margin of sales per firm–product. Sales per firm–product are unrelated to distance but significantly related to destination–market size as measured by GDP. Inasmuch models of multi-product exporting strive to match this pattern, they will need to decouple the response of exporter scope to destination–market characteristics at the second extensive margin from the response of exports per firm–product at the remaining intensive margin. Models with flexible fixed product-entry cost functions, such as Arkolakis and Muendler (2010) but conceivably also several others, can achieve the decoupling under specific parameter restrictions.

The absence of a statistical association between sales per firm–product and distance complements recent findings on the relationship between unit prices and distance (Bastos and Silva 2010; Manova and Zhang 2012; Martin 2012; Görg et al. 2010).[10] Martin (2012), for instance, finds for French exporters that doubling the distance to the destination country is associated with an average increase in a firm–product's unit price by 3% for a given firm–product. For Chinese exporters, Manova and Zhang (2012) find in a comparable regression that firms charge a 1% higher unit price for a given firm–product when the destination–country distance doubles. In light of those unit-price regression results, our finding of no response of sales per firm–product to increasing distance may imply that sold quantities decline by roughly the same percentage as prices increase for the same firm–product when distance doubles.

5.2 Mean exporter scope and market size

We now investigate further the relationship between the (second) extensive margin of product entry by firms and destination–market characteristics. In particular, we revisit the earlier finding that the partial correlation

10 We thank an anonymous referee for pointing out this connection.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

between log average exporter scope $G_d(\omega)$ and market size is close to zero. We use manufacturing absorption as a more rigorous measure of market size, consistent with our earlier evidence in Section 3 and in line with the preferred market-size measure in Eaton et al. (2004, 2011). Accordingly, we pursue a new graphical representation here and plot mean exporter scope per destination against the destination's manufacturing absorption. Figure 5 depicts the relationship between exporter scope and the destination country's market size.

Figure 5 confirms graphically that there is no relevant association between average exporter scope $G_d(\omega)$ and market size, now using manufacturing absorption in the place of GDP as our market size measure. In the left panel of Figure 5A and C, the scatter plot shows mean exporter scope at the destination on the vertical axis against market size on the horizontal axis, without conditioning on any distance proxy. While the linear fit suggests a slightly positive slope for Chile, the slope coefficient is not statistically different from zero at the 5% significance level for either Brazil or Chile.

Figure 5B and D repeats the scatter plot but uses mean exporter scope divided by the source country's market share at the destination country ($\lambda_{sd} \equiv T_{sd}/X_d$) on the vertical axis. The source country's market share is thought to be associated with the distance between source and destination and thus serves as a rudimentary control for geography, similar to the approach in Figure 2. The slope coefficient is again not statistically different from zero at the 5% significance level for either Brazil or Chile.

We have seen evidence in Section 3 (Figure 2 for Chile) consistent with the idea that large markets may attract the entry of individual firm–products at a somewhat higher elasticity than the entry of firms. We now have assembled additional evidence on exporter scope to revisit that idea. Note that the evidence in Section 3 does not imply that market size raises exporter scope for a given firm. As we have seen in Section 4, the median firm ships just one or two products even to the largest market (the USA). In this section, we have seen that the mean scope per exporter, too, is insensitive to destination-market size. Together, these findings suggest that market size drives firm entry but does not meaningfully alter the subsequent product entry decision of firms. The insensitivity of exporter scope to a destination's market size also reinforces the earlier conclusion from Section 4 (Figure 4) that the scope distribution is similar across destinations. As previously mentioned, models with flexible fixed product-entry cost functions, such as Arkolakis and Muendler (2010) but conceivably also several others, can generate the insensitivity of exporter scope to market size under specific parameter restrictions.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

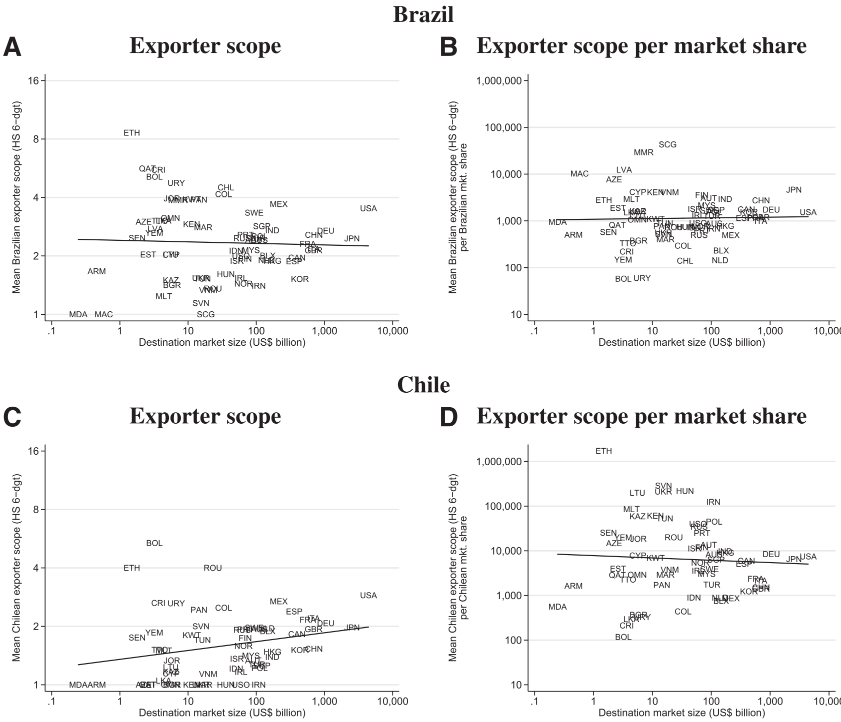

Figure 5 Mean exporter scope and absorption by destination. *Source*: Brazilian *SECEX* 2000, Chilean customs data 2000; manufacturing firms and their manufactured products at the HS 6-digit level, linked to *WTF* (Feenstra et al. 2005) and *Unido* Industrial Statistics (UNIDO 2005). *Note*: Market size is absorption by a country's manufacturing sector. The slopes of the fitted lines are –0.0079 (standard error 0.026) for Brazilian firm's mean exporter scope (A), 0.015 (0.072) for Brazilian firm's mean exporter scope per market share (B), 0.046 (0.023) for Chilean firm's mean exporter scope (C), –0.052 (0.023) for Chilean firm's mean exporter scope per market share (D).

6 Concluding Remarks

We have compared a series of firm-level statistics on product exports across export data sets for four countries. We find a remarkable similarity of the statistics across the four countries, two of which are developing countries and two industrialized. This robustness suggests that these and related firm-level statistics on export products may serve as potential anchors for future theoretical work.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Acknowledgements

We thank Roberto Álvarez for generously sharing Chilean exporter and product data, Ina C. Jäkel for kindly producing statistics from the Danish Globid data base, and Andreas Moxnes for kindly producing statistics from the Norwegian combined exporter-manufacturer data. Muendler and Arkolakis acknowledge NSF support (SES-0550699 and SES–0921673) with gratitude.

References

Álvarez, R., H. Faruq and R. A. López (2007), "New Products in Export Markets: Learning from Experience and Learning from Others", Indiana University, Bloomington, unpublished manuscript.

Anderson, J. E. and E. van Wincoop (2004), "Trade Costs", *Journal of Economic Literature* **42**, 691–751.

Arkolakis, C. (2010), "Market Penetration Costs and the New Consumers Margin in International Trade", *Journal of Political Economy* **118**, 1151–1199.

Arkolakis, C. and M.-A. Muendler (2011), "The Extensive Margin of Exporting Products: A Firm-level Analysis", NBER Working Paper, 16641, National Bureau of Economic Research Cambridge, Massachusetts.

Arkolakis, C., A. Costinot, D. Donaldson and A. Rodríguez-Clare (2012), "The Elusive Pro-competitive Effects of Trade", Yale University, unpublished manuscript.

Bastos, P. and J. Silva (2010), "The Quality of a Firm's Exports: Where You Export to Matters", *Journal of International Economics* **82**, 99–111.

Bergoeing, R., A. Micco and A. Repetto (2011), "Dissecting the Chilean Export Boom", Departamento de Economía de la Universidad de Chile Documentos de Trabajo: Universidad de Chile, Santiago de Chile, 339.

Bernard, A. B., J. B. Jensen and P. K. Schott (2009), "Importers, Exporters, and Multinationals: A Portrait of Firms in the U.S. that Trade Goods", in T. Dunne, J. B. Jensen and M. J. Roberts, eds., *Producer Dynamics: New Evidence from Micro Data,* Vol. 68 of *Studies in Income and Wealth*, University of Chicago Press, Chicago, pp. 513–552.

Bernard, A. B., J. B. Jensen, P. K. Schott, S. J. Redding and P. K. Schott (2007), "Firms in International Trade", *Journal of Economic Perspectives* **21**, 105–30.

Bernard, A. B., J. Eaton, J. B. Jensen and S. Kortum (2003), "Plants and Productivity in International Trade", *American Economic Review* **93**, 1268–1290.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Bernard, A. B., St. J. Redding and P. K. Schott (2011), "Multi-product Firms and Trade Liberalization", *Quarterly Journal of Economics*, advance access July 19, 2011.

Broda, C. and D. E. Weinstein (2006), "Globalization and the Gains from Variety", *Quarterly Journal of Economics* **121**, 541–585.

Chaney, T. (2008), "Distorted Gravity: The Intensive and Extensive Margins of International Trade", *American Economic Review* **98**, 1707–1721.

Chaney, T. (2011), "The Network Structure of International Trade", NBER Working Paper, 16753, National Bureau of Economic Research Cambridge, Massachusetts.

Dhingra, S. (2010), "Trading Away Wide Brands for Cheap Brands", University of Wisconsin-Madison, unpublished manuscript.

Dinopoulos, E., S. Kalyvitis and M. Katsimi (2012), "The Behavior of Exporting Firms under Credit Constraints", Athens University of Economics and Business, unpublished manuscript.

Eaton, J., S. Kortum and F. Kramarz (2004), "Dissecting Trade: Firms, Industries, and Export Destinations", *American Economic Review: Papers and Proceedings* **94**, 150–154.

Eaton, J., S. Kortum and F. Kramarz (2011), "An Anatomy of International Trade: Evidence from French Firms", *Econometrica* **79**, 1453–498.

Eckel, C. and J. P. Neary (2010), "Multi-Product Firms and Flexible Manufacturing in the Global Economy", *Review of Economic Studies* **77**, 188–217.

Fajgelbaum, P. D., G. M. Grossman and E. Helpman (2009), "Income Distribution, Product Quality, and International Trade", NBER Working Paper, 15329, National Bureau of Economic Research Cambridge, Massachusetts.

Feenstra, R. C. and H. Ma (2008), "Optimal Choice of Product Scope for Multiproduct Firms under Monopolistic Competition", in E. Helpman, D. Marin and T. Verdier, eds., *The Organization of Firms in a Global Economy*, Harvard University Press, Cambridge, MA, pp. 173–199.

Feenstra, R., R. E. Lipsey, H. Deng, A. C. Ma and H. Mo (2005), "World Trade Flows: 1962–2000", NBER Working Paper, 11040, National Bureau of Economic Research Cambridge, Massachusetts.

Görg, H., L. Halpern and B. Muraközy (2010), "Why Do Within Firm-product Export Prices Differ Across Markets?", Kiel Working Paper, 1596, Kiel Institute for the World Economy, Kiel.

Iacovone, L. and B. S. Javorcik (2008), "Shipping Good Tequila Out: Investment, Domestic Unit Values and Entry of Multi-product Plants into Export Markets", University of Oxford, unpublished manuscript.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Irarrazabal, A., A. Moxnes and L. D. Opromolla (2010), "The Margins of Multinational Production and the Role of Intra-firm Trade", New York University, unpublished manuscript.

Irarrazabal, A., A. Moxnes and L. D. Opromolla (2011), "The Tip of the Iceberg: A Quantitative Framework for Estimating Trade Costs", Dartmouth College, unpublished manuscript.

Jäkel, I. C. (2012), "Exporting and the Dynamics of Process and Product Innovation", Aarhus University, unpublished manuscript.

Manova, K. and Z. Zhang (2012), "Export Prices Across Firms and Destinations", *Quarterly Journal of Economics* **127**, 379–436.

Martin, J. (2012), "Markups, Quality, and Transport Costs", *European Economic Review* **56**, 777–791.

Mayer, T., M. J. Melitz and G. I. P. Ottaviano (2011), "Market Size, Competition, and the Product Mix of Exporters", NBER Working Paper, 16959, National Bureau of Economic Research Cambridge, Massachusetts.

Melitz, M. J. (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity", *Econometrica* **71**, 1695–1725.

Melitz, M. J. and G. I. P. Ottaviano (2008), "Market Size, Trade, and Productivity", *Review of Economic Studies* **75**, 295–316.

Nocke, V. and S. R. Yeaple (2006), "Globalization and Endogenous Firm Scope", NBER Working Paper, 12322, National Bureau of Economic Research Cambridge, Massachusetts.

Rauch, J. E. (1999), "Networks Versus Markets in International Trade", *Journal of International Economics* **48**, 7–35.

Roberts, M. J. and J. R. Tybout (1997), "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs", *American Economic Review* **87**, 545–564.

Simonovska, I. (2010), "Income Differences and Prices of Tradables", NBER Working Paper, 16233, National Bureau of Economic Research Cambridge, Massachusetts.

UNIDO, 2005 "INDSTAT3 2005 ISIC Rev. 2 User's Guide", Vienna.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Appendix

Table A1 presents short gravity regressions for the three main export margins and their relationship to log GDP and log distance. There are

Table A1: Gravity and the triple exports decomposition

	Log Total exp. T_{sd} (1)	Log # Firms M_{sd} (2)	Log # Products/ firm $\bar{G}_{sd}$ (3)	Log sales/# prod./ firm $\bar{a}_{sd}$ (4)
Brazilian exports 2000 (HS 6-digit level)				
Log GDP	0.97	0.56	0.03	0.38
	(0.05)*	(0.04)*	(0.01)	(0.03)*
Log distance	−2.03	−1.95	−0.42	0.34
	(0.26)*	(0.18)*	(0.07)*	(0.16)
Obs.	174	174	174	174
R^2	0.67	0.63	0.19	0.48
Brazilian exports 2000 (NCM 8-digit level)				
Log GDP	0.98	0.57	0.04	0.38
	(0.05)*	(0.04)*	(0.01)	(0.03)*
Log distance	−2.01	−1.93	−0.46	0.38
	(0.26)*	(0.18)*	(0.07)*	(0.16)
Obs.	175	175	175	175
R^2	0.67	0.63	0.21	0.48
Chilean exports 2000 (HS 6-digit level)				
Log GDP	0.86	0.52	0.03	0.31
	(0.08)*	(0.05)*	(0.01)	(0.05)*
Log distance	−1.02	−1.21	−0.37	0.56
	(0.41)	(0.22)*	(0.06)*	(0.25)
Obs.	161	161	161	161
R^2	0.40	0.47	0.19	0.22
Chilean exports 2000 (HS 8-digit level)				
Log GDP	0.86	0.52	0.03	0.31
	(0.08)*	(0.05)*	(0.01)	(0.05)*
Log distance	−1.02	−1.21	−0.38	0.57
	(0.41)	(0.22)*	(0.06)*	(0.25)
Obs.	161	161	161	161
R^2	0.40	0.47	0.19	0.22

Sources: Brazilian *SECEX* 2000, Chilean customs data 2000; manufacturing firms and their manufactured products. Products at the HS 6-digit and NCM 8-digit levels for Brazil; at HS 6-digit and 8-digit levels for Chile.

Note: Total exports T_{sd} are decomposed into $T_{sd} = M_{sd}\,\bar{G}_{sd}\,\bar{a}_{sd}$, where M_{sd} is the number of exporters in s with shipments to destination d, $\bar{G}_{sd} \equiv \sum_{\omega\in\Omega_{sd}} G_d(\omega)/M_{sd}$ is the exporter's mean exporter scope, and $\bar{a}_{sd} \equiv \bar{t}_{sd}/\bar{G}_{sd}$ is their varietie's mean exporter scale. Results from country-level ordinary least squares regressions of the dependent variable noted at the top of each column on the covariates. Estimates for the constant suppressed. Standard errors in parentheses: * marks statistically significant difference from zero at the 1% level.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

two extensive margins of export activity: first firm entry at a given destination, and second product entry by the same firm at a given destination. The second extensive margin gives rise to a firm's *exporter scope* at a destination. The remaining third, intensive margin captures individual firm–product sales at a destination. A related three-way decomposition of total exports to a destination (see Equation (3)) breaks export sales down into the number of firms shipping to the destination, their average exporter scope, and their average exporter scale (their mean firm–product sales). Those are the three dependent variables in Table A1. Following Bernard et al. (2011), the companion Table 1 in the text considers a quadruple decomposition (see Equation (5)).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

CESifo Economic Studies, Vol. 59, 2/2013, 249–276 doi:10.1093/cesifo/ifs002
Advance Access publication 15 February 2012

Is the International Border Effect Larger than the Domestic Border Effect? Evidence from US Trade

Cletus C. Coughlin* and Dennis Novy†

*Office of the President, Federal Reserve Bank of St Louis, PO Box 442, St Louis, MO 63166-0442, USA. e-mail: coughlin@stls.frb.org
†Department of Economics and Centre for Competitive Advantage in the Global Economy (CAGE), University of Warwick, Coventry CV4 7AL, UK; Centre for Economic Policy Research (CEPR); Centre for Economic Performance (CEP); CESifo. e-mail: d.novy@warwick.ac.uk

Abstract

Many studies have found that international borders represent large barriers to trade. But how do international borders compare to domestic border barriers? We investigate international and domestic border barriers in a unified framework. We consider a data set of exports from individual US states to foreign countries and combine it with trade flows between and within US states. After controlling for distance and country size, we estimate that relative to state-to-state trade, crossing an individual US state's domestic border appears to entail a *larger* trade barrier than crossing the international US border. Due to the absence of governmental impediments to trade within the United States, this result is surprising. We interpret it as highlighting the concentration of economic activity and trade flows at the local level. (JEL codes: F10, F15)

Keywords: International border, intranational home bias, domestic border, gravity, trade costs, distance

1 Introduction

In a seminal article, McCallum (1995) found that Canadian provinces trade up to 22 times more with each other than with US states. This astounding result, also known as the *international border effect*, has led to a large literature on the trade impediments associated with international borders. More recently, Anderson and van Wincoop (2003) revisited the US–Canadian border effect with new micro-founded estimates. Although they are able to reduce the border effect considerably, there is widespread consensus that the international border remains a large impediment to trade.[1]

[1] Anderson and van Wincoop (2004) report 74% as an estimate of representative international trade costs for industrialized countries (expressed as a tariff equivalent). About two-thirds of these costs can be attributed to border-related trade barriers such as tariffs and non-tariff barriers. The remainder represents transportation costs. While McCallum (1995) compares trade between Canadian provinces and US states to inter-provincial trade, Anderson and van Wincoop (2003) add inter-state trade data.

© The Author 2012. Published by Oxford University Press
on behalf of Ifo Institute for Economic Research, Munich. All rights reserved.
For permissions, please email: journals.permissions@oup.com

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

A parallel, smaller literature has documented that border effects also exist within a country, known as the *domestic border effect* or intranational home bias. For example, Wolf (2000) finds that trade within individual US states is significantly larger than trade between US states even after he controls for economic size, distance, and a number of additional determinants. Similarly, despite the absence of formal international trade barriers associated with the single market, Nitsch (2000) finds that domestic trade within the average European Union country is about 10 times larger than trade with another EU country.[2]

It is important to understand the nature of domestic and international trade barriers since they might impede the integration of markets and have negative welfare consequences. Accurately identifying the magnitudes of border effects at the domestic and international levels is a necessary step for assessing their economic significance. The contribution of this article is to merge the two strands of literature about border effects into a unified framework. We construct a data set that includes three tiers of US trade flows: (i) trade within individual US states, e.g., Minnesota–Minnesota; (ii) trade between US states, e.g., Minnesota–Texas; and (iii) trade between US states and foreign countries, e.g., Minnesota–Canada.[3]

We use gravity theory to estimate the relative size of the domestic and international border effects. As is typical in the literature, the domestic border effect indicates how much a US state trades with itself relative to state-to-state trade, while the international border effect indicates how much a US state trades with foreign countries relative to state-to-state trade. After controlling for distance and economic size, we find that relative to state-to-state trade, crossing an individual US state's domestic border entails a *larger* trade barrier than crossing the international US border. Put differently, although trading internationally is of course more costly in total than trading intranationally, our results indicate that the estimated marginal increase in trade barriers when leaving the domestic state is relatively larger than the increase associated with leaving the United States.

What are the economic reasons behind the large domestic border effect? International trade economists traditionally emphasize trade barriers

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

2 An earlier study by Wei (1996) finds similar results for OECD countries. Nikolaus Wolf (2009) finds sizeable domestic border barriers in the historical context for Germany in the late 19th and early 20th centuries. Chen (2004) documents significant intra-European Union border effects at the industry level.

3 Other papers, such as Hillberry and Hummels (2008), have used geographically more finely disaggregated US trade data. However, these data and the related papers pertain only to the question of the domestic border effect. They are silent on the international border effect. Our innovation is in combining US domestic and international trade data for the first time.

associated with international borders such as tariffs, bureaucratic hurdles, and informational barriers. Although beginning with Wolf (2000) and Nitsch (2000) the empirical literature has also demonstrated that borders within a country are associated with a significant trade-impeding effect, it is much harder to think of administrative and informational barriers that coincide with state borders within the same country. Instead, one plausible explanation is related to work by Hillberry and Hummels (2008). Based on ZIP-code-level domestic US trade flows, they document that trade within the United States is heavily concentrated at the local level. In particular, trade within a single ZIP code is on average three times higher than trade with partners outside the ZIP code. This concentration might be due to the prevalence of trade in intermediate goods at the local level, arguably as a result of supply chain optimization as companies seek to minimize transportation costs and suppliers co-locate with final goods producers. This high concentration of trade at the local level implies large domestic border barrier estimates. In that interpretation, the estimated domestic border effect does not reflect state-border barriers per se but rather local agglomeration effects. But of course, the fact that firms cluster in areas as small as a single ZIP code might be indicative in itself of trade costs associated with relatively short distances. As we discuss in Section 5, other reasons for the strong local concentration of trade include informational and search costs, for example in the form of business, social, and immigration networks, increasing returns at the local level as well as location-specific tastes.

Given the large literature on border effects, it can arguably be seen as a logical extension to estimate international and domestic border effects in a joint framework so that they can be directly compared. In fact, research by Fally, Paillacar, and Terra (2010) is related to our work. As part of a study examining wage differences across Brazilian states, they estimate a gravity equation in which bilateral trade flows are explained by a set of trade cost variables that include both domestic and international border effects. Consistent with our results for the United States, their estimates imply that the average Brazilian state border has a relatively *larger* negative impact on bilateral trade flows than the international border.[4]

[4] Given the three sets of trade flows and two dummy variables reflecting border effects, it is necessary to decide which set of trade flows to use as the base or omitted category. In our paper the base is trade between US states, while Fally, Paillacar, and Terra (2010) use trade within Brazilian states as the base. Thus, we generate a positive estimate for the ownstate border effect and a negative estimate for the international border effect, while Fally, Paillacar, and Terra (2010) generate negative estimates for both border effects. In other words, relative to state-to-state trade, we find that within-state trade is relatively higher and international trade is relatively lower. For Fally, Paillacar, and Terra (2010), relative to within-state trade, both state-to-state trade and international trade are lower. In the first column of their Table 2, they report an estimate of −2.594 for their internal border dummy and an estimate of −4.326 for their international border dummy in a

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

On the other hand, results using Chinese trade data indicate that in a number of instances the domestic (i.e., provincial) border tends to have a relatively *smaller* negative effect on trade flows than the international border. For example, Poncet (2003) finds that the international border effect exceeds the domestic border effect for 1987 and 1992 (but not for 1997). Similarly, the results by De Sousa and Poncet (2011) indicate that the international border effect exceeds the domestic border effect for the years 1995, 1999, 2002, 2005, and 2007.[5] In contrast, Hering and Poncet (2010) find that the domestic border effect exceeds the international border effect for 1997.

The article is organized as follows. In Section 2, we carefully examine the general equilibrium theory of trade with trade barriers to derive our empirical estimation framework. In Section 3, we describe the data set which we use in Section 4 to estimate international and domestic border effects. In Section 5, we discuss a number of potential explanations for our empirical results. Section 6 concludes.

2 Gravity Theory and the Estimation Framework

2.1 Gravity theory

Gravity equations can be derived from a variety of trade models, such as the gravity framework with multilateral resistance by Anderson and van Wincoop (2003), the Ricardian trade model by Eaton and Kortum (2002), Chaney's (2008) extension of the Melitz (2003) heterogeneous firms model as well as the heterogeneous firms model by Melitz and Ottaviano (2008) with a linear demand system.[6] To obtain results that are easily comparable to the previous literature on border effects, we adopt the widely used gravity framework by Anderson and van Wincoop (2003). Our results, however, could also be generated with the other frameworks.

Anderson and van Wincoop's (2003) parsimonious model rests on the Armington assumption that countries produce differentiated goods and

log-linear regression with exporter and importer fixed effects and controls for distance and other bilateral trade costs. Their border estimates are directly comparable to ours due to the Frisch-Waugh theorem. Their estimates imply that trade within Brazilian state is on average 13.4 times larger than trade between Brazilian states [exp(2.594) = 13.4], whereas trade between Brazilian states is only 5.7 times larger than trade with foreign countries [exp(4.326−2.594) = 5.7]. In that sense, their results also imply that the domestic border appears to entail a larger trade barrier than the international border.

5 It is unclear though whether the differences between the domestic and international border effect point estimates are statistically significant, especially for the earlier years. Similar to the previous footnote, the coefficients have to be transformed appropriately to make them directly comparable to ours.

6 See Chen and Novy (2011) for an overview.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

trade is driven by consumers' love of variety. They derive the following gravity equation for exports x_{ij} from region i to region j:

$$x_{ij} = \frac{y_i y_j}{y^W} \left(\frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma}, \tag{1}$$

where y_i and y_j denote output of regions i and j, y^W denotes world output, t_{ij} is the bilateral trade cost factor (one plus the tariff equivalent), Π_i is the outward multilateral resistance term, and P_j is the inward multilateral resistance term. The parameter $\sigma > 1$ is the elasticity of substitution. The bilateral trade costs t_{ij} capture a variety of trade frictions such as transportation costs, tariffs, and bureaucratic barriers, and they also include the border barriers.

2.2 The estimation framework

We follow McCallum (1995) and other authors by hypothesizing that trade costs t_{ij} are a log-linear function of geographic distance, $dist_{ij}$, and a border dummy, $INTERNATIONAL_{ij}$, which takes on the value 1 whenever regions i and j are located in different countries. In addition, we hypothesize that domestic trade costs within a region's own territory might be systematically different from bilateral trade costs. We therefore include an ownstate dummy variable, $OWNSTATE_{ij}$, that takes on the value 1 for $i = j$. Our trade cost function can thus be expressed as

$$\ln(t_{ij}) = \tilde{\beta}\, INTERNATIONAL_{ij} + \tilde{\gamma}\, OWNSTATE_{ij} + \tilde{\delta} \ln(dist_{ij}), \tag{2}$$

where $\tilde{\beta}$ and $\tilde{\gamma}$ reflect the international and the ownstate (i.e., domestic) border effects, respectively, and $\tilde{\delta}$ is the elasticity of trade costs with respect to distance.

The trade cost function (2) nests the trade cost functions used by Wolf (2000), Hillberry and Hummels (2003), and Anderson and van Wincoop (2003). Wolf (2000) and Hillberry and Hummels (2003) only consider trade flows within the US so that an international border effect cannot be estimated. This corresponds to $\tilde{\beta} = 0$ in Equation (2). Anderson and van Wincoop (2003) follow McCallum's (1995) specification that does not allow for a domestic border effect ($\tilde{\gamma} = 0$).

We log-linearize Equation (1) so that we obtain

$$\ln(x_{ij}) = \ln(y_i) + \ln(y_j) - \ln(y^W) + (1 - \sigma)\ln(t_{ij}) + (\sigma - 1)\ln(\Pi_i P_j). \tag{3}$$

Substituting the trade cost function (2) yields the following estimating equation:

$$\begin{aligned} \ln(x_{ij}) = {} & \ln(y_i) + \ln(y_j) + \beta\, INTERNATIONAL_{ij} + \gamma\, OWNSTATE_{ij} \\ & + \delta \ln(dist_{ij}) + (\sigma - 1)\ln(\Pi_i P_j) + \alpha + \varepsilon_{ij}, \end{aligned} \tag{4}$$

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

where $\beta = (1-\sigma)\ \tilde{\beta}$, $\gamma = (1-\sigma)\ \tilde{\gamma}$, and $\delta = (1-\sigma)\ \tilde{\delta}$ and where the logarithm of world output is captured by the constant α and where we add a white-noise error term ε_{ij}.

2.3 Border effects in theory

The empirical literature typically finds that international borders impede trade. This corresponds to $\beta < 0$ in estimating Equation (4). Trading within a state is typically associated with higher trade flows, corresponding to $\gamma > 0$. We first examine whether gravity theory allows us to predict whether the international border effect β is larger or smaller in absolute value than the domestic border effect γ, i.e., whether $|\beta| \gtrless |\gamma|$.

As we explain below in more detail, our data set comprises three tiers of trade flows:

(i) ownstate trade: trade flows within a US state, for example within Minnesota, such that $OWNSTATE_{ij}=1$ and $INTERNATIONAL_{ij}=0$,
(ii) national trade: trade flows between two US states, for example from Minnesota to Texas, such that $OWNSTATE_{ij}= INTERNATIONAL_{ij}=0$, and
(iii) international trade: trade flows from a US state to a foreign country, for example from Minnesota to Canada, such that $OWNSTATE_{ij}=0$ and $INTERNATIONAL_{ij}=1$.

The second tier is thus the omitted category in Equation (4), implying that the ownstate border effect is estimated relative to the benchmark of trade between US states. We choose this benchmark to obtain coefficients which are directly comparable to those in the literature (Nitsch, 2000; Wolf, 2000). Therefore, the sign and magnitude of the ownstate border effect can be gauged by comparing trade costs t_{ii} within a typical US state i to bilateral trade costs t_{ij} with another US state j. We draw this comparison by considering their ratio t_{ii}/t_{ij}. Equation (1) for ownstate trade x_{ii} and bilateral trade x_{ij} and Equation (2) for t_{ii} and t_{ij} imply that this ratio is given by

$$\frac{t_{ii}}{t_{ij}} = \left(\frac{x_{ij}y_i}{x_{ii}y_j}\right)^{\frac{1}{\sigma-1}}\frac{P_i}{P_j} = \frac{\exp(\tilde{\gamma})(dist_{ii})^{\tilde{\delta}}}{(dist_{ij})^{\tilde{\delta}}}.$$

Using $\tilde{\gamma}=\gamma/(1-\sigma)$ and $\tilde{\delta}=\delta/(1-\sigma)$ this can be rewritten as

$$\exp(\gamma) = \frac{x_{ii}}{x_{ij}}\frac{y_j}{y_i}\left(\frac{P_j}{P_i}\right)^{\sigma-1}\left(\frac{dist_{ij}}{dist_{ii}}\right)^{\delta}. \tag{5}$$

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

As a simple example, first assume the symmetric case where $y_i = y_j$, $P_i = P_j$, and $dist_{ii} = dist_{ij}$. A positive ownstate effect $\gamma > 0$ would follow only if $x_{ii}/x_{ij} > 1$. Now assume the more representative case where bilateral distance $dist_{ij}$ exceeds domestic distance $dist_{ii}$. Given that the distance elasticity of trade is negative ($\delta < 0$), an even bigger ratio x_{ii}/x_{ij} would be required to ensure $\gamma > 0$. More generally, we conclude that given the distance element of trade costs as well as the output and multilateral resistance variables, the sign and magnitude of the domestic border effect parameter γ will primarily depend on the extent of domestic trade x_{ii} relative to bilateral trade x_{ij}.

As in the literature, we also use the benchmark of trade between US states for estimating the international border effect. To gauge its sign and magnitude, we compare bilateral trade costs t_{ik} between a typical US state i and a typical foreign country k to trade costs t_{ij} between two US states. Their ratio is given by

$$\frac{t_{ik}}{t_{ij}} = \left(\frac{x_{ij}}{x_{ik}}\frac{y_k}{y_j}\right)^{\frac{1}{\sigma-1}}\frac{P_k}{P_j} = \frac{\exp(\tilde{\beta})(dist_{ik})^{\tilde{\delta}}}{(dist_{ij})^{\tilde{\delta}}},$$

or

$$\exp(\beta) = \frac{x_{ik}}{x_{ij}}\frac{y_j}{y_k}\left(\frac{P_j}{P_k}\right)^{\sigma-1}\left(\frac{dist_{ij}}{dist_{ik}}\right)^{\delta}. \tag{6}$$

As before, assume the simple symmetric case where $y_k = y_j$, $P_k = P_j$, and $dist_{ik} = dist_{ij}$. A negative international border effect $\beta < 0$ would follow only if $x_{ik}/x_{ij} < 1$. In the more common case where international distance $dist_{ik}$ (say, between Minnesota and Japan) exceeds inter-state distance $dist_{ij}$ (say, between Minnesota and Texas), an even smaller ratio x_{ik}/x_{ij} would be required to ensure $\beta < 0$. Given distances as well as the output and multilateral resistance variables, the international border effect parameter β will therefore mainly depend on the extent of international trade x_{ik} relative to inter-state trade x_{ij}.

Thus, Equations (5) and (6) can in principle yield either sign for γ and β. The fact that most empirical studies find $\gamma > 0$ or $\beta < 0$ is consistent with but by no means implied by gravity theory. Neither does gravity theory make a prediction about the absolute magnitudes of β and γ. *A priori* we therefore cannot infer whether $|\beta| \gtrless |\gamma|$.[7]

[7] The conclusion that β and γ are not bounded by theory would also go through if we relaxed the symmetry assumption for the output and multilateral resistance variables.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

3 Data

To obtain comparable results, we use the same data sets as Wolf (2000) and Anderson and van Wincoop (2003) for domestic trade flows within the United States. The novelty of our approach is to combine these domestic trade flows with international trade flows from individual US states to the 50 largest US export destinations. Thus, our data set comprises, for instance, trade flows within Minnesota, exports from Minnesota to Texas as well as exports from Minnesota to Canada.[8] We take data quality seriously, and below we describe in detail the data sources, potential concerns, and how we address these concerns.

3.1 Domestic exports: Commodity Flow Survey

For our measures of the shipments of goods within and across US states, we use aggregate trade data from the Commodity Flow Survey, which is a joint effort of the Bureau of Transportation Statistics and the Bureau of the Census. We use survey results from 1993, 1997, and 2002. The survey covers the origin and destination of shipments of manufacturing, mining, wholesale trade, and selected retail establishments. The survey excludes shipments in the following sectors: services, crude petroleum and natural gas extraction, farm, forestry, fishery, construction, government, and most retail. Shipments from foreign establishments are also excluded; import shipments are excluded until they reach a domestic shipper. US export (i.e., trans-border) shipments are also excluded.

3.2 International exports: Origin of Movement

Our data on exports by US states to foreign destinations are from the Origin of Movement series.[9] These data are compiled by the Foreign Trade Division of the US Bureau of the Census. The data in this series identify the state from which an export begins its journey to a foreign country. However, we would like to know the state in which the export was produced. Below we provide details on the Origin of Movement series and its suitability as a measure of the origin of production.[10]

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

[8] There are similarities and differences between the data sets used in Anderson and van Wincoop (2003) and our work. As noted, we both combine domestic and international trade flows. For example, we both use state-to-state trade flows (48 states in our case and 30 states in Anderson and van Wincoop) as well as trade flows that cross international borders. The key difference is that our data set additionally includes intra-state flows. As a result, we are able to estimate both state and international border effects, while Anderson and van Wincoop focus on the latter only.

[9] Other studies that have used the Origin of Movement series include Smith (1999), Coughlin and Wall (2003), Coughlin (2004), and Cassey (2011).

[10] The highlighted details as well as much additional information can be found in Cassey (2009).

Beginning in 1987, the Origin of Movement series provides the current-year export sales, or free-alongside-ship (f.a.s.) costs if not sold, for 54 ‘states’ to 242 foreign destinations. These export sales are for merchandize sales only and do not include services exports. The 54 ‘states’ include the 50 US states plus the District of Columbia, Puerto Rico, US Virgin Islands, and unknown. Following Wolf (2000), we use the 48 contiguous US states. Rather than all 242 destinations, we use the 50 leading export destinations for US exports for 2005.[11] We use the annual data from 1993, 1997, and 2002 for total merchandize exports.[12]

Concerns about using the Origin of Movement series to identify the location of production are especially pertinent for agricultural and mining exports.[13] Cassey (2009) has examined the issue of the coincidence of the state origin of movement and the state of production for manufactured goods.[14] The reason for restricting the focus to manufacturing is that the best source for location-based data on export production, ‘Exports from Manufacturing Establishments,’ covers only manufacturing.[15]

Cassey’s key finding relevant to our analysis is that, overall, the Origin of Movement data is of sufficient quality to be used as the origin of the production of exports. Nonetheless, the data for specific states may not be of sufficient quality as the origin of production. These states are Alaska, Arkansas, Delaware, Florida, Hawaii, New Mexico, South Dakota, Texas, Vermont, and Wyoming. He recommends the removal of Alaska and Hawaii in particular. As we use the 48 contiguous US states, our data set is consistent with this recommendation. The next two candidates for removal would be Delaware and Vermont. Cassey further highlights that the consolidation of export shipments might systematically affect the Origin of Movement estimates (relative to the origin of production). Specifically, consolidation tends to bias upward the estimates for Florida and Texas and to bias downward the estimates for Arkansas

11 Alphabetically, the countries are Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, China, Colombia, Costa Rica, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Finland, France, Germany, Guatemala, Honduras, Hong Kong, India, Indonesia, Ireland, Israel, Italy, Japan, Kuwait, Malaysia, Mexico, Netherlands, New Zealand, Norway, Panama, Peru, Philippines, Russia, Saudi Arabia, Singapore, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Turkey, United Arab Emirates, UK, and Venezuela.

12 We have also tried the data for manufacturing only (as opposed to total merchandise). The two series are very highly correlated (99%). The regression results are almost identical, and we therefore do not report them.

13 See http://www.trade.gov/td/industry/otea/state/technote.html.

14 For the initial work on this issue, see Coughlin and Mandelbaum (1991) as well as Cronovich and Gazel (1999).

15 The data in the ‘Exports from Manufacturing Establishments’ are available at http://www.census.gov/mcd/exports/ but does not contain destination information, so it cannot be used for the current research project.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

and New Mexico. As a robustness check, we drop these states from the sample (see Section 4.3).

3.3 Adjustments for data comparability

Our simultaneous use of the intra-state and inter-state shipments data from the Commodity Flow Survey and the merchandize international trade data from the Origin of Movement series requires an adjustment to increase the comparability of these data sets. As in Anderson and van Wincoop (2003), such an adjustment arises because of three important differences between the data sources, the net effect of which is to increase the commodity flow estimates relative to the international trade flow estimates. First, the merchandize international trade data capture a shipment from the source to the port of exit just once, whereas the commodity flow data likely captures a good in a shipment more than once, recorded in more than one shipment. For example, a good may be shipped from a plant to a warehouse and, later, to a retailer. In this case, the value of the good will be counted twice. But if the good had been exported, its value would have been counted just once as it was shipped from the source to the port of exit. Second, goods destined for foreign countries, when they are shipped to a port of exit, are included in domestic shipments. Third, the coverage of sectors differs between the data sources. The Commodity Flow Survey includes shipments of manufactured goods, but it excludes agriculture and part of mining. Meanwhile, the merchandize trade data include all goods.

Identical to Anderson and van Wincoop (2003), we scale down the data in the Commodity Flow Survey by the ratio of total domestic merchandize trade to total domestic shipments from the Commodity Flow Survey. Total domestic merchandize trade is approximated by gross output in the goods-producing sectors (i.e., agriculture, mining, and manufacturing) minus international merchandize exports.[16] This calculation yields an adjustment factor of 0.495 for 1993, 0.508 for 1997, and 0.430 for 2002.[17] Similar to Anderson and van Wincoop (2003), our adjustment to the commodity flow data does not solve all the measurement problems, but it is the best feasible option.

3.4 Other data

The rest of the data used in our estimations can be characterized as either well-known or straightforward. For individual US states, we use state

[16] See Helliwell (1997, 1998) and Wei (1996).

[17] The difference between our adjustment factor for 1993 and that of Anderson and van Wincoop, 0.495 versus 0.517, is due to data revision.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

gross domestic product data from the US Bureau of Economic Analysis. For foreign countries, we use data on gross domestic product taken from the IMF World Economic Outlook Database (October 2007 edition).

We use the standard great circle distance formula to cover inter-state and international distances between capital cities in kilometers. As intra-state distance, we use the distance between the two largest cities in a state. As alternatives for intra-state distance, we also try the measure used by Wolf (2000) that weights the distance between a state's two largest cities by their population, as well as the measure suggested by Nitsch (2000) that is based on land area. Finally, we also use a distance measure that is related to actual shipping distances, based on data for individual shipments used by Hillberry and Hummels (2003), see Section 4.3 for details.

4 Empirical Results

We form a sample that is balanced over the years 1993, 1997, and 2002. This yields 1801 trade observations per cross-section within the US.[18] Adding 50 foreign countries as export destinations increases the number of trade observations by 2338 so that our sample includes 4139 observations per cross-section, or 12 417 in total.[19] Recall that due to the data quality concerns as well as for consistency reasons, Alaska, Hawaii, and Washington, DC were dropped, so we use the 48 contiguous US states.

First, we show that our data exhibit a substantial domestic border effect, as established by Wolf (2000). In separate regressions, we also show that the data exhibit a significant international border effect, as established by McCallum (1995). Second, we combine the domestic US trade data with the international observations. This allows us to estimate the domestic and international border effects jointly and to directly compare their magnitudes. Finally, we carry out a number of robustness checks.

4.1 Domestic and international border effects estimated separately

In Columns 1 and 2 of Table 1, we show results that replicate the intranational home bias. For comparison with Wolf (2000) who uses a sample

18 The maximum possible number of US observations would be $48 \times 48 = 2304$. The 503 missing observations are due to the fact that a number of Commodity Flow Survey estimates did not meet publication standards because of high sampling variability or poor response quality.

19 The maximum possible number of international observations would be $48 \times 50 = 2400$. Sixty-two observations are missing mainly because exports to Malaysia were generally not reported in 1993. Only 18 of the observations not included in our sample are most likely zeros (as opposed to missing).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 1 Domestic and international border effects, estimated separately

Sample	United States only			United States and 50 countries	
Year	1993	1993, 1997, 2002	1993, 1997, 2002	1993	1993, 1997, 2002
	(1)	(2)	(3)	(4)	(5)
$ln(y_i)$			0.92**	1.29**	1.22**
			(0.02)	(0.02)	(0.02)
$ln(y_j)$			0.91**	0.83**	0.83**
			(0.02)	(0.01)	(0.01)
$ln(dist_{ij})$	−1.08**	−1.07**	−0.94**	−0.86**	−0.85**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
$OWNSTATE_{ij}$	1.46**	1.48**	1.76**		
	(0.20)	(0.19)	(0.19)		
$INTERNATIONAL_{ij}$				−1.19**	−1.04**
				(0.06)	(0.05)
National trade (reference group)	Yes	Yes	Yes	Yes	Yes
Ownstate trade	Yes	Yes	Yes	No	No
International trade	No	No	No	Yes	Yes
Observations	1801	5403	5403	4091	12 273
Clusters	–	1801	1801	–	4091
Fixed effects	Yes	Yes	No	No	No
Random effects	No	No	Yes	No	Yes
R-squared	0.90	0.90	0.82	0.79	0.78

Notes. The dependent variable is $ln(x_{ij})$. OLS estimation. Robust standard errors are reported in parentheses, clustered around country pairs *ij* in Columns 2, 3, and 5. Exporter and importer fixed effects in Columns 1 and 2, time-varying in Column 2; random effects in Columns 3 and 5. Constants and year dummies are not reported. ** Significant at 1% level.

for 1993, in Column 1 we only use data for that year. In Column 2, we add the data for 1997 and 2002. Like Hillberry and Hummels (2003), we use (year-specific) exporter and importer fixed effects so that the output regressors drop out. Our point estimates in Columns 1 and 2 (1.46 and 1.48) are virtually identical to Wolf's baseline coefficient of 1.48 for the ownstate indicator variable. The interpretation of this coefficient is that given distance and economic size, ownstate trade is 4.4 times higher than state-to-state trade [exp(1.48) = 4.4]. As we will use random effects in subsequent tables (see below), we also run a random effects specification that

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

corresponds to the fixed effects specification in Column 2. The results are reported in Column 3. Output regressors are now included. The ownstate coefficient is slightly higher (1.76 compared to 1.48 in column 2) but we cannot reject the hypothesis that the ownstate coefficient is equal to 1.48 (p-value = 0.15).

Hillberry and Hummels (2003) reduce the ownstate coefficient by about a third when excluding wholesale shipments from the Commodity Flow Survey data. The reason is that wholesale shipments are predominantly local so that their removal disproportionately reduces the extent of ownstate trade.[20] However, Nitsch (2000) reports higher home bias coefficients by comparing trade within European Union countries to trade between EU countries. He finds home bias coefficients in the range of 1.8–2.9.

In Columns 4 and 5, we do not consider ownstate trade but rather focus on the international border effect. These regressions use the sample of 50 foreign countries. In Column 4, we estimate an international border coefficient of −1.19 for the year 1993, implying that after we control for distance and economic size, exports from US states to foreign countries are about 70% lower than exports to other US states [$\exp(-1.19) = 0.30$]. This coefficient is somewhat lower in magnitude than the estimate of −1.65 obtained by Anderson and van Wincoop (2003) with trade data between US states and Canadian provinces. When we pool the data over the years 1993, 1997, and 2002 in Column 5, the border effect is estimated at −1.04. Estimation in that column is carried out with random effects since fixed effects at the country level would be collinear with the international border dummy variable.

Overall, we conclude that we obtain estimates for domestic and international border effects in Table 1 that are broadly consistent with the literature.

4.2 Is the international border effect larger than the domestic border effect?

In Table 2, we turn to estimating the domestic and international border effects jointly so that their magnitudes are directly comparable. For this purpose, we simultaneously use domestic and international trade flows, while continuing to use inter-state trade as the reference group as in Table 1. When we pool the data over the years 1993, 1997, and 2002 in

[20] We do not have access to the private-use coding of wholesale shipments and thus cannot replicate their finding with our data. However, our main results in Table 2 on the relative size of the domestic and international border effects is qualitatively robust to a reduction by a third in the ownstate coefficient magnitudes. Hillberry and Hummels (2003) further reduce the ownstate coefficient by using an alternative distance measure that is based on actual shipping distances. We refer to Section 4.3 where we employ such a measure, but our main result is unchanged.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 2 Domestic and international border effects, estimated jointly

Sample	United States and 50 countries	
Year	1993	1993, 1997, 2002
	(1)	(2)
$ln(y_i)$	1.28**	1.21**
	(0.02)	(0.02)
$ln(y_j)$	0.82**	0.82**
	(0.01)	(0.01)
$ln(dist_{ij})$	−0.83**	−0.82**
	(0.03)	(0.03)
$OWNSTATE_{ij}$	2.04**	2.05**
	(0.20)	(0.20)
$INTERNATIONAL_{ij}$	−1.24**	−1.10**
	(0.06)	(0.05)
$\lvert OWNSTATE_{ij}\rvert=\lvert INTERNATIONAL_{ij}\rvert$	[0.00]	[0.00]
National trade (reference group)	Yes	Yes
Ownstate trade	Yes	Yes
International trade	Yes	Yes
Observations	4139	12 417
Clusters	–	4139
Random effects	No	Yes
R^2	0.79	0.79

Notes. The dependent variable is $ln(x_{ij})$. OLS estimation. Robust standard errors are reported in parentheses, clustered around country pairs ij in Column 2. Random effects in Column 2. Constants and year dummies are not reported. **Significant at 1% level. The numbers in brackets report p-values for the test $\lvert OWNSTATE_{ij}\rvert=\lvert INTERNATIONAL_{ij}\rvert$.

Column 2, we add random effects instead of country fixed effects. The reason is again that country fixed effects would be perfectly collinear with the ownstate and international dummy variables. Exporter and importer fixed effects would also be impractical because of collinearity with the international border dummy.[21]

Columns 1 and 2 show that the ownstate coefficients are estimated at 2.04 and 2.05, while the international coefficients are estimated at −1.24 and −1.10. The hypothesis that the two border coefficients in each column are equal in absolute magnitude is clearly rejected (p-value = 0.00). Thus, a key finding in Table 2 is that the domestic border effect is larger in

[21] The collinearity arises because the foreign countries in our data set are only importers but never exporters.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

absolute magnitude than the international border effect. That is, relative to inter-state trade, crossing an individual US state's domestic border is estimated to entail a *larger* trade barrier than crossing the international US border.

Another observation is that the joint estimation in Table 2 yields somewhat different estimates of the domestic border effect. The coefficient on $OWNSTATE_{ij}$ is 1.48 when estimated separately with fixed effects (see Table 1, Column 2) and 1.76 when estimated separately with random effects (see Table 1, Column 3), and 2.05 when estimated jointly (see Table 2, Column 2).[22] Note that the distance coefficient in those columns changes from −1.07 and −0.94, respectively, to −0.82, and the latter value is close to the distance coefficients in Columns 4 and 5 of Table 1. Likewise, the income elasticities are also similar to those estimated in Columns 4 and 5 of Table 1.

4.3 Robustness

Various authors, such as Helliwell and Verdier (2001) and Head and Mayer (2009), have pointed out that the estimation of border effects is sensitive to how distance is measured. For example, if the relevant economic distance within a US state is much shorter than indicated by conventional measures—perhaps because economic activity is highly concentrated in two nearby cities—then it might no longer be surprising if a state trades considerably more within its boundaries than with partners further away. To address this concern we employ three alternative distance measures that have been suggested in the literature.

Column 1 of Table 3 uses the alternative measure for ownstate distance proposed by Wolf (2000). This measure weights the distance between a state's two largest cities by their population. It thus better reflects heavy concentration of economic activity in relatively small areas. For example, most economic activity in Utah is concentrated around Salt Lake City such that the conventional great circle distance measure could easily overstate actual shipping distances. As expected, on average this alternative measure results in shorter ownstate distances (109 km versus 179 km) so that it reduces the domestic border effect compared to Table 2. In particular, the coefficient on $OWNSTATE_{ij}$ declines from 2.05 (Table 2,

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

[22] Given that the estimates from the different tables (in particular, 1.48 from Column 2 of Table 1 and 2.05 from Column 2 of Table 2) stem from separate regressions, it is of course not possible to carry out a direct test of whether they are statistically different from each other. But although the point estimate of 2.05 is significantly different from the value 1.48 and the point estimate of 1.48 is significantly different from the value 2.05, it is possible to find an intermediate value, say, 1.76 as in Column 3 of Table 1, from which neither 1.48 in Column 2 of Table 1 nor 2.05 in Column 2 of Table 2 are significantly different.

Table 3 Robustness checks

Sample	United States and 50 countries			United States and 50 countries		United States and 50 countries			
Years: 1993, 1997, 2002				Distance > 200 m.	Fewer states	Adjacency	Language	Currency	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$ln(y_i)$	1.21**	1.21**	1.21**	1.22**	1.21**	1.21**	1.21**	1.21**	1.21**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$ln(y_j)$	0.81**	0.82**	0.79**	0.82**	0.81**	0.81**	0.80**	0.81**	0.79**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$ln(dist_{ij})$: Wolf (2000)	−0.80**								
	(0.03)								
$ln(dist_{ij})$: Nitsch (2000)		−0.84**							
		(0.03)							
$ln(dist_{ij})$: Actual shipping distance			−0.76**						
			(0.02)						
$ln(dist_{ij})$				−0.81**	−0.78**	−0.65**	−0.86**	−0.82**	−0.70**
				(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
$adjacency_{ij}$						1.11**			1.01**
						(0.07)			(0.07)
$language_{ij}$							0.88**		0.83**
							(0.06)		(0.06)
$currency_{ij}$								−0.10	−0.12
								(0.09)	(0.08)
$OWNSTATE_{ij}$	1.64**	2.23**	1.50**	2.08**	2.03**	1.48**	1.96**	2.05**	1.44**
	(0.22)	(0.17)	(0.20)	(0.20)	(0.22)	(0.19)	(0.20)	(0.20)	(0.19)
$INTERNATIONAL_{ij}$	−1.13**	−1.06**	−0.24**	−1.10**	−1.21**	−1.28**	−0.31**	−1.19**	−0.62**
	(0.06)	(0.05)	(0.07)	(0.06)	(0.06)	(0.06)	(0.07)	(0.10)	(0.10)

(continued)

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 3 Continued

Sample				United States and 50 countries		United States and 50 countries			
Years: 1993, 1997, 2002	United States and 50 countries			Distance > 200 m.	Fewer states	Adjacency	Language	Currency	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\lvert OWNSTATE_{ij}\rvert=\lvert INTERNATIONAL_{ij}\rvert$	[0.02]	[0.00]	[0.00]	[0.00]	[0.00]	[0.30]	[0.00]	[0.00]	[0.00]
National trade (reference group)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ownstate trade	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
International trade	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12 417	12 417	12 417	12 117	10 368	12 417	12 417	12 417	12 417
Clusters	4139	4139	4139	4039	3456	4139	4139	4139	4139
Random effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.78	0.79	0.79	0.78	0.79	0.79	0.80	0.79	0.80

Notes. The dependent variable is $ln(x_{ij})$. OLS estimation. Robust standard errors are reported in parentheses, clustered around country pairs *ij*. Random effects in all columns. Constants and year dummies are not reported. **Significant at 1% level. The numbers in brackets report *p*-values for the test $\lvert OWNSTATE_{ij}\rvert=\lvert INTERNATIONAL_{ij}\rvert$. Column 4 drops all pairs of US states that are less than 200 miles apart. Column 5 drops states with inferior data quality (AR, DE, FL, NM, TX, VT). Column 6 adds an adjacency dummy that is 1 if two regions have a land border. Column 7 adds a common language dummy. Column 8 adds a common currency dummy. Column 9 combines the adjacency, language, and currency dummies.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Column 2) to 1.64 (Table 3, Column 1). Despite the smaller magnitudes of the domestic border effect, it is still significantly different from the absolute value of the international border estimate in Column 1 of Table 3 (the *p*-value is 0.02). Also note that compared to Column 2 of Table 2, the result for the international border effect in Column 1 of Table 3 is virtually the same (−1.13 compared to −1.10). A similar observation can be made concerning the distance coefficient.

In Column 2 of Table 3, we employ a measure of ownstate distance which is based on land area as in Nitsch (2000). His measure is based on a hypothetical circular economy with three equal-sized cities, one in the center and the other two on opposite sides of the circle. The average internal distance of such an economy, and also other economies with more complex structures, can be approximated by the radius of the circle. In the data, this is computed as $1/\sqrt{\pi} = 0.56$ times the square root of the area in km^2, and on average this results in roughly similar ownstate distances (170 km versus 179 km). Nevertheless, the ownstate dummy estimate increases slightly to 2.23 compared to 2.05 in Column 2 of Table 2.

In Column 3, we employ a third alternative distance measure that is closer to actual shipping distances by ground transportation observed within the United States. Based on private-use Commodity Flow Survey data at the ZIP code level, Hillberry and Hummels (2003, Equation 4 and Table 1) provide a statistical relationship between the distance measure used by Wolf (2000), an ownstate dummy and an adjacency dummy. They estimate the following equation:

$$\begin{aligned}\ln(actual\,dist_{ij}) = \lambda_1 \ln(Wolf\,dist_{ij}) + \lambda_2 OWNSTATE_{ij} \\ + \lambda_3 adjacency_{ij} + e_{ij}\end{aligned} \tag{7}$$

with $\lambda_1 = 0.821$, $\lambda_2 = -0.498$, and $\lambda_3 = -0.404$. We use these coefficients to approximate actual shipping distances within the United States as well as to Canada and Mexico, and we then use them as an explanatory variable. The resulting distances are on average considerably shorter compared to the great circle distances, both within US states (18 km versus 179 km) as well as between US states and to Canada and Mexico (450 km versus 1556 km). The distances to overseas countries are not affected as those routes are not covered by ground transportation. In Column 3 of Table 3, both border coefficients are reduced in magnitude to 1.50 from 2.05 for the ownstate coefficient and to −0.24 from −1.10 for the international coefficient compared to Column 2 of Table 2. But note that the absolute difference between the coefficients remains highly significant.

Results for additional robustness checks are reported in the remaining columns of Table 3. Hillberry and Hummels (2008) document that trade is highly concentrated at the local level and that it consists to some extent of

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

local wholesale shipments. In Column 4, we provide results for trade between locations that are not within immediate proximity to limit the potential influence of wholesale shipments. In particular, we drop all state-to-state observations that are less than 200 miles apart to check whether they distort the sample. This check removes 100 cross-sections from the panel. Nonetheless, the regression results are virtually the same as in Column 2 of Table 2. We obtain similar results if we also drop all within-state observations less than 200 miles apart (not reported here).

As we explain in Section 3.2, Cassey (2009) raises doubts as to whether the Origin of Movement data are sufficiently similar to the actual origin of production in the case of Arkansas, Delaware, Florida, New Mexico, Texas, and Vermont. In Column 5, we drop these six states from our sample. Once again, the regression results are overall quite similar to those in Table 2.

In Column 6, we follow Wolf (2000) by adding an adjacency dummy that takes on the value 1 whenever two states are neighboring (say, Minnesota and Wisconsin).[23] Similar to Wolf (2000), we find that adding an adjacency dummy reduces the ownstate coefficient. Nevertheless, the domestic border effect remains larger in the absolute value than the international border effect. However, in Column 6, we can no longer reject the hypothesis that their absolute values are equal (p-value $= 0.30$).

In Column 7, we control for a common language dummy that takes on the value 1 whenever countries have English as an official language according to the CIA World Factbook. In our sample, these countries are Australia, Canada, Hong Kong, India, Ireland, New Zealand, Singapore, South Africa, and the UK. For all intra-US observations, the common language dummy is also set to take on the value 1. We note that a dummy variable for common legal origin (common law) would be exactly the same in our sample. Thus, it should arguably be interpreted as a broader measure of cultural and political similarity. As typical in the gravity literature, the language dummy is positive and highly significant. Compared to Column 6, its inclusion increases the ownstate coefficient to 1.96, and the international dummy coefficient is considerably reduced in the absolute value to -0.31.

In Column 8, we use a dummy variable for a common currency. It takes on the value 1 whenever one of the foreign countries uses the US dollar as their official currency, or where the local currency is freely exchanged against the US dollar, or where countries tied their currency against the US dollar for at least one of the years of our sample. In our sample these

23 All ownstate observations are defined to also count as adjacent observations in our sample.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

countries are Argentina, Ecuador, El Salvador, Hong Kong, and Panama, and we also include all intra-US observations.[24] However, the common currency dummy turns out insignificant.

Finally, in Column 9, we combine the three additional trade cost regressors from Columns 6–8. The domestic border effect coefficient follows as 1.44, and the international border effect coefficient stands at −0.62. Statistically their absolute values are strongly different from each other (p-value = 0.00). This result shows that once we use a more complete trade cost function that controls for a wider range of trade cost elements, our main finding is corroborated: the domestic border effect appears larger in absolute value than the international border effect.

In Table 4, we carry out a number of additional robustness checks that alter the trade cost function (2). The results in Table 1 are characterized by a larger distance elasticity in absolute value for the domestic border effect regressions than for the international border effect regressions. This suggests that the trade cost function (2), which is log-linear in distance, could be problematic when applied to the pooled sample in Table 2. Instead, it might be more appropriate to use a trade cost function that allows for a larger distance elasticity at relatively short distances (typically associated with domestic border effect regressions) and for a smaller distance elasticity at relatively longer distances (typically associated with international border effect regressions). In Column 1 of Table 4, we adopt such a trade cost function in the form of a double-logarithmic specification for distance.[25] Of course, the distance coefficient now takes on a different value (−5.78 as opposed to a value in the vicinity of −1 as in the previous regressions) but it remains highly significant. The regression retains its explanatory power, yielding an R-squared of 79%. Most importantly, although the coefficient on $OWNSTATE_{ij}$ declines from 2.05 (Table 2, Column 2) to 1.53, it is still larger in absolute value than the $INTERNATIONAL_{ij}$ coefficient. But their difference is no longer statistically significant given the corresponding p-value of 0.27.

For completeness, in Column 2 of Table 4, we consider the opposite case of a trade cost function that implies a smaller distance elasticity at shorter distances. This specification uses the square of logarithmic distance. It results in a larger domestic border effect estimate equal to 2.56 so that the difference to the absolute value of the international border effect estimate becomes significant.

---

[24] The source of this information is available at http://www.gocurrency.com/countries/united_states.htm.

[25] If the trade cost function depends on $\tilde{\delta}\ln[\ln(dist_{ij})]$ instead of $\tilde{\delta}\ln(dist_{ij})$ in Equation (2), then the elasticity of trade costs with respect to distance becomes $d\ln(t_{ij})/d\ln(dist_{ij}) = \tilde{\delta}/\ln(dist_{ij})$. This elasticity is decreasing in distance.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 4 Robustness checks for the functional form of distance

Sample	United States and 50 countries		United States and 50 countries	
Years: 1993, 1997, 2002			Intervals by km	Intervals by obs.
	(1)	(2)	(3)	(4)
ln(y_i) $ln(y_i)$	1.21**	1.21**	1.21**	1.21**
	(0.02)	(0.02)	(0.02)	(0.02)
$ln(y_j)$	0.81**	0.82**	0.81**	0.84**
	(0.01)	(0.01)	(0.01)	(0.01)
$ln[ln(dist_{ij})]$	−5.78**			
	(0.24)			
$[ln(dist_{ij})]^2$		−0.05**		
		(0.00)		
$ln(dist_{ij})$: interval 1			−0.07	−0.76**
			(0.06)	(0.06)
$ln(dist_{ij})$: interval 2			−0.20**	−0.80**
			(0.06)	(0.05)
$ln(dist_{ij})$: interval 3			−0.27**	−0.85**
			(0.05)	(0.05)
$ln(dist_{ij})$: interval 4			−0.35**	−0.88**
			(0.05)	(0.04)
$ln(dist_{ij})$: interval 5			−0.37**	−0.78**
			(0.04)	(0.04)
$OWNSTATE_{ij}$	1.53**	2.56**	2.77**	1.96**
	(0.25)	(0.17)	(0.18)	(0.20)
$INTERNATIONAL_{ij}$	−1.25**	−1.08**	−0.93**	−0.80**
	(0.06)	(0.06)	(0.07)	(0.06)
$\lvert OWNSTATE_{ij}\rvert=\lvert INTERNATIONAL_{ij}\rvert$	[0.27]	[0.00]	[0.00]	[0.00]
National trade (reference group)	Yes	Yes	Yes	Yes
Ownstate trade	Yes	Yes	Yes	Yes
International trade	Yes	Yes	Yes	Yes
Observations	12 417	12 417	12 417	12 417
Clusters	4139	4139	4139	4139
Random effects	Yes	Yes	Yes	Yes
R^2	0.79	0.78	0.80	0.81

Notes. The dependent variable is $ln(x_{ij})$. OLS estimation. Robust standard errors are reported in parentheses, clustered around country pairs *ij*. Random effects in all columns. Constants and year dummies are not reported. **Significant at 1% level. The numbers in brackets report *p*-values for the test $\lvert OWNSTATE_{ij}\rvert=\lvert INTERNATIONAL_{ij}\rvert$. Column 1 uses the logarithm of $ln(dist_{ij})$ as a regressor. Column 2 uses the square of $ln(dist_{ij})$ as a regressor. Column 3 uses five distance intervals delineated by 750 km, 1500 km, 3000 km, and 6000 km (see the text for details). Column 4 uses five distance intervals with an equal number of observations each (see the text for details).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Inspired by Eaton and Kortum (2002), in the remaining columns of Table 4, we distinguish between several distance intervals and allow the distance coefficients to vary over these intervals. This approach represents a more flexible trade cost function. As a reference point, we note that the average distance in the domestic border effect regressions in Columns 2 and 3 of Table 1 is 1485 km with a median of 1284 km, and the average distance in the international border effect regression in Column 5 of Table 1 is 5451 km with a median of 3816 km.[26] We allow for five intervals that are supposed to reflect these different ranges. In particular, in Column 3 of Table 4, the first interval captures all bilateral observations with the shortest distances in the sample of up to 750 km. The second interval captures distances between 750 km and 1500 km, the third interval those between 1500 km and 3000 km, the fourth those between 3000 km and 6000 km, and the fifth interval captures all distances above 6000 km.[27]

It turns out that the first individual distance coefficient is not significant, suggesting that at very short distances trade is hardly sensitive to slightly longer routes. In contrast, Hillberry and Hummels (2008) document a highly nonlinear distance effect, with the distance elasticity falling as distance rises. But this effect applies to extremely short distances. For example, Hillberry and Hummels (2008) show that trade within a single US ZIP code is on average three times higher than trade with partners outside the ZIP code. But the average ZIP code has a median radius of only four miles. Likewise, Llano-Verduras, Minondo, and Requena-Silvente (2011) document a similar relationship at very short distances for geographically finely disaggregated Spanish trade data.[28] However, our sample does not focus on such short ranges. In fact, the average distance in the shortest distance interval in our sample is 439 km and thus substantially higher. The most important aspect of Column 3 for our purposes is that the domestic border effect estimate is significantly larger than that for the international border effect in absolute value. The corresponding coefficients are 2.77 and −0.93.

Finally, we allow for five intervals that contain an equal number of observations. These intervals are delineated by the 1166 km, 2589 km, 6323 km, and 9835 km marks. The results are reported in Column 4 of

[26] As he only considers trade for Canada and the United States, McCallum (1995) compares trade flows over a similar range of distances. Our data set includes US trade with many countries outside North America so that the average distance for international flows is longer.

[27] These intervals capture 1371, 1878, 2148, 1845, and 5175 observations, respectively.

[28] Figure 1 in Hillberry and Hummels (2008) shows that the value of trade drops almost tenfold between 1 and 200 miles, with most of that decline occurring at the first few miles. Llano-Verduras, Minondo, and Requena-Silvente (2011) report sharp reductions in the value of trade for shipments between 25 and 250 km (see their Figures 1 and 2).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 4. It remains the case that the $OWNSTATE_{ij}$ dummy is significantly larger in absolute value than the $INTERNATIONAL_{ij}$ dummy. The values are 1.96 and −0.80, respectively.

Overall, we conclude that although the point estimates of the domestic and international border effects can change depending on the distance measure, the distance function, and the subsample, it is a robust feature of the data that the absolute magnitude of the domestic border effect exceeds that of the international border effect. Their difference is highly significant in almost all specifications.

5 Discussion

We discuss a number of potential explanations for our empirical result that the domestic border effect is comparatively large in magnitude. One major explanation is related to work by Hillberry and Hummels (2008). Based on ZIP-code-level domestic US trade flows, they document that trade within the United States is heavily concentrated at the local level: trade within a single ZIP code is on average three times higher than trade with partners outside the ZIP code. As a major reason, they point out the co-location of producers in supply chains to exploit informational spillovers, to minimize transportation costs, and to facilitate just-in-time production.[29] The local concentration of trade might also be related to external economies of scale in the presence of intermediate goods and associated agglomeration effects (see Rossi-Hansberg, 2005), as well as to hub-and-spoke distribution systems and wholesale shipments (see Hillberry and Hummels, 2003). Such spatial clustering of economic activity can lead to large domestic border barrier estimates, as we find in our results.[30] In that case, the domestic border effect should be interpreted as reflecting the local concentration of economic activity rather than border barriers associated with crossing a state border.

The concentration of trade at the local level is also borne out in other types of data. Using individual transactions data from online auction websites, Hortaçsu, Martínez-Jerez, and Douglas (2009) find that purchases tend to be disproportionately concentrated within a short distance perimeter, with many counterparties based in the same city. Some of these

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

[29] Historically, competition on US state-to-state transportation routes was heavily restricted by the Interstate Commerce Commission well into the post-World War II era, giving companies an additional incentive to co-locate (see Levinson, 2006).

[30] The concentration of trade at the local level might also be related to firms' slicing up their production chains (multistage production and vertical specialization). Yi (2010) offers an explanation of the border effect using the vertical specialization argument in a Ricardian framework.

purchases can be explained by their location-specific nature, for example, in the case of opera tickets. But the evidence also suggests that lack of trust and lack of direct contract enforcement in the case of breach may be major reasons behind the same-city bias, which the authors subsume under 'contracting costs.' They also find evidence for culture and local tastes as factors that shape the local concentration of trade. For example, the same-city effect is most pronounced for local interest items such as sports memorabilia (also see Blum and Goldfarb, 2006).

Business networks and immigration patterns might also be related to strong trade flows between relatively close locations. Combes, Lafourcade, and Mayer (2005) report that business and immigrant networks significantly facilitate trade within France. They cite the reduction of information costs and the diffusion of preferences as two main economic mechanisms through which networks may operate. This includes the reduction of search costs associated with matching buyers and sellers (Rauch and Casella, 2003). As an additional facilitating factor for trade, Rauch and Trindade (2002) also mention the possibility of community sanctions that could be imposed among members of an ethnic network. In the context of the border effect in US data, Millimet and Osang (2007) find that incorporating migration flows within the US diminishes the estimated intranational home bias. Business and immigrant networks therefore likely play an important role in explaining the trade-reducing effect of distance.[31]

6 Conclusion

We collect a data set of US exports that combines three types of trade flows: trade within an individual state (Minnesota–Minnesota), trade between US states (Minnesota–Texas), and trade flows from an individual US state to a foreign country (Minnesota–Canada). This data set allows us to jointly estimate the effect on trade of crossing the domestic state border and the effect of crossing the international border.

While we obtain point estimates consistent with those generally found in the literature, we show that the international border effect is in fact *smaller* than the state border effect. That is, while trading internationally is still the most costly in absolute terms, overcoming the first few miles that are associated with leaving the home state appears harder than crossing the international border once the home state has been left. This result is robust

[31] The impact of ethnic networks on exports from US states has been explored recently by Bandyopadhyay, Coughlin, and Wall (2008). One of their findings is that the inclusion of a common network effect reduces the negative impact of distance on exports.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

to alternative distance measures, alternative functional forms for distance, additional trade cost factors and different subsamples. Our article thus sheds new light on the relative size of border effects as typically estimated in gravity applications. In particular, our finding of a relatively strong domestic border effect can be interpreted as reflecting the concentration of economic activity and trade flows at the local level.

Acknowledgements

The article has benefited greatly from the comments of two anonymous referees and the editor. We are also grateful for comments by Keith Head, John Helliwell, David Jacks, Christopher Meissner, Peter Neary, Krishna Pendakur, Frédéric Robert-Nicoud, Nikolaus Wolf, as well as seminar participants at Simon Fraser University, the Midwest Trade conference at Penn State and the CESifo Measuring Economic Integration conference. We also thank Lesli Ott for outstanding research assistance. Novy gratefully acknowledges research support from the Economic and Social Research Council, Grant RES-000-22-3112, and support as a visiting scholar at the Federal Reserve Bank of St Louis. The views expressed are those of the authors and do not necessarily represent official positions of the Federal Reserve Bank of St Louis or the Federal Reserve System.

References

Anderson, J. and E. van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", *American Economic Review* **93**, 170–192.

Anderson, J. and E. van Wincoop (2004), "Trade Costs", *Journal of Economic Literature* **42**, 691–751.

Bandyopadhyay, S., C. Coughlin and H. Wall (2008), "Ethnic Networks and US Exports", *Review of International Economics* **16**, 199–213.

Blum, M. and A. Goldfarb (2006), "Does the Internet Defy the Law of Gravity?", *Journal of International Economics* **70**, 384–405.

Cassey, A. (2009), "State Export Data: Origin of Movement vs. Origin of Production", *Journal of Economic and Social Measurement* **34**, 241–268.

Cassey, A. (2011), "State Foreign Export Patterns", *Southern Economic Journal* **78**, 308–329.

Chaney, T. (2008), "Distorted Gravity: The Intensive and Extensive Margins of International Trade", *American Economic Review* **98**, 1707–1721.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Chen, N. (2004), ‘‘Intra-National Versus International Trade in the European Union: Why Do National Borders Matter?’’, *Journal of International Economics* **63**, 93–118.

Chen, N. and D. Novy (2011), ‘‘Gravity, Trade Integration, and Heterogeneity Across Industries’’, *Journal of International Economics* **85**, 206–221.

Combes, P., M. Lafourcade and T. Mayer (2005), ‘‘The Trade-Creating Effects of Business and Social Networks: Evidence from France’’, *Journal of International Economics* **66**, 1–29.

Coughlin, C. (2004), ‘‘The Increasing Importance of Proximity for Exports from U.S. States’’, *Federal Reserve Bank of St. Louis Review* **86**, 1–18.

Coughlin, C. and T. Mandelbaum (1991), ‘‘Measuring State Exports: Is There a Better Way?’’, *Federal Reserve Bank of St. Louis Review* **73**, 65–79.

Coughlin, C. and H. Wall (2003), ‘‘NAFTA and the Changing Pattern of State Exports’’, *Papers in Regional Science* **82**, 427–450.

Cronovich, R. and R. Gazel (1999), ‘‘How Reliable Are the MISER Foreign Trade Data?’’, Unpublished manuscript.

De Sousa, J. and S. Poncet (2011), ‘‘How Are Wages Set in Beijing?’’, *Regional Science and Urban Economics* **41**, 9–19.

Eaton, J. and S. Kortum (2002), ‘‘Technology, Geography and Trade’’, *Econometrica* **70**, 1741–1779.

Fally, T., R. Paillacar and C. Terra (2010), ‘‘Economic Geography and Wages in Brazil: Evidence from Micro-Data’’, *Journal of Development Economics* **91**, 155–168.

Head, K. and T. Mayer (2009), ‘‘Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade’’, in P. van Bergeijk and S. Brakman, eds., *The Gravity Model in International Trade: Advances and Applications*, Cambridge University Press, Cambridge, United Kingdom.

Helliwell, J. (1997), ‘‘National Borders, Trade and Migration’’, *Pacific Economic Review* **2**, 165–185.

Helliwell, J. (1998), *‘‘How Much Do National Borders Matter?’’*, Brookings Institution, Washington, DC.

Helliwell, J. and G. Verdier (2001), ‘‘Measuring Internal Trade Distances: A New Method Applied to Estimate Provincial Border Effects in Canada’’, *Canadian Journal of Economics* **34**, 1024–1041.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Hering, L. and S. Poncet (2010), ‘‘Market Access and Individual Wages: Evidence from China’’, *Review of Economics and Statistics* **92**, 145–159.

Hillberry, R. and D. Hummels (2003), ‘‘Intranational Home Bias: Some Explanations’’, *Review of Economics and Statistics* **85**, 1089–1092.

Hillberry, R. and D. Hummels (2008), ‘‘Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data’’, *European Economic Review* **52**, 527–550.

Hortaçsu, A., F. Martínez-Jerez and J. Douglas (2009), ‘‘The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre’’, *American Economic Journal: Microeconomics* **1**, 53–74.

Levinson, M. (2006), ‘‘The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger’’, Princeton University Press, Princeton, NJ, USA.

Llano-Verduras, C., A. Minondo and F. Requena-Silvente (2011), ‘‘Is the Border Effect an Artefact of Geographic Aggregation?’’, *World Economy* **34**, 1771–1787.

McCallum, J. (1995), ‘‘National Borders Matter: Canada-U.S. Regional Trade Patterns’’, *American Economic Review* **85**, 615–623.

Melitz, M. (2003), ‘‘The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity’’, *Econometrica* **71**, 1695–1725.

Melitz, M. and G. Ottaviano (2008), ‘‘Market Size, Trade, and Productivity’’, *Review of Economic Studies* **75**, 295–316.

Millimet, D. and T. Osang (2007), ‘‘Do State Borders Matter for U.S. Intranational Trade? The Role of History and Internal Migration’’, *Canadian Journal of Economics* **40**, 93–126.

Nitsch, V. (2000), ‘‘National Borders and International Trade: Evidence from the European Union’’, *Canadian Journal of Economics* **33**, 1091–1105.

Poncet, S. (2003), ‘‘Measuring Chinese Domestic and International Integration’’, *China Economic Review* **14**, 1–21.

Rauch, J. and A. Casella (2003), ‘‘Overcoming Informational Barriers to International Resource Allocation: Prices and Ties’’, *Economic Journal* **113**, 21–42.

Rauch, J. and V. Trindade (2002), ‘‘Ethic Chinese Networks in International Trade’’, *Review of Economics and Statistics* **84**, 116–130.

Rossi-Hansberg, E. (2005), ‘‘A Spatial Theory of Trade’’, *American Economic Review* **95**, 1464–1491.

Smith, P. (1999), ‘‘Are Weak Patent Rights a Barrier to U.S. Exports?’’, *Journal of International Economics* **48**, 151–177.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Wei, S. (1996), ‘‘Intra-National Versus International Trade: How Stubborn are Nations in Global Integration?’’, NBER Working Paper #5531, Cambridge, MA, USA.

Wolf, H. (2000), ‘‘Intranational Home Bias in Trade’’, *Review of Economics and Statistics* **82**, 555–563.

Wolf, N. (2009), ‘‘Was Germany Ever United? Evidence from Intra- and International Trade 1885–1933’’, *Journal of Economic History* **69**, 846–881.

Yi, K. (2010), ‘‘Can Multistage Production Explain the Home Bias in Trade?’’, *American Economic Review* **100**, 364–393.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

CESifo Economic Studies, Vol. 59, 2/2013, 277–305 doi:10.1093/cesifo/ifs035
Advance Access publication 24 October 2012

Border Effects and European Integration

Angela Cheptea*†

*INRA, UMR1302, F-35000 Rennes, France
†AGROCAMPUS OUEST, UMR1302, F-35000 Rennes, France

Abstract

A new method for measuring trade potential from border effects is developed and applied to manufactured trade between the old 15 European Union (EU) members and 12 Central and East European (CEE) economies. Border effects are estimated with a theoretically compatible trade specification and much larger trade potentials are obtained than usually predicted by standard trade potential models. Even after a decade of regional trade liberalization, the integration of CEE and EU economies is two to three times weaker than intra-EU integration, revealing a large potential for East–West European trade. Adjusting for the impact of borders on multilateral resistance, yields lower trade potentials, but considerably larger than the magnitudes obtained with traditional approaches. (JEL codes: F10; F12; F14; F15)

Keywords: trade potential, regional integration, border effects

1 Introduction

Economic relationships between Central and Eastern European (CEE) countries and their Western partners during the 1990s have been marked by the premises of European Union (EU) enlargement. In the early 1990s most CEE countries have formulated officially their desire to integrate the Union, and have received an affirmative response conditional on the fulfillment of several economic criteria. About a decade later, they have acquired the membership status and benefit from all insiders' advantages. The evolution of their economic exchanges between these two dates reflected a gradual elimination of trade costs, and a concentration of trade with the old EU15 partners. Regional integration between Eastern and Western European nations has been accompanied by important trade creation effects that continue even after CEE countries have joined the EU. Indeed, it takes time for firms to grasp trading opportunities offered by the modified economic environment. The economic literature employs the term *trade potential* to designate these effects.

The additional trade arising from an economic integration initiative is traditionally estimated in the literature by trade potential models that rely on the empirical success of the gravity equation. The essence of these models consists in comparing actual trade to the gravity-predicted or so-called 'normal' level of trade, with the difference between the two capturing the trade potential. Wang and Winters (1992), Hamilton and Winters (1992), Baldwin (1993), Gros and Gonciarz (1996), Fontagne

© The Author 2012. Published by Oxford University Press
on behalf of Ifo Institute for Economic Research, Munich. All rights reserved.
For permissions, please email: journals.permissions@oup.com

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

et al. (1999), Nilsson (2000), and Papazoglou et al. (2006) use this approach to estimate European trade potential during the 1990s. One drawback of this method is the misspecification of the gravity equation used in these models with respect to trade theory, and the sensitiveness of results upon the gravity specification employed. Another weakness of trade potential models is that they disregard the large amount of trade taking place inside national borders and base their predictions on an analysis carried exclusively on international trade.

The present article introduces a new method for measuring trade integration and quantifying future increases in intra-regional trade, inspired from the literature on border effects (McCallum 1995; Wei 1996; Helliwell 1996; Head and Mayer 2000; Nitsch 2000; Wolf 2000; Head and Ries 2001; Chen 2002; Wolf 2009; Coughlin and Novy 2012). Differently from traditional trade potential models, I define the level of trade integration of two or more countries by referring to their domestic trade. The closer is the volume of trade between two countries to their domestic trade, when controlling for standard variables such as supply, demand, and trade costs, the more integrated are the two countries. In other words, I compute trade potentials from all cross-border trade costs, taking into account domestic trade.

Technically, the method consists of two steps. Firstly, I estimate the level of cross-border trade costs using each country's domestic trade as benchmark for its trade with partner countries. The rationale for this is the following: a country is a highly integrated and homogeneous economic space, where full economic integration is achieved. Indeed, in the light of some recent studies (e.g. Brunetti et al. 1997; Rauch 2001), the presence of a single legislative system, central administration, currency, communication network, and set of economic policies contributes to an important reduction of transaction costs and fosters exchange. This argument is confirmed by empirical works revealing that higher volumes of trade take place inside countries (i.e. within national borders) than between them (i.e. across borders). McCallum (1995) refers to this as the *border effect* and finds that even highly integrated countries as Canada and US trade about 20 times less with each other than with themselves. Later work has proven this figure to be unrealistically high: e.g. Anderson and van Wincoop (2003) find a border effect ranging between 2.24 and 10.7 for the same countries. Still, domestic trade remains a convenient benchmark for international trade flows. In this article I make the assumption that trade costs other than those induced by the distance are null for transactions taking place within the same country, and express international trade costs in terms of border effects, i.e. the ratio of international-to-domestic volume of trade.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Secondly, I compare international trade costs for the integrating and the reference group of countries. The group of countries with the lowest level of intra-group trade costs serves as reference for all other regional trade flows. I compute the level of trade integration or trade potential as the ratio of estimated within- and cross-group border effects, with a lower ratio corresponding to a higher level of trade integration. I choose the reference group to be formed by countries with the lowest international trade costs and I assume that further integration within the region reduces trade costs to the level observed for the reference group. In the particular case of European integration, trade between the 15 core-EU members is subject to lower distortions and I use it as a reference for other European flows, as in the literature on trade potentials. The fact that the share of intra-EU trade in total EU trade remained at a steady level during the last two decades suggests that the latter might well correspond to the long-term equilibrium. The East–West European trade creation may or not be accompanied by trade diversion in the detriment of intra-CEE integration. After the EU enlargement in 2004 and 2007 trade between new member states (NMSs) became intra-EU trade, and trade costs associated with these flows should also converge, at least in the long run, to the level of intra-EU costs prior to enlargement.

Another question tackled in this article is that of the correct specification of the gravity equation. Although gravity is shown to be compatible with both traditional and new trade theories, each theoretical model produces a different final trade specification. This aspect, ignored by trade potential models, is incorporated here through the use of a theoretically derived trade equation in the estimation of border effects.

For the simplicity of the exposal I refer hereafter to trade between old/core EU15 countries as *intra-EU trade*, to trade between NMSs that joined the EU in the last decade as *intra-CEE trade*, and to trade between the two groups of countries as *CEE–EU* or *East–West European trade*. Thus, the CEE–EU trade potential or trade integration is obtained as the ratio between the border effect estimated for CEE–EU trade and for intra-EU trade.

Trade of the 12 NMSs, both with each other and with the 15 core-EU countries increased remarkably over the last two decades. The results predict much higher trade potential values for CEE–EU and intra-CEE trade than usually found in the literature with traditional trade potential models. At the beginning of the 21st century trade between CEE and EU countries represented about two-thirds of its attainable level, suggesting a possible 39% increase with further EU integration. Adjusting for the impact of borders on multilateral resistance (MR), yields lower trade potentials, but the latter continue to be above the magnitudes obtained with traditional approaches. The possible upsurge of intra-CEE trade in

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

the following years, despite the impressive reduction of bilateral border effects reached by the beginning of the century, is even higher.

The article is organized as follows. The next section presents the trade model and the trade specification used to estimate border effects. Border effect estimates within and between country groups are presented and discussed in Section 3. In Section 4, I discuss trade potentials for European trade flows produced by the different approaches and their evolution in time. Section 5 summarizes the conclusions.

2 Theoretical Discussions

I start by describing an underlying preference structure for trade in differentiated goods. The obtained trade equation includes variables that are unobserved or inaccurately measured, i.e. is unsuitable for direct estimations. To address this issue I follow Rose and van Wincoop (2001) and Redding and Venables (2004) and use country-specific effects to capture importer and exporter variables.

2.1 A differentiated-goods trade structure

First, I consider a trade structure with a differentiated good and n_i varieties produced in each country i. The model has a slightly different interpretation depending on the data used. Each industry (when using industry-level data) or the entire manufactured sector (when using aggregate data) is considered to be composed of a single differentiated product of which multiple varieties are available. Product differentiation can be at country or firm level. National product differentiation was introduced by Armington (1969) who proposed an utility function in which consumers distinguish products by their origin. It can also arise from a Heckscher–Ohlin model with no factor price equalization as in Deardorff (1998). An alternative approach is that of Dixit–Stiglitz–Krugman (DSK) type monopolistic competition models. In the latter each variety is produced by a distinct firm, and the number of varieties n_i (identical to the number of firms) is endogenously determined by the model.

Consumer preferences are homothetic and represented by a CES utility function. Importing country j's representative consumer utility is given by

$$u_j = \left[\sum_i n_i \left(a_{ij} x_{ij} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{1}$$

with a_{ij} representing country j consumers' preference for country i products, x_{ij} the volume of goods produced in i and consumed in j, and σ the

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

substitution elasticity between any two varieties. Coefficients a_{ij} are introduced in order to allow for different preferences across countries.[1]

I assume that consumers of each product are charged with the same price augmented by trade costs. The difference in the price of the same good in two different locations is therefore entirely explained by the difference in trade costs to these locations. For simplicity an *iceberg* trade costs function is used. The price to country j consumers of a good produced in i, p_{ij}, is the product of its mill price p_i and the corresponding trade cost t_{ij}. Two elements of bilateral trade costs are considered: transport costs proportional to the shipping distance d_{ij}, and costs due to the presence of trade barriers such as tariffs, non-tariff barriers, information costs, partner search costs, institutional costs, etc.:

$$t_{ij} = \underbrace{d_{ij}^{\rho}}_{\text{transport costs}} \underbrace{\exp\left[\left(1 - home_{ij}\right)b_{ij}\right]}_{\text{border-specific costs}}. \tag{2}$$

The second type of costs arises exclusively for trade across national borders. $home_{ij}$, is a dummy variable equal to one for internal trade and to zero for trade between countries. $[\exp(b_{ij}) - 1] \times 100$, gives the tariff equivalent of border-specific trade barriers on country i exports to destination j. In Section 3, I introduce a more complex trade cost function by decomposing the second left-hand side term of equation (2) in order to account for the presence of a common land border or language, and different trade flow types.

Consumers of each country j spend a total amount E_j on domestic and foreign products:

$$\sum_i n_i x_{ij} p_{ij} = E_j, \tag{3}$$

and choose quantities that maximize their utility function (1) under the budget constraint (3). Country j's total demand for country i products is given by

$$m_{ij} \equiv x_{ij} p_{ij} = a_{ij}^{\sigma-1} \left(\frac{p_i t_{ij}}{P_j}\right)^{1-\sigma} n_i E_j, \tag{4}$$

$$\text{where } P_j \equiv \left[\sum_k a_{kj}^{\sigma-1} \left(p_k t_{kj}\right)^{1-\sigma} n_k\right]^{\frac{1}{1-\sigma}} \tag{5}$$

[1] Two forms of preferences are usually found in the literature: identical for all countries, $a_{ij} = a_i \forall j$, yielding symmetric utility functions (e.g. Anderson and van Wincoop 2003), and more pronounced for domestic products, $a_{ij} = \exp(e_{ij})$ if $i \neq j$ and $a_{jj} = \exp(e_{jj} + \beta)$, producing asymmetric demand functions (e.g. Bergstrand 1989; Head and Mayer 2000).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

is a price index of the importing country j nonlinear with respect to the unknown parameter σ. The estimation of equations (4)–(5) is possible only for particular values of the substitution elasticity σ. But even then the presence of a nonlinear price index P_j, and the difficulty of measuring the number of varieties produced in each country limit the accuracy of results. Slightly different specifications are reached with national and firm-level product differentiation.

Consumer preferences can also be expressed as a function of bilateral variables, similar to trade costs. However, there are no means to disentangle the impact of the same variable on preferences from its impact on trade costs. Estimated coefficients on the latter will actually reflect the global effect on both trade costs and consumer preferences. For exposal simplicity I assume throughout the rest of the article identical preferences for all products and consumers and interpret any increase (drop) in the term (t_{ij}/a_{ij}) as a reduction (raise) of trade costs. The main implication of this assumption is that border effect measures will capture the trade gap arising from stronger preferences of consumers for domestic goods, in addition to the effect induced by larger costs for trading across borders.[2] Alternatively, one could consider that an identical equally priced good from source country s is perceived differently by consumers in country i and consumers in country j. A strong (weak) taste for good s leads consumers to overvalue (undervalue) the virtues of the product and shifts their demand function upward (downward). Thus, the actual price to which respond consumers in country j is $a_{sj}^{\sigma-1} p_{sj}$ rather than p_{sj}.

2.2 The fixed-effects specification

An estimable trade specification can be derived directly from (4) by grouping i and j terms of the equation and taking logarithms on both sides:

$$\ln m_{ij} = FE_i + (1 - \sigma) \ln t_{ij} + FM_j, \tag{6}$$

Country fixed effects are used as proxies for supply and demand terms of the equation with $FE_i \equiv \ln(n_i p_i^{1-\sigma}) + \ln(a_{ij}^{\sigma-1})$, and $FM_j \equiv \ln(E_j P_j^{\sigma-1})$, and coefficients of bilateral variables are only estimated. Integrating the expression of trade costs given by identity (2), I reach the trade specification:

$$\ln m_{ij} = FE_i + FM_j + \rho(1 - \sigma) \ln d_{ij} + (1 - \sigma) b_{ij} - (\sigma - 1)\, b_{ij}\, home_{ij} \tag{7}$$

[2] The assumption of identical preferences does not alter the main conclusion of the article. The aim of the article is to illustrate the integration between old and new EU countries over the past two decades. While differences in consumer preferences may inflate the level of border effects estimated for each year, they leave unaffected the evolution trend since changes in tastes and consumption habits arise on much longer time horizons than the one considered in the article.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

A larger coefficient on the last variable designates higher cross-border barriers for country i's exports to j. Thus, higher barriers to international trade can arise not only from larger trade costs (larger b_{ij}), but also from a higher elasticity of substitution σ. This means that even very small trade barriers may generate important deviations of trade toward the domestic market when the substitution elasticity is sufficiently high. Note that one cannot identify both b_{ij} and σ in (7). However, this does not represent a major drawback, since I am interested only in the estimation of the overall border-specific effect: $(\sigma - 1)b_{ij}$.

A convenient advantage of this approach is that it relies uniquely on the differentiated-good structure, without introducing additional assumptions regarding the market structure or the production process. Thus, it is equally compatible with constant and increasing returns to scale, national, and firm-level differentiation of products.

Differently, one can first derive a gravity-type trade equation following Anderson and van Wincoop (2003)'s approach for national product differentiation, and only afterwards group supply and demand variables separately into country-specific effects. This will produce identical estimation equations and results; the difference lays in the interpretation of country and partner effects FE_i and FM_j.

Summing bilateral imports (4) across destinations gives the production level at origin Y_i. Then the obtained identity can be further used to express the unknown amount $p_i^{1-\sigma}$ ($n_i = 1, \forall i$ in this particular case), which is then re-introduced in trade equation (4). Unlike Anderson and van Wincoop (2003) this can be accomplished without imposing market clearance ($Y_i = E_i$) and using data on importer's expenditure.[3] A nice gravity equation is then obtained:[4]

$$m_{ij} = \frac{Y_i E_j \phi_{ij}}{\bar{P}_i^{1-\sigma} \tilde{P}_j^{1-\sigma}} \tag{8}$$

$$\text{with } \bar{P}_i^{1-\sigma} \equiv \sum_k \phi_{ik} P_k^{\sigma-1} E_k, \text{ and } \tilde{P}_j^{1-\sigma} = \sum_k p_k^{1-\sigma} \phi_{kj}. \tag{9}$$

[3] Market clearance is a quite restrictive assumption as it implies balanced international trade, which occurs only at national level and in the long run. This assumption is inconsistent with the CEE–EU industry-level pattern of trade. In 2000, 80% of the trade between EU and CEE countries at the industry level was intra-industry trade. Trade imbalances are less important for the entire manufactured sector, but not sufficiently low to suggest that realistic predictions shall be obtained by assuming market clearance at the aggregate level. Therefore, it is preferable to use industry-level expenditures computed as the sum of domestic production and foreign imports.

[4] Deardorff (1998) reaches a similar trade equation from a Heckscher–Ohlin trade model with differences in factor prices across countries and complete specialization.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

$\tilde{P}_j$ is an importer-specific price index reflecting the average price of country j's imports. A higher average price paid by consumers of the importing country increases the value of exports to that market. $\tilde{P}_j^{1-\sigma}$, on the contrary, corresponds to the relative isolation of a country in terms of trade costs and/or consumer preferences, and reduces bilateral flows. $\bar{P}_i$ is an exporter-specific weighted average of price indexes of all its trading partners including itself. The expression of $\bar{P}_i^{1-\sigma}$ in (9) is very similar to the remote market access used in economic geography models: the access of country i's products to all markets, including the domestic one. In other words, $\bar{P}_i$ reflects the purchasing power of i's partners and is positively related to trade. An improved global market access for country i products translates into higher total shipments to its partners. Symmetric trade costs $(t_{ij} = t_{ji}, \forall i, j)$, and identical preferences across countries $(a_{ij} = a_i, \forall i, j)$ yield the symmetric solution $\bar{P}_i = \tilde{P}_i$ used by Anderson and van Wincoop (2003) to reach a more elegant version of (8). Rewriting equation (8) in logarithmic form and using country and partner binary variables to capture demand and supply terms[5], yields equation (6).

Country and partner fixed effects capture not only demand, supply, and price (remoteness) effects, but also any unilateral origin- and destination-specific component of trade costs (e.g. strong non-tariff barriers, poor domestic institutions), which might result in an underestimation of actual border effects. However, this method offers a fair evaluation of bilateral components of trade costs, the ones most easily to be reduced through regional integration.

3 Estimating Border Effects Across Europe

The method proposed in this article computes trade potentials from border effects within and between country groups. This section is dedicated to the estimation of border effects.

3.1 The empirical model

I divide trade between European countries into three types: CEE–EU trade, intra-EU trade, and intra-CEE trade, and estimate border effects for each type of flows. I allow for differences in international trade costs across the three types of trade and for countries sharing a land border or speaking the same language. Initially I also assume that barriers to

[5] $FE_{\mathrm{i}} \equiv \ln\left(Y_i \bar{P}_i^{\sigma-1}\right)$ and $FM_j \equiv \ln(E_j \sum_k p_k^{1-\sigma 1} \phi_{kj})$

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

trade within national borders are the same in all countries, but relax this constraint later. For that I decompose the last term of equation (2) as follows:

$$\begin{aligned}\ln t_{ij} = \delta \ln d_{ij} + b_0 home_{ij} + b_1 CEE\_EU_{ij} + b_2 intraEU_{ij} \\ + b_3 intraCEE_{ij} + c_1 contig_{ij} + c_2 comlang_{ij}\end{aligned} \tag{10}$$

As previously, $home_{ij}$ stands for domestic trade and $b_0 < 0$. Using a single dummy for domestic trade, I implicitly assume that intranational trade costs are the same for EU and CEE country groups. This assumption is addressed in detail later in the section. Dummies CEE_EU_{ij}, $intraEU_{ij}$, and $intraCEE_{ij}$ indicate the affiliation of each observation to a particular type of trade. Variables $contig_{ij}$ and $comlang_{ij}$ denote, respectively, a common land border and language for countries i and j. As both linguistic and neighbor relations are likely to reduce trade costs, I expect coefficients c_1 and c_2 to be negative.

The trade equation to be estimated is obtained by integrating the trade costs function (10) in equation (6). Observe that the first four dichotomic variables in (10) sum to unity, just like the full sets of country and partner fixed effects. Therefore, using all these variables together in the same equation does not permit the identification of all parameters. The inclusion of all country-specific effects is imperative for the estimation of average effects for the entire sample, not relative to an excluded country pair. I choose to drop the variable $home_{ij}$ and use domestic trade as reference for the estimation of coefficients b_1 through b_3. Thus, the constant term reflects the level of domestic trade and other trade flows are expressed as deviations from this level:

$$\begin{aligned}\ln m_{ij} = FE_i + FM_j + \alpha \ \ln \ d_{ij} + \beta_0 + \beta_1 CEE\_EU_{ij} + \beta_2 intraEU_{ij} \\ + \beta_3 intraCEE_{ij} + \gamma_1 contig_{ij} + \gamma_2 comlang_{ij} + \varepsilon_{ij}.\end{aligned} \tag{11}$$

The border effect for each type of trade is obtained by taking the exponential of the absolute value of the corresponding coefficient. For example, $\exp(-\beta_1)$ shows how much more on average an EU15 member state buys from itself compared to purchases from other EU15 countries.

Notice that the use of country and group dummies in equation (11) does not permit to differentiate between EU15 exports toward NMSs and trade in the opposite direction. This is due to the fact that indicator variables for these two types of flow are simply a linear combination of other dummies already included in the equation. Thus, CEE–EU border effects estimated below measure average international trade costs in either direction.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

3.2 Data and descriptive statistics

The present study carries over a sample of 27 countries: 15 core-EU members with Belgium and Luxembourg aggregated under a single observation, and 12 CEE countries. Trade between and within these countries is investigated over a 14-year period: from 1994 to 2007. Of the 12 CEE countries in the panel 10 have joined the EU in May 2004 and 2 in January 2007. Two levels of aggregation are considered: total manufacturing industry, and 26 product industries according to the ISIC Rev.2 classification.

Data on bilateral unidirectional trade are obtained from the BACI database of Cepii. An important challenge for estimating equation (11) is the absence of internationally comparable data on intranational trade flows. To overcome these difficulties, I follow Wei (1996) and other empirical works studying border effects and compute domestic trade volumes as the difference between national production and total exports.

Industry-level productions used in the computation of intranational trade are from Eurostat; missing data are complemented with data from the Trade and Production database of UNIDO (World Bank). In order to ensure the compatibility of different data sources, data have been adjusted by applying a conversion rate equal to the average ratio of the value from the base source and the value from the secondary source, and estimated separately for each country on observations present in both databases.

On average two countries from our panel exchange annually USD 2.5 million worth of manufactured products. The bulk of international trade within Europe (77–90% depending on the year) comes from trade between EU15 countries, with an average annual bilateral trade flow of USD 7.5 million. Meanwhile, two CEE countries exchange on average manufactured products for a value of USD 205 thousand only. The average bilateral flow for intra-CEE trade amounts to USD 765 thousand. The intensification of intra-CEE and CEE–EU trade integration over the last two decades led to a 7-fold and, respectively, a 5-fold increase in these flows (expressed in constant terms) from 1994 to 2007, in contrast to a modest 89% increase in intra-EU trade over the same period.

Still, most of the trade arises within national borders. The ratio of domestic-to-international trade ranges from 2.7 in 1994 to 1.6 in 2007 and testifies of a decline in cross-border trade costs within Europe. This evolution is even more prominent for the group of CEE countries: from 12.4 in 1994 to 4.3 in 2007. Domestic trade for the average country in the panel represents USD 136 million. Again, the difference between the two country groups is very pronounced. Trade in manufactured goods between buyers and sellers from the same EU15 country amounts on average to USD 230 million, but is 10 times smaller for a NMS.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Bilateral distances are geodesic distances between capital cities. The use of domestic trade in estimations implies the necessity to compute intranational distances d_{jj}. The latter are obtained as a weighted average of inter-regional distances. Both measures, as well as data on contiguity and common language, are from the Cepii distance database.

3.3 Border-effect estimates

Estimates of equation (11) for industry-level bilateral imports of EU15 countries and NMSs with country, partner, and year fixed effects are reported in Table 1. I employ the Poisson Pseudo Maximum Likelihood (PPML) estimator suggested by Santos Silva and Tenreyro (2006), which yields the most robust point estimates for group variables across different panels of countries. In addition, this technique permits to control for the presence of heteroscedasticity and nil trade flows.[6] Standard deviations are obtained with a robust clustering technique that allows error terms for the same country pair to be correlated. This permits to control at least partially for autocorrelation in the data.[7]

First, I estimate equation (11) with international trade flows only and display results in the first column. All group dummies cannot be included on the right-hand side of the equation due to collinearity. I choose to drop the one corresponding to intra-EU trade and express the rest of the flows with reference to the average level of trade between two EU15 countries. Group dummy coefficients give then estimates of international trade costs for different types of flow relative to intra-EU costs. Negative values for CEE–EU and intra-CEE trade indicate that intra-EU trade is subject to lower trade costs, justifying its use as reference for other European trade flows. According to results in the first column a EU15 country trades on average 89% $[= (1 - \exp(-2.18)) \times 100]$ less with a NMS than with another EU15 country, while two NMS trade 98% $[= (1 - \exp(-3.86)) \times 100]$ less than two EU15 countries equally large and distant. According to these results, the gap between East–West European and

6 When trade is broken down by industries, an important number of zero value trade flows is observed. The problem with nil trade flows is that they do not occur randomly, but are the outcome of a selection procedure, e.g. a low supply or demand for a particular product. Therefore, apart the PPML estimator I have also employed the two-step Heckman estimator and its maximum-likelihood (Tobit) version, which corrects for this sample self-selection bias by giving a positive weight to the zero trade mass. The independent variables of the selection equation are the same as for the trade-level equation. A statistically significant coefficient of Mills' ratio in the second stage is obtained in all estimations, indicating the necessity of this adjustment. For brevity, estimates are not displayed in the article but can be obtained upon request.

7 Adding industry-level fixed effects in estimations only slightly amends the value of estimated parameters, but convergence is not always achieved. Therefore, these estimations are not presented in the article.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 1 European trade integration: industry-level imports

	(1) All int'l flows	(2) All flows	(3) CEE–EU and all domestic	(4) Intra-EU and all domestic	(5) Intra-CEE and all domestic
ln distance	−0.37[a]	−0.55[a]	−0.70[a]	−0.52[a]	−0.98[a]
	(0.07)	(0.05)	(0.09)	(0.06)	(0.15)
Common land frontier	0.44[a]	0.38[a]	0.82[a]	0.29[a]	0.73[a]
	(0.08)	(0.08)	(0.16)	(0.08)	(0.24)
Common language	0.37[b]	0.40[a]	1.05[a]	0.48[a]	
	(0.17)	(0.13)	(0.16)	(0.12)	
CEE–EU	−2.18[a]	−2.94[a]	−2.77[a]		
	(0.26)	(0.13)	(0.23)		
Intra-EU		−2.41[a]		−2.43[a]	
		(0.11)		(0.12)	
Intra-CEE	−3.86[a]	−3.12[a]			−2.75[a]
	(0.52)	(0.16)			(0.34)
Reference flows	Intra-EU	All domestic	All domestic	All domestic	All domestic
Number of observations	236 600	243 634	129 338	73 282	55 082
Share of nil flows (%)	6.7	6.5	5.6	0.6	14.9

Note: The explained variable in all columns are bilateral imports at industry-level according to the ISIC Rev. 2 classification of manufactured products. Estimations are obtained with the Poisson PML technique and importer, exporter and year fixed effects. Standard errors in parentheses: [a], [b], and [c] represent statistical significance at the 1, 5, and 10% levels, respectively.

intra-EU trade is very prominent and intra-CEE trade integration lies below the level reached by the EU15 members or between EU15 and NMSs.

The second column shows estimates for the entire sample of international and intranational trade flows. All coefficients have the expected sign and are statistically significant. Importer and exporter effects are highly significant. The distance elasticity of trade is lower (in absolute terms) than values found in most empirical studies, but this is not surprising for countries located within the same geographic area. Countries sharing a common border or speaking the same language trade more with each other. This arises not only from lower trade costs between these countries, but also from more similar consumer preferences. The parameters of

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

interest are the coefficients on group (trade type) dichotomic variables. Setting all group variables equal to zero yields an estimation of domestic trade. Trade costs for each type of international trade flows are obtained relative to this reference level. According to results in column (2), a EU15 member country buys on average about 11.2 [= exp(2.41)] times more from itself than from another EU15 country and 19.0 [= exp(2.94)] times more than from a NMS. Similarly, a NMS buys about 22.6 [= exp(3.12)] times more domestically than from another NMS. Thus, trade between EU15 and NMS countries is 1.7 times less intense than between two EU15 countries comparable in terms of market and supply capacities, linguistic, and geographical proximity.

The next three columns replicate this estimation on the subsamples of CEE–EU, intra-EU, and intra-CEE flows. Note that coefficients of group variables are very close to the ones obtained for the entire sample in column (2), making border-effect estimates comparable across the three subsamples. The common language dummy is dropped in column (5) because none of two countries from the CEE group share this characteristic.

In Table 1 intranational trade of both EU and CEE countries serve as reference for cross-border flows, thereby assuming that intranational barriers, except for distance, are equal across the two groups of countries. This assumption is tested by replicating regressions in the last two columns but excluding extra-group domestic flows, and finds confirmation in the data.[8]

Table 2 summarizes border effects for each type of trade flow. Each column corresponds to the last four sub-panels employed in estimations

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

[8] To test whether intranational barriers are equal for CEE and EU countries one can simply split the $home_{ij}$ dummy in two, one for each group: homeEU_{ij} and $\mathrm{homeCEE}_{ij}$. One of the two variables will drop due to collinearity; a significant coefficient on the remaining variable will indicate that EU and CEE countries face different levels of intranational costs. Unfortunately, this technique cannot be applied to the current framework because homeEU_{ij} and $\mathrm{homeCEE}_{ij}$ can each be expressed as a linear combination of country and partner fixed effects. However, one does not have this problem when focusing on intra-EU or intra-CEE trade alone. A test of the equal intranational costs assumption is obtained by estimating equation (11) once on intra-EU (intra-CEE) trade and all domestic trade and once on intra-EU (intra-CEE) trade and EU–CEE domestic trade only. A difference in the estimates of group dummy effect would indicate that the assumption is not verified. I find virtually identical estimates for all parameters, implying that our initial assumption is confirmed empirically. Using the OLS, Heckman or Tobit estimator rather than the PPML method in Table 1 leads to the same conclusion. Less convenient, OLS estimates yield large differences in group dummy coefficients when equation (11) is estimated on the entire sample or on subsamples, making coefficients on group dummy less comparable across subsamples. Results can be provided upon request.

Table 2 European trade integration: border effects with industry-level data

	(2)	(3)	(4)	(5)
Border effects for country pairs that do not share a common land border and do not speak the same language				
CEE–EU	19.0	16.0		
Intra-EU	11.2		11.3	
Intra-CEE	22.6			15.7
Reference flows	All domestic	All domestic	All domestic	All domestic
Border effects for country pairs that share a common land border and speak the same language				
CEE–EU	8.7	2.5		
Intra-EU	5.1		5.2	
Intra-CEE	10.3			7.5
Reference flows	All domestic	All domestic	All domestic	All domestic

Note: Border effects are computed using estimated coefficients of equation (11) for each year with industry-level data. Columns (2)–(5) correspond to the last four columns in Table 1.

in Table 1. For EU15 countries with no common border or language domestic trade exceeds cross-border trade by a factor of 11. For similar NMS countries, this factor ranges between 16 and 23. On average, domestic trade of EU15 and NMS countries is 16 to 19 times larger than trade between the two groups of countries in either direction. This ratio of domestic-to-foreign trade drops by more than half for neighbor countries that speak the same language.

Similar but slightly lower border-effect estimates are obtained with data aggregated for the entire manufactured sector.[9] This outcome testifies that most European trade liberalization was concentrated in a small number of large size industries. The use of aggregate manufacturing data underestimates the amount of 'missing' international trade because it disproportionately reflects large sectors with low barriers to trade. Although the number of nil flows is much lower in this case (7 out of 9441 observations for the entire sample), differences between estimates with the PPML estimator and OLS, Heckman and Tobit estimators persist. This reveals the impact of heteroscedasticity in the data on final results, as demonstrated by Santos Silva and Tenreyro (2006).

[9] Estimates can be provided upon request.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

While trade integration between EU15 countries is significantly deeper than between NMSs and across the two groups of countries, the gap between the latter two is less important. These findings suggest that there is place for major reductions in international trade costs across Europe.[10] The next section addresses the computation of trade potentials, the central issue of this article.

4 Trade Potential and East–West European Integration

The important steps undertaken by Eastern and Western European countries for the removal of politically imposed distortions on bilateral exchanges at the beginning of 1990s, as well as efforts engaged with the scheduled EU enlargement translated into a continuous increase in trade between these countries. The drop in European cross-border trade costs is well pictured by the evolution on regional border effects. Figure 1 shows that border effects for both CEE–EU and intra-EU trade reduced considerably from 1994 to 2007. Differently, intra-EU trade costs declined only little throughout the period. By 2007 intra-EU trade was still about 9 times more expensive than domestic trade. This ratio dropped from over 40 to around 13 for intra-CEE and CEE–EU trade costs. This evolution shows that the reduction of trade costs continued even after 10 of these CEE countries integrated the EU in 2004.

While strengthening trade between old and new members, EU enlargement affected as well trade between NMSs. According to the literature, the reintegration of CEE countries into the world economy in the early 1990s was accompanied by their disengagement from intra-CEE integration.[11] The decline of trade with other CEE partners was beyond its normal level, pointing out the strong competition between former socialist economies for obtaining a higher share of the larger and more attractive core-EU market, and for increasing their chances for accession. With most of CEE countries joining the union, this rivalry has been significantly reduced, and intra-CEE trade has regained attraction. Indeed, as shown in Figure 1, intra-CEE border effects dropped by more than 30 points from the mid-90s until 2007.

---

[10] In a previous version of the article I have also estimated border effects and trade potentials employing alternative theoretical specifications of bilateral unidirectional trade, in particular the *odds* and *friction* specifications according to the terminology employed by Combes et al. (2005). Similar conclusions are obtained, although the magnitude of estimated border effects and trade potentials differ. For details please consult http://www6.rennes.inra.fr/smart/Media/Working-papers/WP10-15.

[11] Baldwin (1993), Gros and Gonciarz (1996), Maurel and Cheikbossian (1998), Nilsson (2000).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

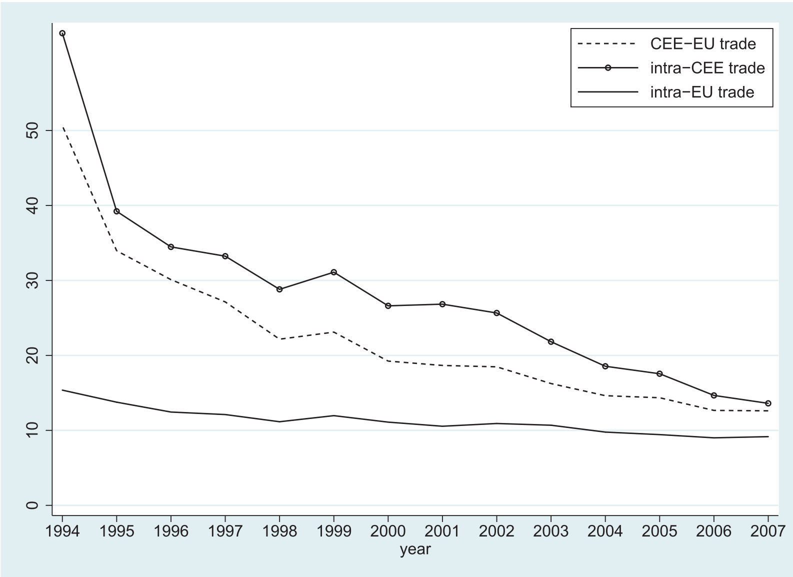

Figure 1 European trade integration: border effects.
Note: Border effects are computed using coefficients of equation (11) estimated with PPML for each year with industry-level data. Effects for countries with no common land border or language are represented.

4.1 An alternative measure of trade potentials

Traditionally trade potentials are obtained in the literature as the difference between actual trade and the gravity prediction (normal level of trade). In this subsection I propose an alternative method that computes trade potentials as the ratio of border effects for investigated and reference trade flows.

With the EU enlargement to the East, the convergence of countries from Central and Eastern Europe toward the EU15 market is expected to take place in all areas of economic activity. It is thus reasonable to assume the share of domestic relative to foreign purchases of NMSs will approach that of the EU15. Indeed, intra-EU trade integration remained almost unchanged (Figure 1) over the last two decades, advocating its use as reference for other regional trade flows. In other words, I assume that in the long run both CEE–EU and intra-CEE trade integration will reach the intra-EU level. Accordingly, I compute the level of trade integration across Europe and the expected further increase in these flows (trade potentials) by comparing the trade costs associated with each trade type to intra-EU costs. I define the potential of CEE–EU and intra-EU

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

trade as the ratio of corresponding border effects estimated with equation (11), and I subtract one in order to express potentials in terms of actual trade:

$$\text{CEE}-\text{EU potential} = \frac{\text{CEE}-\text{EU border effect}}{\text{CEE}-\text{EU border effect}} - 1 = \frac{\exp(-\beta_1)}{\exp(-\beta_2)} - 1 \quad (12)$$

$$\text{intra-CEE potential} = \frac{\text{intra-CEE border effect}}{\text{intra-CEE border effect}} - 1 = \frac{\exp(-\beta_3)}{\exp(-\beta_2)} - 1 \quad (13)$$

Trade potentials obtained in this way reflect a trade integration in terms of border effects. The latter can be interpreted as a measure of trade costs for cross-border flows relative to domestic flows. If these costs are smaller for the reference group, the ratio of border effects for investigated and reference trade flows in (12) and (13) is >1, yielding a positive trade potential. Enhanced CEE–EU (intra-CEE) trade integration due to a drop in relative trade costs for the corresponding flows lowers the border-effects ratio and the trade potential. The integration peak (the maximum trade level and a trade potential of zero) is reached when relative trade costs become equal to those for reference flows. I compute trade potentials using equations (12) and (13) and border effect estimates obtained in Section 3.

For comparison, I also compute trade potentials using the traditional methods employed in the literature. Again, trade flows between EU15 countries are used as reference. Unlike the border-effect-ratio method presented above, traditional trade potential models rely exclusively on cross-border flows. Most of this literature employs traditional gravity equations. Such an equation can be obtained by replacing country and partner fixed effects in (11) with industry-level production of the exporter and consumption for the importer:

$$\begin{aligned} \ln\ m_{ij} = {} & \alpha_0 + \alpha_1 prod_i + \alpha_2 cons_j + \alpha_3 d_{ij} + \beta_1 CEE\_EU_{ij} + \beta_2 intraEU_{ij} \\ & + \beta_3 intraCEE_{ij} + \gamma_1 contig_{ij} + \gamma_2 comlang_{ij} + \epsilon_{ij} \end{aligned} \quad (14)$$

When estimating equation (11) on international trade flows, the three group dummies sum to unity and I choose to drop the one corresponding to intra-EU trade, i.e. the reference group.[12] The difference between trade values observed in the data and the ones predicted with equation (11) for

[12] Industry-level expenditures are computed as the sum of demand for domestic goods and imports from all trading partners.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 3 European trade potentials

	CEE–EU trade		Intra-CEE trade	
	1994	2007	1994	2007
Traditional trade potential models[a] (in percentage of actual trade)				
Standard (atheoretical) gravity estimated on international trade flows				
– In the entire sample	60	17	73	47
– In the intra-EU sub-sample	60	10	88	45
Less traditional trade potential models[a] (in percentage of actual trade)				
The fixed-effects specification estimated				
– On international intra-EU trade	91	19	168	92
– On international and domestic intra-EU trade	97	−14	219	90
Border-effects-ratio method[b] (in percentage of trade relative to EU15)				
Border effects estimated with				
– Standard (atheoretical) gravity	205	59	197	76
– *Fixed-effects* specification	225	39	302	51

Note: Trade potentials are computed with industry-level data.

[a]Trade potentials computed as the difference between actual and predicted trade.

[b]Trade potentials computed according to equations (12) and (13).

each type of flows is attributed to the trade potential. In line with the existing literature,[13] I also estimate equation (14) on trade flows between reference group countries (intra-EU trade) and use obtained coefficients along with data on production, consumption, bilateral distance, and bilateral linkages (common language and land border) to predict the *normal* level of trade for the rest of flows. The difference between actual and predicted trade levels gives the potential of trade. Results with both methods for the first and last year in the panel are displayed in the upper part of Table 3.

The difference between the traditional approach for computing trade potentials and the one developed in this article is threefold. First, the new approach integrates trade taking place within country borders. Secondly, it relies on a measure of trade costs in terms of border effects rather than as the difference between actual and predicted levels. Finally, the trade

[13] Wang and Winters (1992), Hamilton and Winters (1992), Baldwin (1993), Gros and Gonciarz (1996), Fontagne et al. (1999), and Nilsson (2000).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

equation employed in estimations is derived from a theoretical trade model. To disentangle the impact of each of these aspects on the magnitude of trade potentials, in the middle part of Table 3 trade potentials are computed according to the traditional method, but using equation (11) instead of standard gravity, and adding intranational flows. More precisely, I estimate (11) on intra-EU trade without domestic flows, predict values for CEE–EU and intra-CEE trade and obtain trade potentials as the difference between predicted and observed levels. Then, I repeat the exercise including intranational EU trade in initial estimations.

Trade potentials obtained with the innovative approach introduced in this article are shown in the lower part of the table. Similarly, the computation of trade potentials as a border-effect ratio can be applied using estimates obtained with specification (11) or with an atheoretical gravity equation. Results corresponding to these two trade potentials are displayed in the last two rows of the table.

A first conclusion that stems from Table 3 is that traditional methods employed in the literature yield small trade potentials. For all types of trade flow these values are considerably lower than trade potentials obtained with the border-effects-ratio method. This suggests that traditional methods might overestimate the level of trade integration in the region. According to the traditional approach found in the literature, CEE–EU trade in 1994 represented 40% of the level of intra-EU trade for comparable countries, corresponding to a trade potential of 60%. Using an equation derived from a theoretical model to obtain predicted trade values increases the trade potential by half (up to 91%). Including intranational trade flows in reference group estimations brings an additional 6 percentage points. At the same time the ratio of border effects produces a trade potential of over 200%. Differences between trade potentials for intra-CEE trade with the different approaches are even more stringent and follow the same pattern.

All methods show a significant deepening of the CEE–EU and the intra-CEE trade integration from 1994 to 2007, but evaluate very differently the potential for further improvement. Traditional models suggest that trade between EU15 and NMSs has almost reached its potential.[14] In particular, when intranational trade flows are taken into account, a negative trade potential is obtained, suggesting that 2007 trade volumes have outpassed the predicted levels. This outcome testifies that the trade levels predicted

[14] When GDP and population data are used instead of industry-level production and consumption in equation (14), a simplification frequently adopted in the traditional literature, trade potentials predicted by traditional models are even lower (results not displayed).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

by these models are not the good reference and that an alternative is needed. Although traditional methods yield lower potentials for trade between NMSs, they are considerably higher than for the CEE–EU trade. Notice that using a specification based on a theoretical trade model is also important. Indeed, trade potentials obtained with the *fixed-effects* specification are considerably larger than the ones obtained with simple gravity. Omitting intranational trade from estimations alters results mostly at the beginning of the period, i.e. for low levels or integration. Larger trade potentials obtained when intranational trade flows are included in estimations confirm that traditional methods might tend to underestimate the normal (reference) level of trade.

The new method for measuring trade potentials introduced earlier situates the CEE–EU trade potential in 2007 at 39%. This figure was five times larger in the mid 1990s, indicating that by 2007 trade between the two groups of countries regained a large share of its potential. According to this approach NMS countries trade very little with each other in the early 1990s. In 1994 intra-CEE trade amounted to 25% $[=100-302/(302+100)\times 100]$ of its potential level, but increased significantly since then. This reflects the drastic reorientation of foreign trade of these countries in the years following the collapse of the socialist system. Advances in the process of transition and the development of regional economic agreements (CEFTA, the Free Trade Agreement of Baltic states) are encouraged regional trade, which tripled in terms of its potential.

Table A1 of the Appendix exhibits industry-level effects of European integration on trade.[15] Border effects for trade in each industry are estimated with the *fixed-effects* specification and trade potentials are obtained according to equations (12) and (13). The first thing to notice is that with a few exceptions trade creation effects are observed for all industries for both CEE–EU and intra-CEE trade. The largest trade creation for both two-way East–West European trade and intra-CEE trade was achieved for rubber products. Beverages, pottery, china and glassware, metals, and machinery also enjoyed important trade creation. The lowest trade integration is found in the tobacco industry, subject to specific domestic regulations, especially in EU15 countries. For footwear products European trade has even lost some of its potential. This can be explained by the increased competition in these industries with products from emerging Asian countries and in particular China. Moderate effects on trade are obtained for the rest of industries. By the year 2007 CEE–EU trade remains largely inferior to its potential (less than one-third) only in four

[15] The term European integration is used for all 26 European countries considered in this article. This is different from its wide but inaccurate use in the literature to designate only EU integration.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

industries: beverages, tobacco, chemicals, and footwear. Intra-CEE integration was also low in these industries.

Negative 2007 trade potentials in a few industries indicate that for these products CEE–EU, respectively intra-CEE, trade costs declined beyond the level of intra-EU costs. For example, in 2007 trade in metal products and electric machinery was subject to lower costs between a NMS and a EU15 country and between two NMSs than when it involved two EU15 countries.

The reduction of trade barriers and trade potentials for CEE–EU trade coincided with an even more impressive evolution for trade between NMSs. These results disseminate the fears formulated by politicians and some authors that CEE–EU trade integration will be accompanied by a lower commitment of CEE countries to regional integration, reflected by larger intra-CEE border effects and trade potentials at the beginning of the period. Still, according to Table 3 manufactured trade between CEE countries may increase by half in the following years. In selected industries intra-CEE trade may even expand to as much as two to three times the actual volume.

4.2 Advantages and limits of the border-effect-ratio measure

The large difference in trade potentials between the upper and lower part of Table 3 comes from the use of different criteria for evaluating trade integration. Traditional trade potential models ignore domestic trade and assign *normal* trade to the prediction of the gravity equation. The method introduced in the last subsection compares directly trade costs arising in CEE–EU and intra-CEE transactions to costs existing between EU15 trade partners. Trade within the domestic market is used as benchmark for the very estimation of these costs. Therefore this method accounts for the discrepancy between domestic and cross-border trade integration. It is important to signal that not all ‘missing’ international trade is attributed to the trade potential, but only the proportion which corresponds to the difference in trade impediments for specific types of flow. Regional integration is evaluated here in terms of trade costs, expected to converge to the lower intra-EU level. This uniformization of costs will result in increased trade with more distant partners and weaker concentration of trade in the immediate neighborhood.

Larger potentials obtained with the new method confirm the necessity to account for domestic trade in predicting the trade creation effects of regional integration. The disregard of internal trade opportunities is likely to largely underestimate trade potentials. Our method has the advantage of accounting for total international barriers to trade and therefore produces results more in compliance with integration efforts

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

made by countries. Although the access of CEE goods to the EU15 markets improved considerably from 1994 to 2007 and a large part of the potential European trade creation was already accomplished, by the year 2007 the left CEE–EU trade potential was still very large, implying that an important deepening of trade integration within the region is possible in the years to follow.

The implementation of the new approach for measuring trade potentials is subject to several limits. The magnitude of border effects estimated with equation (11) shows only the partial effect of borders on trade, i.e. on the ratio of intranational to international trade, but not the effect on international trade of removing the borders. Anderson and van Wincoop (2003) show that the former is influenced by the relative size differences between countries through multilateral resistance (MR). In particular, a uniform decrease in trade barriers reduces MR more for small countries than for large countries who have larger trade opportunities at home. Therefore, trade between large economies and within small countries is more affected by an equal change in trade costs.

NMSs are considerably smaller than EU15 countries, both individually and as a group,[16] implying that an adjustment of border effects and trade potentials estimated previously in the article is needed. In order to predict by how much trade flows would increase if border-specific barriers within Europe were to be eliminated, one can estimate directly the (8)–(9) equation system with nonlinear techniques as do Anderson and van Wincoop (2003). Recently, Baier and Bergstrand (2009) proposed an alternative estimation method using approximations of MR terms $\bar{P}_i$ and $\tilde{P}_j$ with a first-order log-linear Taylor-series expansion:

$$MR_k = (1-\sigma)\left(\sum_l \theta_k \ \ln \ t_{lk} - \frac{1}{2}\sum_k \sum_l \theta_l \theta_k \ \ln \ t_{lk}\right), \ \forall k, l = i, j \qquad (15)$$

and showed that the estimate of the effect of borders on trade is very close to the one obtained by Anderson and van Wincoop (2003). I follow their approach and estimate equation (8) in logarithmic form using exporter's production $prod_i$ and importer's consumption $cons_j$ as proxies for supply and demand effects. MR terms are obtained by integrating the trade costs function (10) in equation (15); they are the same for each country k whether it acts as exporter or importer. θ_i and θ_j are the shares of countries i and j in world GDP, and the overall MR for trade between any pair of countries i and j is the sum of exporter and importer MRs: $MR_{ij} = MR_i + MR_j$. Notice that these terms cannot be computed directly,

[16] In 1994 the 12 NMSs accounted for 3.5% of the region GDP; this percentage doubled by 2007.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

as parameters δ, b_0 to b_3, c_1 and c_2 are unknown. To overcome this difficulty, a MR term is computed for each element of the trade costs function.[17] The resulting equation:

$$\begin{aligned}\ln\ m_{ij} = {} & \alpha_0 + \alpha_1 prod_i + \alpha_2 cons_j + \lambda_1\ \ln\ d_{ij} - \lambda_1\ MR\_\ln\ d_{ij} \\ & + \lambda_2 contig_{ij} - \lambda_2 MR\_contig_{ij} + \lambda_3 comlang_{ij} - \lambda_3 MR\_comlang_{ij} \\ & + \lambda_4 CEE\_EU_{ij} - \lambda_4 MR\_CEE\_EU_{ij} + \lambda_5 intraEU_{ij} \\ & - \lambda_5 MR\_intraEU_{ij} + \lambda_6 intraCEE_{ij} - \lambda_6 MR\_intraCEE_{ij} + \epsilon_{ij}\end{aligned} \tag{16}$$

is estimated under the constraint that the coefficient of each trade cost component in equation (16) is equal to the opposite of the coefficient for the corresponding MR term.

Coefficients estimated with equation (16) are displayed in Table A2 of the Appendix. The first two columns of the table correspond to the first two columns of Table 1. With (15) one can test for the difference in the level of intranational trade costs for EU15 and NMSs and depending on the direction of flow. This is done in columns (3)–(5) by adding on the RHS a dummy variable equal to one for trade flows within each of the 12 NMSs and by breaking down the CEE–EU dummy in 2. The coefficient of the domestic CEE trade dummy is never statistically significant. This indicates that the level of intranational trade costs of these countries is not different from that of EU15. Similarly, CEE–EU trade costs do not seem to differ depending on the direction of trade. The last two columns of the table are simply replications of column (2) for the first and last year of the studied period.

The upper part of Table 4 summarizes average MR terms for EU15 and NMS country groups *with* and *without* border-specific costs. Multilateral resistances with trade barriers (MR) are computed according to equation (15). When some or all international trade barriers are removed, multilateral resistances (MR^*) are obtained by dropping the corresponding term(s). Similarly to Anderson and van Wincoop (2003), I integrated the effect of MR to obtain the overall effect of borders on trade between and within the two groups:

$$\text{adjusted border effect} = b_{ij}\ \exp\left(MR_i - MR_i^* + MR_j - MR_j^*\right) \tag{17}$$

[17] In particular, $\text{MR\_}\ln d_{ij} = \sum_j \theta_j \ln d_{ij} + \sum_i \theta_i \ln d_{ij} - \sum_i \sum_j \theta_i \theta_j \ln d_{ij}$ for distance, and $\text{MR\_dummy}_{ij} = \sum_j \theta_j \text{dummy}_{ij} + \sum_i \theta_i \text{dummy}_{ij} - \sum_i \sum_j \theta_i \theta_j \text{dummy}_{ij}$ for dummy variables *contig*, *comlang*, *CEE_EU*, *intraEU* and *intraCEE*.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 4 Adjustments for impacts on MR

Multilateral resistances (in logarithmic form)	EU	CEE
With: all border-specific costs	−2.31	−3.31
Without:		
– Intra-EU border-specific costs	−0.90	−4.41
– CEE–EU border-specific costs	−2.30	−0.18
– Intra-CEE border-specific costs	−2.32	−3.15
The effect of border-specific costs removal on bilateral trade[a]		
	All sample	2007
CEE–EU trade	1.39	1.50
Intra-EU trade	1.02	1.06
Intra-CEE trade	27.60	17.29

Note: [a]The ratio of trade without to trade with border-specific barriers, computed according to equation (17).

The left-hand side of equation (17) is the ratio of trade without border barriers to trade with border barriers. Asterisks denote border removal and b_{ij} is equal to $\exp(-\lambda_4)$ for CEE–EU trade, to $\exp(-\lambda_5)$ for intra-EU trade and $\exp(-\lambda_6)$ for trade between NMSs. If international trade costs were eliminated, the average MR of EU15 countries would decrease by a factor of $4.1[= \exp(-(-2.31 - (-0.90)))]$ but only by a factor of $1.2[= \exp(-(-3.31 - (-3.15)))]$ for NMSs. This difference in results comes from the difference in the levels of trade costs.

The impact of border-specific barriers on bilateral trade between and within the two groups of countries, adjusted for changes in MR, is reported in the lower part of the table. Removing border barriers would result into a 39% increase in the volume of CEE–EU trade, whereas trade between EU15 countries would only slightly change. When computations are performed for the last year in the sample, one may conclude that CEE–EU trade has reached only two-thirds of its potential, a 50% increase being still possible. This figure is lower than the potential obtained with the border-effect-ratio method, but considerably higher than the level predicted by traditional models in the literature. An adjusted border effect close to unity shows the proximity of intra-EU trade to its potential. This finding confirms once more that trade between old EU members is a good reference for other trade flows in the region. Trade integration is the weakest between NMSs and trade creation associated with these flows is particularly large. In conclusion, even after adjusting

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

for the impact of trade barriers on MRs, trade potentials within Europe remain quite large.

5 Conclusions

Trade between CEE and between CEE and EU countries improved remarkably during the last two decades, both in terms of border effects and trade potentials. The article shows that there is still place for important growth in bilateral CEE–EU transactions. This result is in contrast to most trade potential gravity models that claim that East–West European trade has already reached its highest integration level. Much higher trade potentials for both CEE–EU and intra-CEE trade are obtained when one controls for the amount of trade within national borders. At the beginning of the 21st century trade between CEE and EU countries represented two-thirds of its attainable level, suggesting a possible 39% increase with further EU integration. Adjusting for the impact of borders on MR, yields lower trade potentials, but above the magnitudes obtained with traditional approaches. An even larger trade potential is obtained for trade between NMSs, despite the strong reduction of bilateral border effects between these countries achieved during the 1990s.

Acknowledgements

I would like to thank the editor and two anonymous referees for their comments and remarks, which permitted to improve the quality of the article, the organizers and participants of the CESifo conference on Measuring Economic Integration, 18–19 February 2011, Munich, Germany, as well as Thierry Mayer and Chantal Le Mouel for insightful discussions and suggestions.

References

Anderson, J. E. and E. van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", *American Economic Review* **93**, 170–92.

Armington, P. (1969), "A Theory of Demand for Products Distinguished by Place of Production", *IMF Staff Papers* **16**, 159–76.

Baier, S. L. and J. H. Bergstrand (2009), "Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects Using the Gravity Equation", *Journal of International Economics* **77**, 77–85.

Baldwin, R. (1993), "The Potential for Trade Between the Countries of EFTA and Central and Eastern Europe", CEPR Discussion Paper No. 853.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Bergstrand, J. H. (1989), "The Generalized Gravity Equation, Monopolistic Competition, and the Evolution of the Factor-Proportions Theory", *Review of Economics and Statistics* **71**, 143–53.

Brunetti, A., G. Kisunko and B. Weder (1997), "Institutional Obstacles to Doing Business", World Bank Research Paper No. 1759.

Chen, N. (2002), "Intra-National Versus International Trade in European Union: Why Do National Borders Matter?", CEPR Discussion Paper No. 3407.

Combes, P.-P., M. Lafourcade and T. Mayer (2005), "The Trade Creating Effects of Business and Social Networks: Evidence from France", *Journal of International Economics* **66**, 1–29.

Coughlin, C. C. and D. Novy (2012), "Is the International Border Effect Larger than the Domestic Border Effect? Evidence from U.S. Trade", *CESifo Economic Studies* **59**, 249–76.

Deardorff, A. V. (1998), "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?", in J. A. Frankel, ed., *The Regionalization of the World Economy,* NBER, University of Chicago Press edition, pp. 7–28.

Fontagne, L., M. Freudenberg and M. Pajot (1999), "Le potential d'echanges entre l'Union europeenne et les PECO", *Revue Economique* **50**, 1139–68.

Gros, D. and A. Gonciarz (1996), "A Note on the Trading Potential of Central and Eastern Europe", *European Journal of Political Economy* **12**, 709–21.

Hamilton, C. and L. A. Winters (1992), "Opening up International Trade with Eastern Europe", *Economic Policy* **14**, 77–116.

Head, K. and T. Mayer (2000), "Non-Europe: The Magnitude and Causes of Market Fragmentation in Europe", *Weltwirtschaftliches Archiv* **136**, 285–314.

Head, K. and J. Ries (2001), "Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of US-Canada Trade", *American Economic Review* **91**, 858–76.

Helliwell, J. F. (1996), "Do National Borders Matter for Quebec's Trade?", *Canadian Journal of Economics* **29**, 507–22.

Maurel, M. and G. Cheikbossian (1998), "The New Geography of Eastern European Trade", *Kyklos* **51**, 45–72.

McCallum, J. (1995), "National Borders Matter: Canada-U.S. Regional Trade Patterns", *American Economic Review* **85**, 615–23.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Nilsson, L. (2000), ‘‘Trade Integration and the EU Economic Membership Criteria’’, *European Journal of Political Economy* **16**, 807–27.

Nitsch, V. (2000), ‘‘National Borders and International Trade: Evidence from the European Union’’, *Canadian Journal of Economics* **33**, 1091–105.

Papazoglou, C., E. J. Pentecost and H. Marques (2006), ‘‘A Gravity Model Forecast of the Potential Trade Effects of EU Enlargement: Lessons from 2004 and Path-dependency in Integration’’, *World Economy* **29**, 1077–89.

Rauch, J. E. (2001), ‘‘Business and Social Networks in International Trade’’, *Journal of Economic Literature* **39**, 1177–203.

Redding, S. and A. J. Venables (2004), ‘‘Economic Geography and International Inequality’’, *Journal of International Economics* **62**, 53–82.

Rose, A. and E. van Wincoop (2001), ‘‘National Money as a Barrier to International Trade: The Real Case for Currency Union’’, *American Economic Review* **91**, 386–90.

Santos Silva, J. M. C. and S. Tenreyro (2006), ‘‘The Log of Gravity’’, *Review of Economics and Statistics* **88**, 641–58.

Wang, Z. K. and L. A. Winters (1992), ‘‘The Trading Potential of Eastern Europe’’, *Journal of Economic Integration* **7**, 113–136.

Wei, S.-J. (1996), ‘‘Intranational Versus International Trade: How Stubborn are Nations in Global Integration’’, NBER Working Paper No. 5531.

Wolf, H. C. (2000), ‘‘Patterns of Intra- and Inter-State Trade’’, *Review of Economics and Statistics* **82**, 555–63.

Wolf, N. (2009), ‘‘Was Germany Ever United? Evidence from Intra- and International Trade 1885-1933’’, *Journal of Economic History* **69**, 846–81.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Appendix

Table A1 Trade potential with respect to intra-EU trade (percentage of actual trade)

Industry	CEE–EU trade		Intra-CEE trade	
	1994	2007	1994	2007
Food manufacturing	311	141	412	60
Beverage industries	878	611	657	269
Tobacco manufactures	1181	1138	29	291
Textiles	129	−24	346	23
Wearing apparel, except footwear	−7	−12	223	29
Leather	224	−20	1555	6
Footwear	174	206	685	1131
Wood, except furniture	58	7	11	−3
Furniture	6606[a]	34	692273[a]	236
Paper and paper products	256	48	123	−24
Printing and publishing	565	33	156	39
Industrial chemicals	236	87	110	17
Other chemical products	449	257	105	102
Rubber products	586	14	642	7
Plastic products	235	39	114	35
Pottery, china, and earthenware	486	12	1262	69
Glass and glass products	429	2	766	−29
Other non-metallic mineral products	125	62	92	10
Iron and steel basic industries	402	104	330	56
Non-ferrous metal basic industries	449	47	1071	19
Metal products, except machinery, and equipment	25	−45	31	−48
Machinery except electrical	298	25	386	98
Electrical machinery apparatus and appliances	135	−45	82	−59
Transport equipment	331	35	211	21
Professional and scientific, measuring and controlling equipment, photographic, and optical goods	230	107	214	393
Other manufacturing	205	54	517	97

Note: Trade potentials are obtained as in (12) and (13) using border effects estimated with equation (11) and PPML for each industry.

[a]Estimated coefficients are not statistically significant.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table A2 European trade integration and MRs

	(1) All int'l	(2) All flows	(3) All flows	(4) All flows	(5) All flows	(6) 1994 flows	(7) 2007 flows
ln exporter production	0.64[a]	0.59[a]	0.58[a]	0.59[a]	0.59[a]	0.59[a]	0.59[a]
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
ln importer consumption	0.59[a]	0.54[a]	0.54[a]	0.54[a]	0.54[a]	0.52[a]	0.54[a]
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
ln distance	−0.48[a]	−0.27[a]	−0.28[a]	−0.27[a]	−0.28[a]	−0.18[a]	−0.31[a]
	(0.05)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)
Common land frontier	0.41[a]	0.63[a]	0.63[a]	0.63[a]	0.63[a]	0.73[a]	0.63[a]
	(0.09)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Common language	0.23[c]	0.26[a]	0.26[a]	0.26[a]	0.26[a]	0.14	0.26[a]
	(0.12)	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)	(0.09)
CEE–EU	−0.49[a]	−3.48[a]	−3.49[a]			−4.45[a]	−3.11[a]
	(0.10)	(0.18)	(0.18)			(0.17)	(0.17)
Intra-EU		−2.84[a]	−2.84[a]	−2.84[a]	−2.84[a]	−3.26[a]	−2.63[a]
		(0.16)	(0.16)	(0.16)	(0.16)	(0.15)	(0.16)
Intra-CEE	−0.59[a]	−3.65[a]	−3.67[a]	−3.65[a]	−3.67[a]	−4.47[a]	−3.26[a]
	(0.13)	(0.20)	(0.21)	(0.20)	(0.21)	(0.34)	(0.18)
Domestic CEE			−0.06		−0.06		
			(0.14)		(0.14)		
CEE exports to EU				−3.44[a]	−3.45[a]		
				(0.20)	(0.21)		
EU exports to CEE				−3.52[a]	−3.53[a]		
				(0.19)	(0.19)		
Reference flows	Intra-EU	All domestic	Domestic EU	All domestic	Domestic EU	All domestic	All domestic
Number of observations	202 884	209 918	209 918	209 918	209 918	14 508	14 240

Note: The explained variable in all columns are bilateral imports at industry level according to the ISIC Rev. 2 classification of manufactured products. Estimations are obtained with the Poisson PML technique and year fixed effects. Standard errors in parentheses: [a], [b], and [c] represent statistical significance at the 1, 5, and 10% levels, respectively.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

CESifo Economic Studies, Vol. 59, 2/2013, 306–336 doi:10.1093/cesifo/ifs001
Advance Access publication 24 March 2012

The Role of NAFTA and Returns to Scale in Export Duration

Tibor Besedeš

School of Economics, Georgia Institute of Technology. e-mail: besedes@gatech.edu

Abstract

While exports within NAFTA face a lower hazard of ceasing, its onset has increased the hazard for Mexican and US intra NAFTA exports. Intra NAFTA exports still enjoy a lower hazard relative to exports to non-members. While NAFTA did affect the hazard for Canada's exports in the short run, its effect on Mexican and US exports is persistent. Exports of increasing-returns-to-scale (IRS) manufacturing products face the highest hazard in the case of Canada and Mexico, while IRS natural resource products have the highest hazard for Mexico. The effect of NAFTA on the returns to scale product types is exporter specific. (JEL codes: F10, F14)

Keywords: hazard, export survival, returns to scale, NAFTA

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

1 Introduction

The North American Free Trade Agreement (NAFTA) expanded the Canadian US Free Trade Agreement (CUSFTA) to include Mexico. One argument in favor of CUSFTA was the well-known source of welfare gains in trade models incorporating economies of scale—namely that tariff-free access to the large US market will allow Canadian firms to expand production, take advantage of economies of scale, and increase productivity resulting in welfare gains for both producers and consumers. Trefler's (2004) influential work has shown that the effect of CUSFTA was precisely as predicted by economy of scale models. CUSFTA initiated a contraction of low productivity plants resulting in a 12% decrease in employment, but also ushered a period of rising labor productivity increasing it by 15%.

Creation and expansion of free trade areas create new opportunities for firms to begin exporting their products. Little attention has been paid to NAFTA's effect on the ability of NAFTA members' firms to successfully maintain their exports to the NAFTA area. The goal of this article is to assess the effect NAFTA has had on the hazard of exports ceasing for the three member countries and to assess whether the effect of NAFTA is related to returns to scale in production, given one of the arguments in favor of NAFTA was rooted in returns to scale. I use annual exports disaggregated at the 6-digit Harmonized Schedule level between 1990 and 2007 for Canada, Mexico, and the USA.

© The Author 2012. Published by Oxford University Press
on behalf of Ifo Institute for Economic Research, Munich. All rights reserved.
For permissions, please email: journals.permissions@oup.com

The effect of CUSFTA/NAFTA on productivity is one aspect of the agreements to have been investigated. Trefler (2004) showed that Canadian industries that experienced the largest tariff cuts enjoyed largest productivity gains. Romalis (2007) shows that the two free trade agreements have had a substantial effect on the volume of trade, but a much smaller effect on prices and welfare. Much work has been devoted to identifying whether CUSFTA and NAFTA were trade creating or diverting. Clausing (2001) found that CUSFTA was primarily trade creating, while Trefler (2004) found evidence of both trade diversion and trade creation, with creation the dominant force. Romalis (2007) raised the possibility of a substantial trade diversion effect of NAFTA/CUSFTA which may be responsible for increased North American output and prices in once highly protected sectors.

Other researchers have focused on the effects of NAFTA/CUSFTA on the extensive margin. Kehoe and Ruhl (2009) found that especially in cases of thin trade relationships, such as between Canada and Mexico prior to NAFTA, trade liberalization is a key ingredient in sparking the growth of the extensive margin, an important source of new trade. Debaere and Mostashari (2010) find a small effect of trade liberalization on the extensive margin of US imports for the 1989–1999 and 1996–2006 periods.

I begin by providing a descriptive analysis of the evolution of exports of NAFTA members in a novel way. As the main focus of the article is the hazard of exports ceasing, only data on newly created trade relationships can be used to evaluate it. In the descriptive analysis I differentiate between new and old export relationships, where old are all relationships active in 1990, the first year in the data, while new are all relationships created after 1990. After the descriptive analysis of new and old export relationships, I present a simple motivating model along the lines of Melitz (2003) and then estimate the hazard of exports ceasing.

I find that exports of each NAFTA member to other members face a much lower hazard of ceasing than their exports to non-members. The onset of NAFTA itself has increased the hazard of Mexican and US exports to fellow NAFTA members ceasing and had no net effect on the hazard of Canadian exports to other NAFTA members ceasing. In terms of differences across the nature or returns to scale, exports of increasing-returns-to-scale (IRS) manufacturing products face the highest hazard in the case of Canada and Mexico, while in the case of the USA that distinction belongs to IRS natural resource products. Only in the case of Mexico are there significant differences in the hazard for all three product types, with IRS products having the lowest hazard.

The intra NAFTA exports of the three returns-to-scale product types are affected differently for each member by the onset of NAFTA. In the case of Canadian exports, NAFTA has increased the hazard for both IRS

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

products, but not in a statistically meaningful way. In the case of Mexico, NAFTA has had the strongest effect on the hazard of exports of IRS manufacturing products ceasing, increasing it, while it did not increase the hazard of exports of IRS natural resource and constant-returns-to-scale (CRS) products ceasing in a statistically significant manner. NAFTA has had the strongest effect on every returns-to-scale type product exported by the USA, significantly increasing the hazard of exports of such products ceasing to other NAFTA members, especially for both increasing returns to scale types. The effect of NAFTA has not been consistent over time, as it has increased the hazard in some 3-year periods after its onset and not in others.

The role of returns to scale and free trade agreements in duration of exports has not been examined to date. Thus, this article makes a contribution to the duration of trade literature in addition to making a contribution to the literature on the effects of NAFTA. Duration of trade was first examined by Besedeš and Prusa (2006a, b) who noted that most US import relationships are short lived and that differentiated products are exported to the US in longer lasting relationships than homogeneous goods. Besedeš (2008) showed that uncertainty in international trade and its effect on relationship formation as modeled by Rauch and Watson (2003) can account for many features of duration data. Nitsch (2009) has found similar results for German imports, while Brenton et al. (2010) and Fugazza and Molina (2009) find similar conclusions for a larger set of countries. Cadot et al. (2011) for four African economies and Görg et al. (2012) for Hungary reach similar conclusions for firm-level exports. Hess and Persson (2011a) examine duration of EU imports, while Besedeš (2011) examines how the hazard of exports ceasing of Eastern European transition economies was affected by the transition processes. Jaud et al. (2009) examine the relationship between financial constraints and duration of trade. Besedeš and Prusa (2010) provide a summary of the duration of trade literature.

2 Data and Preliminary Analysis

I use data on Canadian, Mexican, and US export flows recorded at the 6-digit Harmonized Schedule (HS) level. Data for Canada and Mexico come from the UN Comtrade database, while data for the USA were aggregated from 10-digit HS level data available from the US Census *US Exports* CDs/DVDs. I use annual data between 1990 and 2007 for all three countries. In each year I identify new export relationships, converting annual data into spells of active exporting, and track them until they cease to be active. A spell reflects the number of consecutive years

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 1 Data summary

Type of exports		Canada	Mexico	US
Annual observations	All	559,942	311,881	2,250,343
	New	416,970	253,713	1,165,839
	Fraction new	0.75	0.81	0.52
Relationships	All	140,215	71,082	356,969
	New	116,046	61,990	240,942
	Fraction new	0.83	0.87	0.68
Spells	All	231,055	124,300	621,910
	New	206,886	115,208	505,883
	Fraction new	0.90	0.93	0.81

during which a relationship is active. A relationship is defined as the instance of a country exporting a 6-digit HS product to another country, such as Mexican exports of 'Monumental/building stone, cut/sawn' (HS 680221) to Argentina, Australia, Belgium, Brazil, Canada, France, Italy, Japan, and the USA among others.

To identify the production technology and the nature of returns to scale for each product, I use the classification developed by Antweiler and Trefler (2002). They identify four types of returns to scale: IRS manufacturing, IRS natural resources, CRS, and non–robust IRS industries for which they could not establish the exact nature of the returns to scale. The latter group of industries are omitted from the analysis. The share of export volumes and export relationships of products with identified returns to scale varies across the three countries. Canada has some 22% of its volume and a half of its relationships in products with robustly identified returns to scale. Mexico has 40% of its volume and 44% of its relationships, while the US has 65% of volume and 59% of relationships in such products.

Table 1 presents summary information for products with robustly identified returns to scale for each country. Since I focus on exports created after 1990 the table presents information for all exports as well as new exports. A note on the use of the term 'new' is in order—it refers to all export volumes or relationships created after 1990. For example, in 2005 new exports and relationships would be all those created since 1990 and not only those created in 2005 alone. In all figures below old and new values are normalized by the total 1990 values. Thus, values for new exports are fractions of the 1990 value of total exports.

The US has significantly more annual observations, export relationships, and spells of service than Canada and Mexico put together. Perhaps the largest difference across the three countries is in the fraction

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

of observations in exports created since 1990. While Canada and Mexico have 75 and 81% of all annual observations created after 1990, the US has almost a half of all of its observations started prior to 1990. These differences decrease as one looks at export relationships, 83 and 87% for Canada and Mexico and 68% for the US, and spells of service, 90 and 93% for Canada and Mexico versus 81% for the US.

2.1 New versus old exports

Figure 1 examines differences between exports started prior to 1990, old exports, and those started afterwards, new exports. Top panels show the total volume of both old and new exports, while bottom panels show the total number of both old and new export relationships. For every country, new exports embody a significantly lower volume, with the difference the largest for Mexico. While old exports embody more value, they also grow at a much slower rate than new exports. For Canada the 1991–2007 growth rate of the volume of old exports is 141 and 1.512% for new exports. For Mexico the corresponding figures are 727% and 2.172%, while for the US they are only 13% for old exports and 1.372% for new exports. By 2007 some 15% of all Canadian, 12% of all Mexican, and 16% of all US exports are embodied in relationships started since 1990.

While the top row of Figure 1 shows that new exports are significantly smaller in volume than old exports, the opposite is true for the number of export relationships which carry that volume. The number of old relationships declines over time for every country since their ranks cannot increase by definition. The rate of decline from 1991 to 2007 is similar across the three countries: 66% for Canada and 57% for both Mexico and the US. Both Canada and Mexico have slightly more than a half of all relationships in 1991 in old exports, while the US has almost 80% of its relationships in old exports. Canada's new relationships grow by 513%, followed by Mexico at 469%, and the US at 186%. While new exports account for a relatively small share of total exports in 2007, new relationships account for a significant number of all relationships in 2007: 91% for Canada, 92% for Mexico, and 68% for the US.

A final note on the growth in the number of relationships. The rate at which new relationships are created exceeds the rate at which new relationships are ended. This is true in almost every year for every country as the number of relationships started after 1990 grows in almost every year.[1]

[1] The exceptions are 1996 for Canada; 1996, 1997, 2000, and 2001 for Mexico; and 1997, 1998, 2002, and 2007 for the US.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

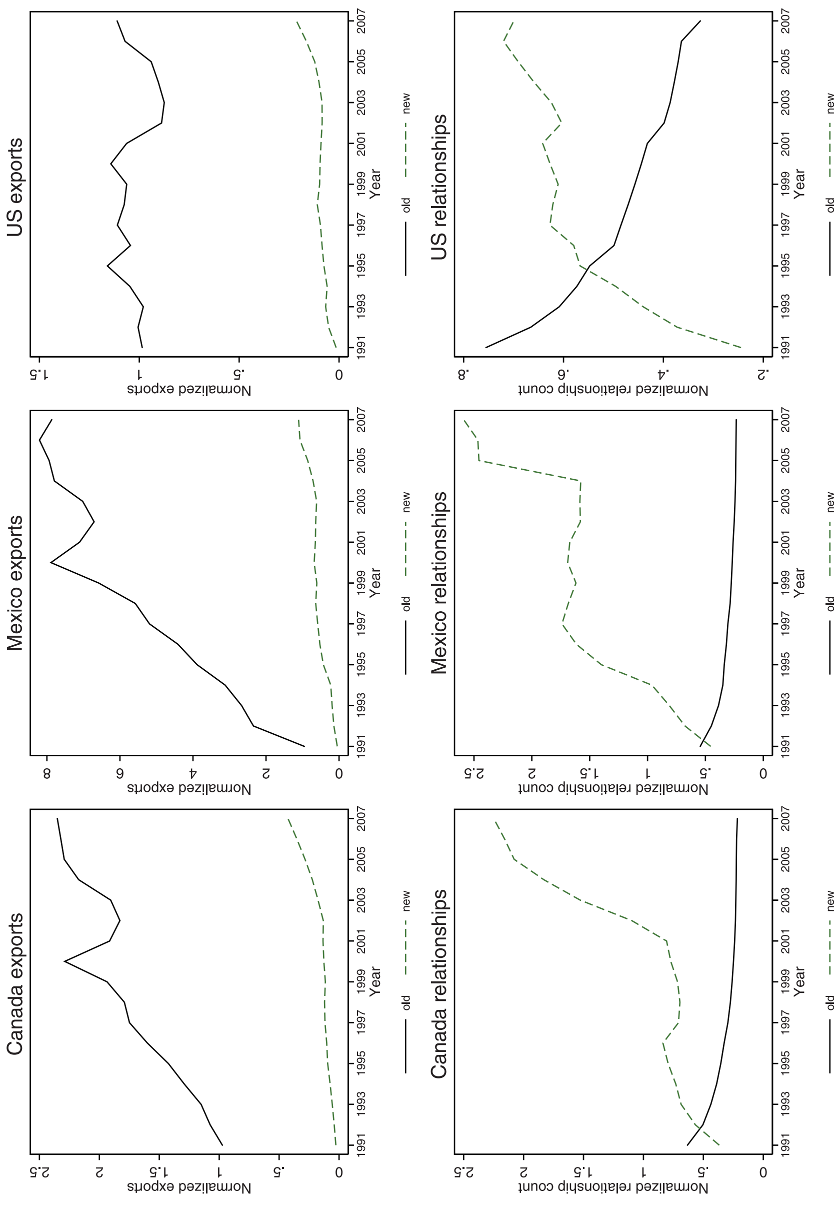

Figure 1 New and old exports and relationships.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

These facts are supportive of the broad conclusion of Besedeš and Prusa (2011), that, for at least some countries, the extensive margin is relatively unimportant in the sense that over time it does not account for much of the volume of exports, even though it accounts for a significant share of relationships. There is a lot of churning at the extensive margin in terms of the number of relationships, but much less in terms of the volume.

2.2 Intra NAFTA export shares

Top panels in Figure 2 illustrate NAFTA shares of export volumes and relationships embodied in old exports, while bottom panels shows NAFTA shares of new exports. In every country much of the effect of NAFTA is by far stronger in exports active in 1990. By 2000 almost 90% of Canada's volume of exports was in old relationships destined for Mexico and the US. The share of NAFTA destined old relationships for Canada has quadrupled from 10% to 40%. Mexico has enjoyed an even stronger dominance of NAFTA destined old exports, with the share of volume increasing to >95% and the share of relationships increasing to 60%. The US has the most diversified structure of exports with the share of the volume of exports in old relationships destined to NAFTA members doubling from 20% to 40% and the share of relationships increasing by some 2 percentage points to 8%.

Patterns for new exports created after 1990 are more varied. While the share of the volume of Canadian exports destined to NAFTA members has increased by a factor of 8–24%, the share of relationships displays an inverted U shape. For Mexico the share of the volume has increased from 5% in 1991 to just under 20% by 2007, but having increased to as much as 40% in the intervening years. The share of relationships was slightly lower in 2007 at roughly 6% than in 1991, but has increased to as much as 9% in the intervening years. The decrease in both Canadian and Mexican exports to NAFTA members in the latter part of this period may have been caused by the displacement of trade due to the rising presence of Chinese exports in the US, as investigated by Iacovone et al. (2010). The US has had the smallest share of new exports destined to NAFTA members, a consequence of a more diversified export structure.

2.3 Export shares across returns to scale

In Figure 3, I examine the evolution of shares of export volumes and relationships across the three types of returns to scale. While the majority of Canadian exports involve IRS manufacturing products, they have experienced a fair amount of change since 1990. The creation of CUSFTA resulted in a rapid drop in the share of IRS manufacturing products from 80% of exports to 50% by 1994. The addition of Mexico

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

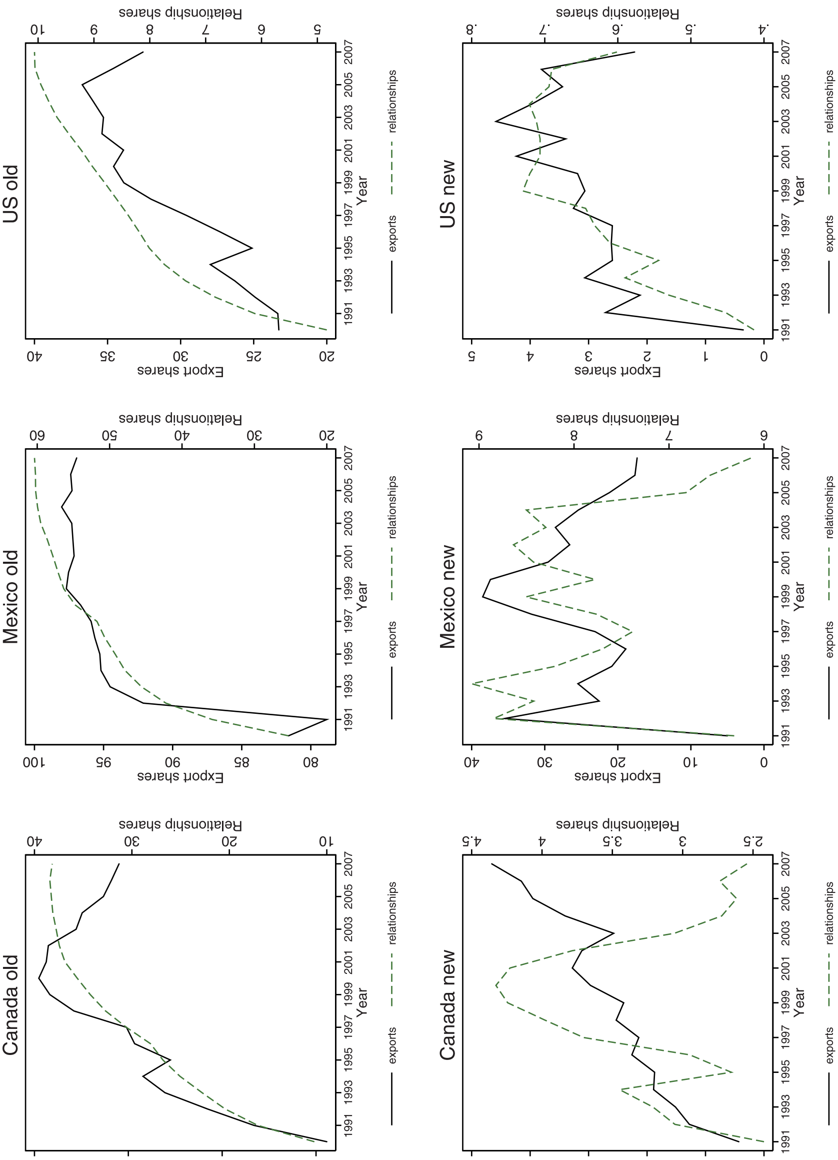

Figure 2 New and old exports and relationships to NAFTA members.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

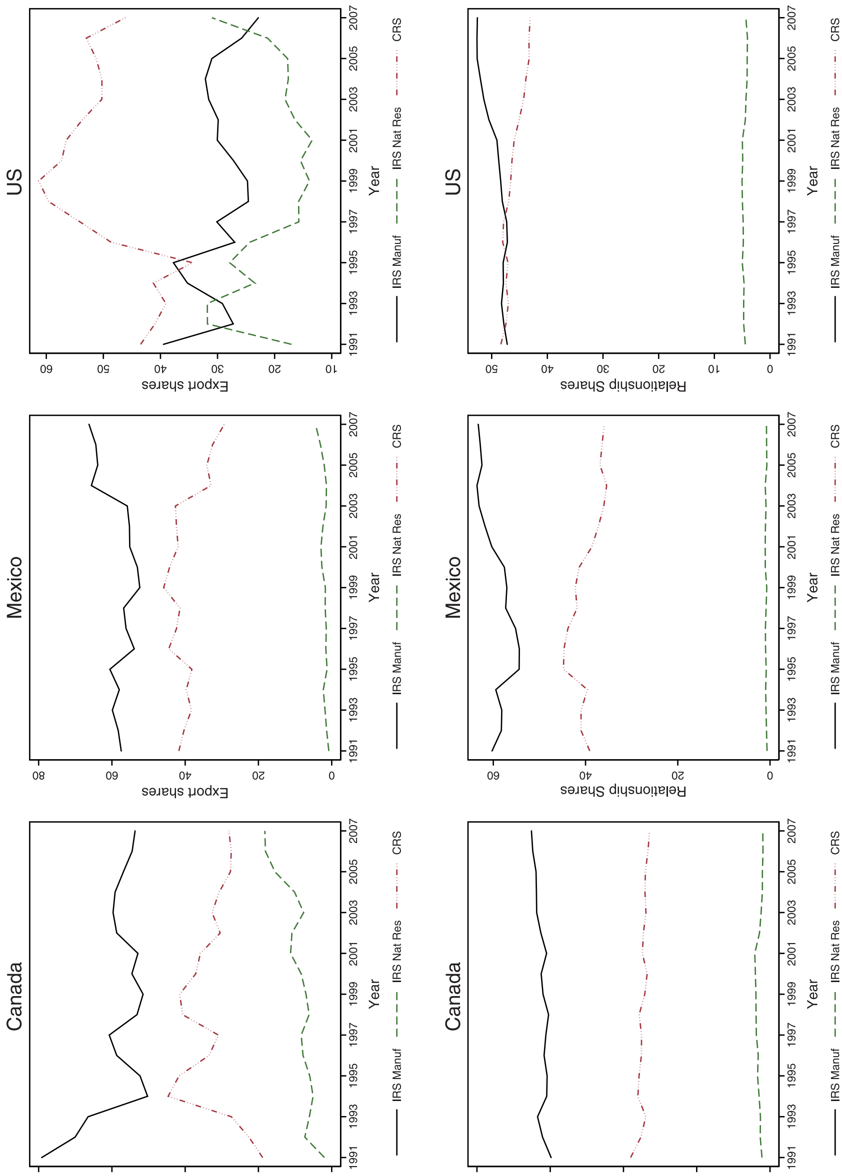

Figure 3 IRS export and relationship shares.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

to CUSFTA resulted in an increase in the share of IRS manufacturing products to 60% which has trended back toward 50% by 2007. The share of exports in CRS products more than doubled from 20% to 45% by 1994, but has trended downward ever since. The share of IRS natural resource products has increased rapidly from $<5\%$ of exports to almost 20% by 2007.

Mexico has had a different experience. While the share of IRS manufacturing products in exports has decreased to 55% immediately after NAFTA took effect, it has since increased to about 65%. The share of CRS products has decreased from slightly $>40\%$ to $\sim 30\%$ after an initial boost from NAFTA. The share of IRS natural resource products has increased consistently, though to a much lesser degree than for Canada. The US has a more balanced distribution of exports across the three types of products. Unlike Canada and Mexico the US has the largest share of its exports in CRS products. Following CUSFTA the share of CRS products decreased from 44% to 35% by 1995, followed by a sharp increase to 60% by 1999. It has decreased steadily since to about 47% in 2007. The share of IRS manufacturing products has fluctuated from some 40% in 1991 and 1995 and has decreased to some 25% in 2007. IRS natural resource products decreased from $>30\%$ in 1994 to $<15\%$ in 2001 before increasing to roughly 30% by 2007.

NAFTA members are more similar in the distribution of relationships. In every country IRS manufacturing products account for the largest share of exports and have followed rather similar paths. For all three the share of IRS natural resources has been relatively stable over time, being the largest for the US at some 5%. As a result, shares of IRS manufacturing and CRS products largely look as mirror images of each other. IRS manufacturing has increased slightly in Canada, and somewhat more for Mexico and the US, though the total increase in the share from 1991 to 2007 is only several percentage points.

2.4 Intra NAFTA export shares and returns to scale

There are larger fluctuations in shares of exports destined for NAFTA members, as seen in Figure 4. CRS products dominate the share of exports for each country. For Canada, it has increased from some 40% in 1991 to almost 80% in 1995 before collapsing to some 20% by 2007. For Mexico their share has fluctuated between some 40% in both 1991 and 2007 and $>60\%$ reached on several occasions. In the US their share has steadily increased from $<40\%$ to $>75\%$ by 2007.

Canada's exports of IRS manufacturing goods rapidly decreased with the formation of CUSFTA from almost 60% in 1991 to $<20\%$ in 1995 before rebounding under NAFTA to $>30\%$ in the early 2000s and settling

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

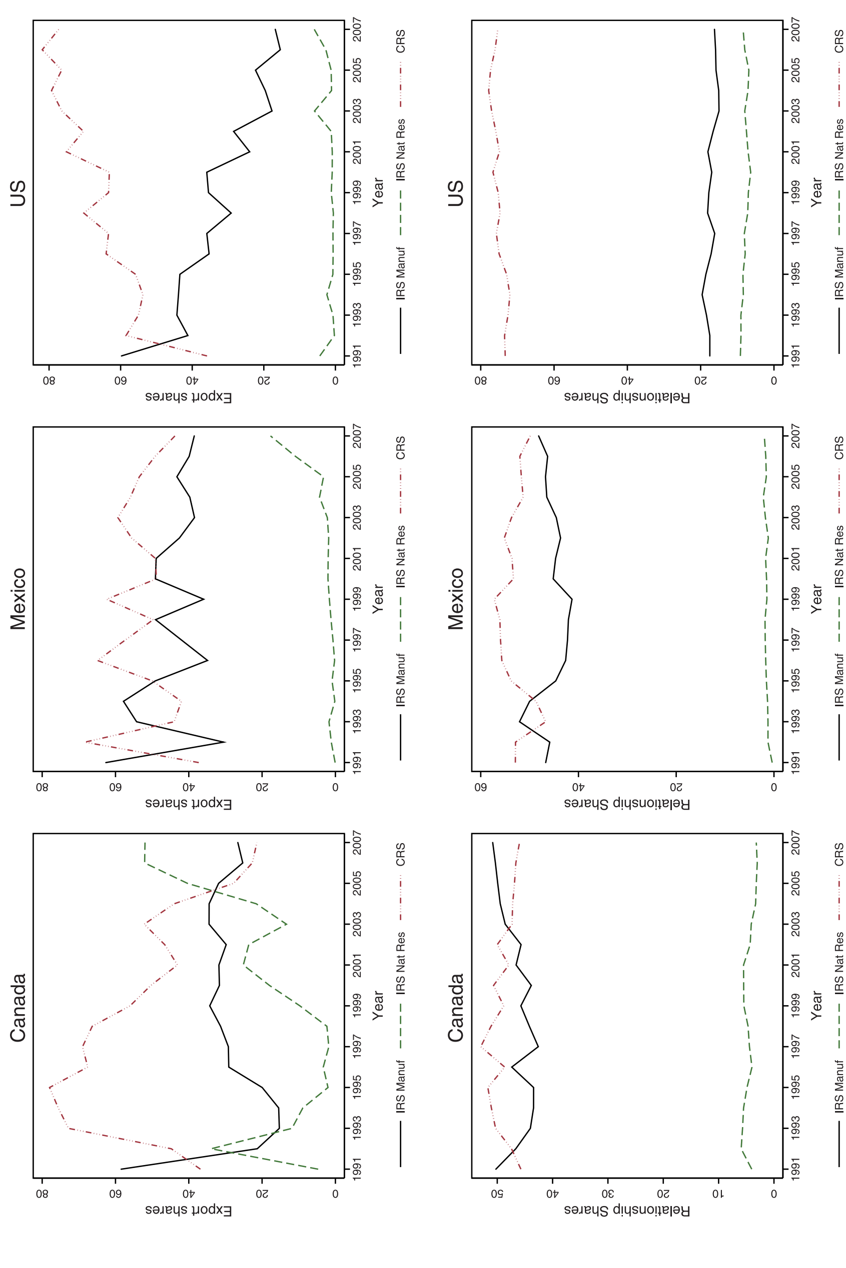

Figure 4 IRS export and relationship shares in exports to NAFTA members.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

at some 25% in 2007. The share of IRS manufacturing products in Mexico's exports to NAFTA members has decreased from 60% to around 40%, while their share in US exports to NAFTA members has decreased from some 60% to around 20%. IRS natural resource products share in Canadian exports increased rapidly between 1991 and 1992, decreased as rapidly by 1995, and finally increased to $>50\%$, dominating Canadian exports by 2007. In Mexico's exports the share of IRS natural resource products increased modestly until 2005 when they enjoyed a rapid increase to almost 20%. Their share in US exports has remained largely constant and small.

Fluctuations have been much smaller in the distribution of relationships. The share of IRS natural resource product relationships has remained steady at or below 10% for all three NAFTA members. The share of CRS product relationships in Canadian exports to NAFTA has increased from 50% to $>60\%$ in the late 1990s before returning to some 50%. The share of IRS manufacturing products has mirrored that of CRS products, first decreasing from 60% before returning to that level by 2007. The share of CRS product relationships in Mexican exports to NAFTA members has fluctuated around 50% and that of IRS manufacturing has fluctuated around 45%. The US offers a different picture with CRS products relationships accounting for a roughly constant 75% share of relationships and IRS manufacturing products accounting for a roughly constant 18%.

3 Motivating Model

To motivate the estimation of the hazard of exports ceasing consider the following extension of the Melitz (2003) model. Firms are characterized by a distribution $\mu(\phi)$ of productivity levels over a subset of $(0, \infty)$.[2] In any given year firms are subject to exogenous productivity or product appeal shocks, μ. These shocks may be positive or negative and their persistence will differ across firms. The effect of the shock is to affect the position of each firm within the distribution of productivity levels, $\mu(\phi)$. A positive shock will move the firm up in the distribution, while a negative shock will move the firm down.

The immediate effect of the shock for trade is that it will affect the firm's ability to export, depending on the firm's initial starting point and the persistence of the shock. For example, for an exporting firm a temporary negative shock may cause the firm to cease exporting. Once the shock dissipates, the firm will resume its exporting activities. Such a shock accounts for the possibility that exporters exit and re-enter exporting,

[2] As in Section 2.3 of Melitz (2003)

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

while never ceasing to produce for the domestic market. Clearly a large and permanent negative shock may affect the firm so drastically that it exits the industry altogether. A temporary positive shock may allow an otherwise purely domestic firm to export for a brief period before it dissipates. Clearly a permanent shock will create a permanent exporter. Depending on the nature of the shock, one can expect to observe a range of firm export behavior, from firms that never export to temporary in-and-out exporters to largely persistent exporters with short exits to finally permanent exporters.

A free trade agreement affects firms through its effect on the productivity shock μ. Free access to a foreign market can either be attributed to a positive productivity shock or a positive product appeal shock as it reduces costs of supplying a market. NAFTA may have had a differential effect based on the nature of the returns to scale used in production. Trade models with economies of scale argue that access to a large foreign market increases firm productivity as Trefler (2004) has found for NAFTA. This would entail an increase in the productivity parameter $\mu(\phi)$, thus making it more likely firms would both start exporting and export longer, even if for some the shock is temporary. Thus, NAFTA can reduce the hazard of exports ceasing. NAFTA may have had an offsetting effect as well, as some firms may be induced to try to export, even though they should not. Some firms may be overly optimistic about their ability to successfully export and attempt to do so only to find themselves abandoning exports quickly. The overall effect of NAFTA can therefore be seen as a the balance of these two opposing forces.

4 Hazard of Exports Ceasing

4.1 Econometric methodology and specification

I estimate the hazard of exports ceasing by estimating a probit model. As argued by Hess and Persson (2011b), using a probit estimator is more appropriate for discrete duration data, as annual trade data are, and does not impose the restrictive proportional hazard assumption.[3] Unobserved heterogeneity is another reason to use a discrete-time model such as probit as it is more easily addressed in such models than in the Cox proportional hazard model. To take unobserved heterogeneity into account, I estimate the random effects probit model of the hazard rate

[3] Brenton et al. (2010) and Hess and Persson (2011b) show that the assumption of proportional hazards is not satisfied in annual trade data. Hess and Persson (2011b) examine several other estimators which relax the proportionality assumption and recommend that probit be used.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

with random effects at the spell level. I model the dependence of the hazard rate on time by including a dummy for each year in the spell.

Advantages of the probit-estimated hazard model come at the cost of a more complicated interpretation of estimated coefficients. Neither coefficients nor the associated marginal effects reveal the true effect of each covariate. Although a given coefficient may be statistically significant, whether it makes a difference for the estimated hazard depends on the standard error of every covariate, all pairwise covariances, and the distributional specification of the probit model. To ascertain whether estimated hazards for different values of covariates of interest are statistically different I plot estimated hazards with the 95% confidence interval, which is denoted with dotted lines in every figure. All estimated hazards are estimated with covariates at their mean levels, with the exception of covariates of interest.

I use several country- and product-level variables to estimate the hazard of exports ceasing. The gross domestic product of the importer is expected to reduce the hazard, while distance to the importer is expected to increase the hazard. I use two measures of common language; one capturing whether two countries share an official language and the other whether $>9\%$ of the population speak the same language. In as much as common language reduces costs, both are expected to reduce the hazard. I use two measures at the country–product level to capture information spillovers: one measures the number of products exported to the same country, while the other measures the number of countries to which the same product is exported. The former measures experience with a country, while the latter measures experience with a product. Both are expected to reduce the hazard.

A lower economic risk of the importer is expected to reduce the hazard. The volume of initial exports should reduce the hazard, reflecting the confidence exporters (or their importing partners) have in their ability to consistently export their products (Rauch and Watson 2003; Besedeš 2008). The coefficient of variation of unit values measures the extent of the variation of unit values for each product across all export destinations. It reflects the extent of product differentiation. I use a dummy to capture any colonial relationship in the past. Finally, I use dummies to capture each multiple instance of a relationship.

I use two dummies to capture the effect of returns to scale, one for IRS manufacturing products and one for IRS natural resource products with CRS products as the baseline. Since spells of export relationships created after 1990 stretch across the period prior to and after the establishment of NAFTA, it is possible to distinguish between the two NAFTA-related effects. Therefore, I use two dummies. One simply captures exports to NAFTA members, while the other captures whether NAFTA itself is in

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

effect. Estimates of these four dummies are of main interest. A data appendix lists all data sources.

4.2 Basic results

Table 2 collects the basic results. Most variables have identical qualitative effects across the three exporters. Similar to other papers in the literature, the larger the GDP of the importer the lower the hazard for both Canadian and US exports. The effect of importer's GDP on the hazard of Mexican exports is positive—Mexican exports to larger economies face a higher hazard. To the extent that Mexican exports to larger and more developed economies are potentially of lower quality may result in them experiencing a higher hazard than exports to smaller economies.

The economic risk variable offers a somewhat puzzling result. With higher values indicating a riskier economy, a negative coefficient implies that US exports to riskier economies face a lower hazard, rather than a higher one as one might expect. This result highlights the presence of uncertainty in international trade as modeled by Rauch and Watson (2003) and empirically investigated by Besedeš (2008). It is possible that exports to highly risky economies are undertaken only once the exporter is relatively certain its exports will be long lived and will generate a profit. Such a strategy minimizes costs associated with exporting especially in situations where there is uncertainty as to whether exports will be profitable. Economic risk has no significant effect on Canadian exports, while Mexican exports face a lower hazard when destined for less risky markets, as expected.

The volume of exports at the start of a relationship has a significant negative effect for all three countries resulting in longer lived spells for relationships starting with a larger volume. Information spillovers have large negative effects—the more products exported to a country or the more countries a product is exported to, the lower the hazard. Both of these results are consistent with Cadot et al. (2011) and Besedeš (2011). The more variable are the unit values for Canadian and US exports the higher the hazard. The effect for Mexican exports is the opposite, with more variable unit values resulting in a lower hazard.

To the extent that distance reflects transportation costs, the further away the export market the higher are the hazard of Canadian and US exports, as expected. For Mexico the effect is the opposite—hazard is lower for exports destined for markets further away from Mexico. Official common language has a significant negative effect for Canadian and US exports resulting in a reduced hazard. Mexican exports to countries with Spanish as the official language face a higher hazard. The minority common language has a statistically significant negative effect for

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 2 Hazard estimates

	Canada	Mexico	US
GDP (ln)	−0.014***	0.045***	−0.047***
	(0.003)	(0.003)	(0.002)
Economic risk (ln)	−0.040	0.108***	−0.072***
	(0.025)	(0.032)	(0.012)
Initial exports (ln)	−0.048***	−0.086***	−0.120***
	(0.002)	(0.002)	(0.002)
Partners by product (ln)	−0.439***	−0.531***	−0.705***
	(0.005)	(0.009)	(0.007)
Products by partners (ln)	−0.269***	−0.360***	−0.437***
	(0.007)	(0.007)	(0.006)
Cov unit values (ln)	0.028***	−0.009***	0.019***
	(0.002)	(0.002)	(0.003)
Distance (ln)	0.047***	−0.092***	0.096***
	(0.009)	(0.013)	(0.006)
Common language (official)	0.006	−0.069**	−0.034***
	(0.015)	(0.028)	(0.006)
Common language (minority)	−0.021	−0.020	−0.026**
	(0.017)	(0.025)	(0.010)
Colonial relationship	0.104***	0.073***	0.023***
	(0.007)	(0.009)	(0.005)
IRS manufacturing	−0.002	−0.107**	0.096***
	(0.018)	(0.043)	(0.010)
IRS natural resources	−0.270***	−0.858***	−1.011***
	(0.054)	(0.066)	(0.073)
NAFTA members	−0.028*	0.058**	−0.052***
	(0.016)	(0.028)	(0.006)
NAFTA in effect	0.034	0.129***	0.168**
	(0.057)	(0.048)	(0.075)
Constant	3.549***	3.423***	7.829***
	(0.114)	(0.146)	(0.090)
Observations	187 188	168 642	667 787
Spells	103 851	81 732	292 694
Log-likelihood	−101 025	−91 296	−376 942
ρ	0.022***	0.190***	0.204***
Year in spell FE	Y	Y	Y
Spell number FE	Y	Y	Y

Robust standard errors clustered by relationships in parentheses with *, **, *** denoting significance at 10, 5, and 1%.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Mexico and the US, indicating longer duration and lower hazard, while it has no effect on Canadian exports. Exports to countries with which Canada and Mexico share a colonial history are no different form exports to other countries, while in the case of the US such exports face a lower hazard.

The relative importance of unobserved heterogeneity is reflected by estimates of ρ, which can be interpreted as the fraction of individual spell variation due to variation in unobserved factors. There are large differences in unobserved heterogeneity between Canada on one side and Mexico and the US on the other. While unobserved heterogeneity plays a statistically significant role in all three cases, the magnitude of ρ is almost an order of magnitude larger in the case of Mexico and the US (0.190 and 0.204 versus 0.022).

4.3 Effects of NAFTA

Exports to a NAFTA member face a significantly lower hazard for all three countries, with the effect strongest for the US and weakest for Canada. Surprisingly, the establishment of NAFTA increased the hazard of Mexican and US exports destined to NAFTA markets, while it had no effect on Canadian exports. To properly evaluate whether NAFTA has had a significant effect, I plot the estimated hazard of exports ceasing for non-NAFTA members and for NAFTA members both in the absence and presence of NAFTA in Figure 5. I include the 95% confidence interval around the estimated hazard for non–members and members in the absence of NAFTA. If the estimated hazard for members in the presence of NAFTA lies outside the confidence interval, then NAFTA has had a significant effect.

The hazard of exports to NAFTA members ceasing is lower than to non-NAFTA countries for all three members with differences statistically significant. The difference between the hazard of exports ceasing to NAFTA members and non–members is the largest for Mexico and the US. The pure effect of joining NAFTA is slightly positive for Canada, increasing the hazard, but is not statistically significant. The enactment of NAFTA has a much larger effect for Mexico and the US, increasing the hazard of exports to NAFTA members ceasing. In addition, the estimated hazard for exports ceasing to NAFTA members with NAFTA in effect lies outside the 95% confidence interval (dotted lines) for the estimated hazard for exports ceasing to NAFTA members, with the difference larger for US exports.

The onset of NAFTA has increased the hazard of exports to member countries ceasing in the case of Mexico and the US. In neither case is this increase sufficiently large to offset the lower hazard associated with

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

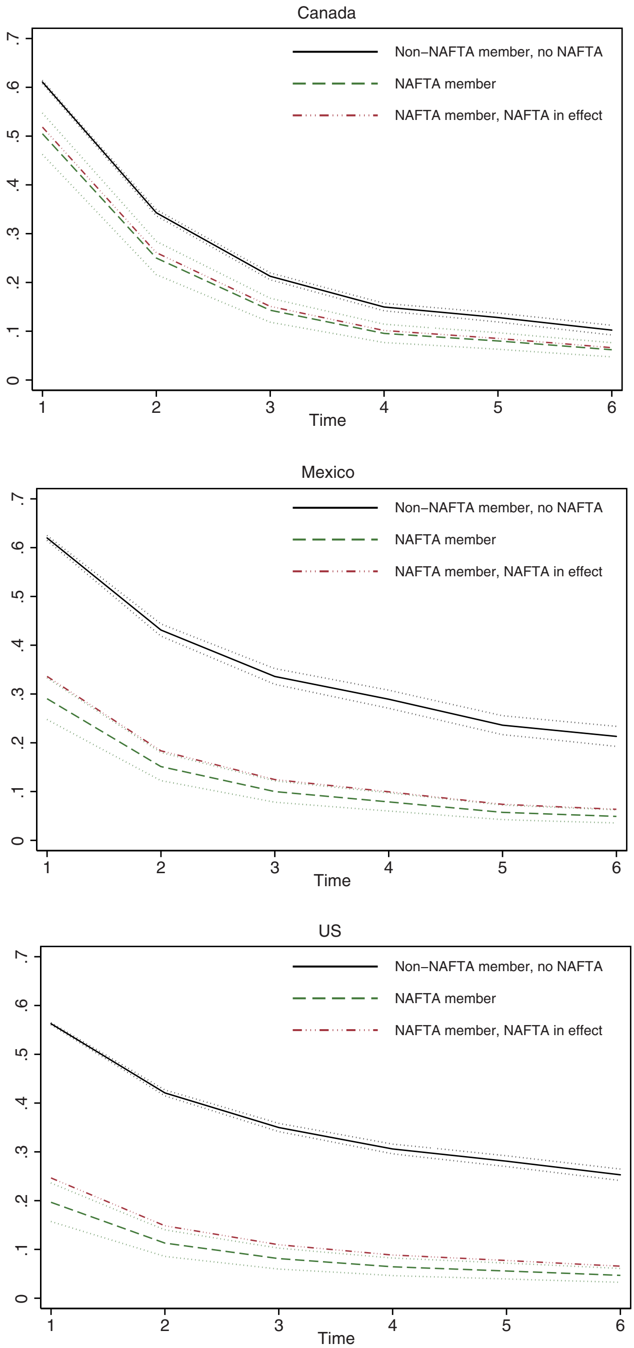

Figure 5 The effect of NAFTA on the estimated hazard.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

exporting to a NAFTA member. One explanation for the NAFTA effect is that its establishment may have induced too many firms to try to take advantage of new opportunities created by NAFTA, thus resulting in more failures. Essentially, NAFTA may have induced some firms who otherwise would not export to do so, only to quickly realize they cannot compete in the foreign market.

4.4 Effects of returns to scale

In terms of returns to scale, exports of IRS manufacturing products face a higher hazard than CRS products for Canadian, Mexican, and US exports. US exports of IRS natural resource products face a higher hazard, while Mexican exports of IRS natural resource products face a significantly lower hazard than their exports of CRS products. There are no differences between the two types of products in the case of Canadian exports.

Figure 6 illustrates estimated hazard for the three types of returns to scale for each country, along with the corresponding 95% confidence intervals. There are virtually no differences between CRS and IRS natural resource products for Canada, while its exports of IRS manufacturing products face a statistically different and higher hazard. In the case of Mexico, IRS manufacturing products also face a statistically significantly different and highest hazard of exports ceasing. Mexican exports of IRS natural resource products face the lowest hazard. Differences between the three returns to scale type products for Mexico are statistically different, especially for the first few years in a spell. The ordering of estimated hazards is different for the US with IRS natural resource products facing the highest hazard of exports ceasing, which is statistically different from the other two types. While IRS manufacturing products exported by the US face a higher hazard, it is not statistically different from that for CRS products.

4.5 The effect of NAFTA over time

To better understand the nature of the effect of the implementation of NAFTA on the hazard of exports ceasing, Table 3 shows the results with the NAFTA-in-effect variable having a time-dependent effect, for 3-year intervals starting in 1994. It is possible NAFTA has had a different effect over time as firms adjust to it and since some of its provisions were phased in over a period of time.[4] Since the coefficients for other variables are

[4] Baier and Bergstrand (2007) show in a gravity framework that the full effect of free trade agreements on the partners' trade volume takes 10 years to accrue.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

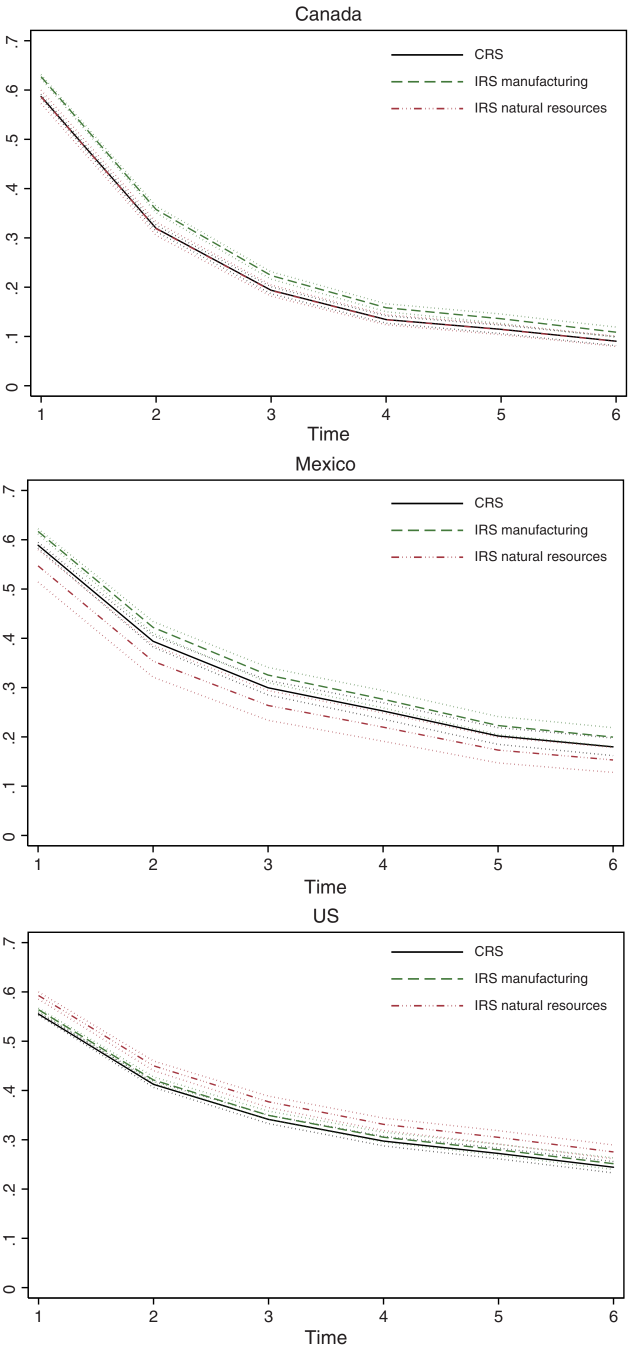

Figure 6 The effect of returns to scale on the estimated hazard.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 3 Time-dependent NAFTA effect

	Canada	Mexico	US
NAFTA members	−0.267***	−0.858***	−1.012***
	(0.055)	(0.066)	(0.073)
NAFTA in effect 1994–1996	0.201***	0.021	0.170**
	(0.071)	(0.055)	(0.086)
NAFTA in effect 1997–1998	−0.058	0.171***	0.067
	(0.074)	(0.054)	(0.086)
NAFTA in effect 2000–2002	−0.063	0.241***	0.320***
	(0.067)	(0.055)	(0.084)
NAFTA in effect 2003–2005	0.050	0.111**	0.107
	(0.061)	(0.055)	(0.086)
ρ	0.0228***	0.184***	0.204***

Robust standard errors clustered by relationships in parentheses with *, **, *** denoting significance at 10, 5, and 1%; year in spell and spell number fixed effects included.

virtually unchanged, I only present the coefficients for the time-dependent NAFTA-in-effect variable.[5] These four dummies indicate in which years did the existence of NAFTA affect the hazard.

While the onset of NAFTA had no net effect on Canadian exports, the time-dependent coefficients indicate that the onset of NAFTA did increase the hazard of Canadian exports to the other two NAFTA members immediately after the onset of NAFTA and had no effect after 1997. The opposite holds for Mexico: there is no effect in the first 3 years of Mexican participation in NAFTA, while in every subsequent 3-year period exports to other NAFTA members face a significantly higher hazard of ceasing. US exports to NAFTA members face a higher hazard consistently since the onset of NAFTA, with the exception of the 3-year period between 1997 and 1999.

Figure 7 plots the estimated hazard of ceasing for exports to NAFTA members in the absence of NAFTA with the 95% confidence interval along with estimated hazards for the four subperiods. In the case of Canada, the effect of NAFTA between 1994 and 1996 is outside the confidence interval indicating that during those 3 years NAFTA did indeed increase the hazard of Canadian exports to NAFTA members ceasing. In the case of Mexico, the effect is statistically significant between 1997 and 2002, while the effect between 2003 and 2005 lies just inside the 95% confidence interval. In the case of the US, NAFTA had a statistically significant effect immediately upon its inception, between 1994 and

[5] Full results are available on request.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

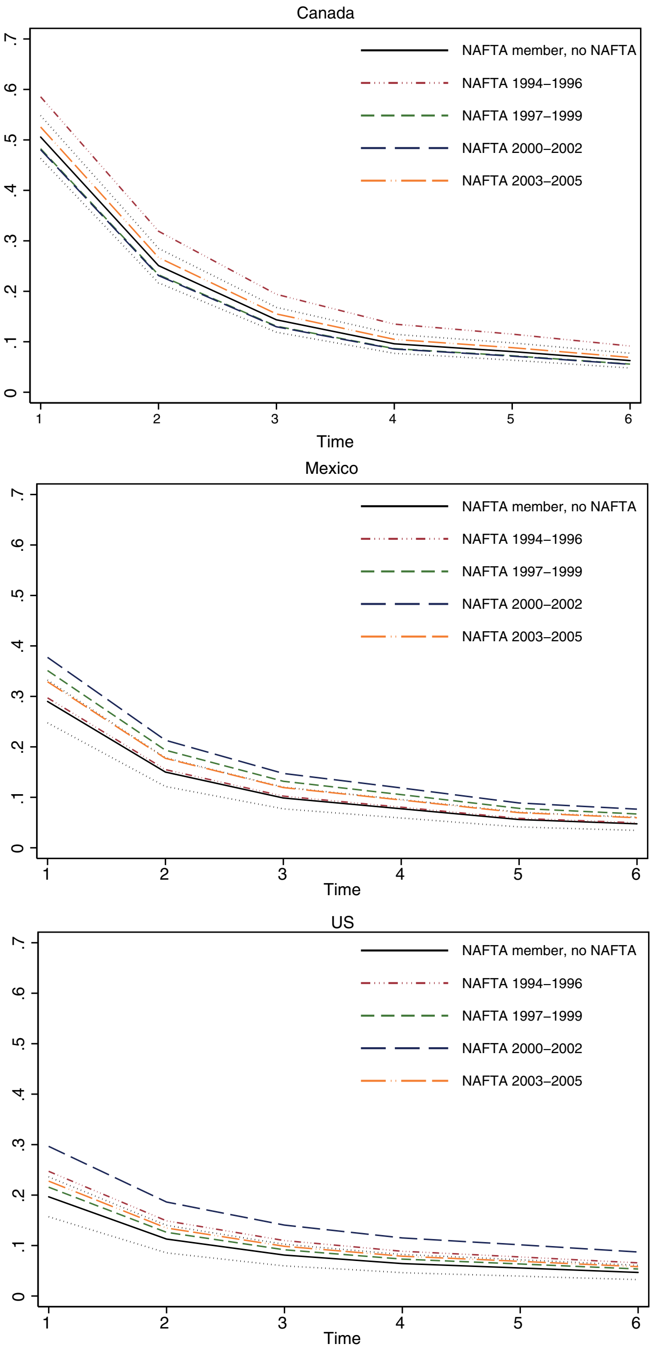

Figure 7 The effect of NAFTA across time.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

1996, and again between 2000 and 2002 when it increased the hazard significantly (by almost 10 percentage points at the start of a spell).

4.6 NAFTA and returns to scale

I have argued above that one might expect the hazard of exports ceasing for IRS manufacturing products to be the lowest when destined to NAFTA markets given advantages offered by access to larger markets. The above results that the IRS manufacturing products face the highest hazard are potentially indicative of the opposite holding. However, the appropriate examination of such a hypothesis entails comparing the hazard of exports to NAFTA members ceasing for each returns-to-scale product separately, to which I now turn. Rather than introducing a number of interacted variables to examine whether the effect of NAFTA is different for different returns-to-scale products, I estimate the hazard of exports ceasing for each of the three returns-to-scale types of products, focusing on the NAFTA-in-effect coefficients, and compare the fitted hazards for each country. In order to conserve space I only present coefficients relevant to NAFTA.[6] Table 4 collects the results.

Intra-NAFTA exports of all three countries in all three returns-to-scale types face a lower hazard. In the case of Canada, the onset of NAFTA has no significant effect on the hazard of exports IRS manufacturing and CRS products ceasing, and only marginally increases the hazard for IRS natural resource products. In the case of Mexico, the net effect of NAFTA which was to increase the hazard associated with exports to NAFTA members (Table 2) seems to be driven by its effect on IRS manufacturing products. The other two types are not affected by the onset of NAFTA. In the case of the US, the onset of NAFTA increases the hazard for all three types of returns to scale.

Figure 8 shows the estimated hazard of exports to NAFTA members ceasing for the different types of returns to scale and the effect of NAFTA on the estimated hazard. I include the 95% confidence interval for the estimated hazard of exports to NAFTA members ceasing. The top panels examine IRS manufacturing products, the middle panels IRS natural resource products, and the bottom panels CRS products. In the case of Canadian exports, despite the fact that the effect of the onset of NAFTA is estimated as significant, the estimated hazard for the onset of NAFTA is not statistically significantly different from that for exports to NAFTA members, with the possible exception of IRS natural resource products.

In the case of Mexico NAFTA does not have a significant effect on the hazard for IRS natural resource and CRS products, as the two estimated

[6] Full results are available on request.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 4 The effect of NAFTA across returns to scale

	IRS manufacturing	IRS natural resources	CRS
		Canada	
NAFTA members	−0.263***	−0.274***	−0.176***
	(0.066)	(0.086)	(0.059)
NAFTA in effect	0.076	0.164*	0.033
	(0.069)	(0.089)	(0.061)
ρ	0.0315***	0.0259***	0.0299***
		Mexico	
NAFTA members	−0.703***	−0.624***	−0.773***
	(0.063)	(0.090)	(0.067)
NAFTA in effect	0.182***	0.082	0.025
	(0.049)	(0.071)	(0.049)
ρ	0.161***	0.174***	0.203***
		US	
NAFTA members	−1.097***	−1.042***	−1.029***
	(0.116)	(0.126)	(0.076)
NAFTA in effect	0.431***	0.457***	0.195***
	(0.120)	(0.130)	(0.078)
ρ	0.119***	0.135***	0.239***

Robust standard errors clustered by relationships in parentheses with *, **, *** denoting significance at 10, 5, and 1%; year in spell and spell number fixed effects included.

hazard are within the 95% confidence interval. However, the effect of NAFTA is statistically significant for IRS manufacturing products. The largest effect that NAFTA has had on the hazard of exports ceasing is for US products across the full spectrum of returns to scale. In the case of each returns-to-scale product type, the difference between exports to NAFTA members in the absence of NAFTA and after its onset is statistically significant. Thus, NAFTA has increased the hazard of US exports to NAFTA members ceasing, by almost 10 percentage points at the start of a spell in the case of IRS products and some 5 percentage points in the case of CRS products.

Table 5 contains the time-dependent effects of NAFTA for each returns to scale type, shedding more light on the exact nature of the effect of NAFTA on the hazard of exports ceasing. For Canada, NAFTA has increased the hazard of IRS manufacturing and CRS products only during its first 3 years, while the hazard for IRS natural resource products was higher in the first 3 years as well as between 2003 and 2005. This is confirmed by plots in Figure 9 where the estimated hazard during these periods is outside the 95% confidence interval.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

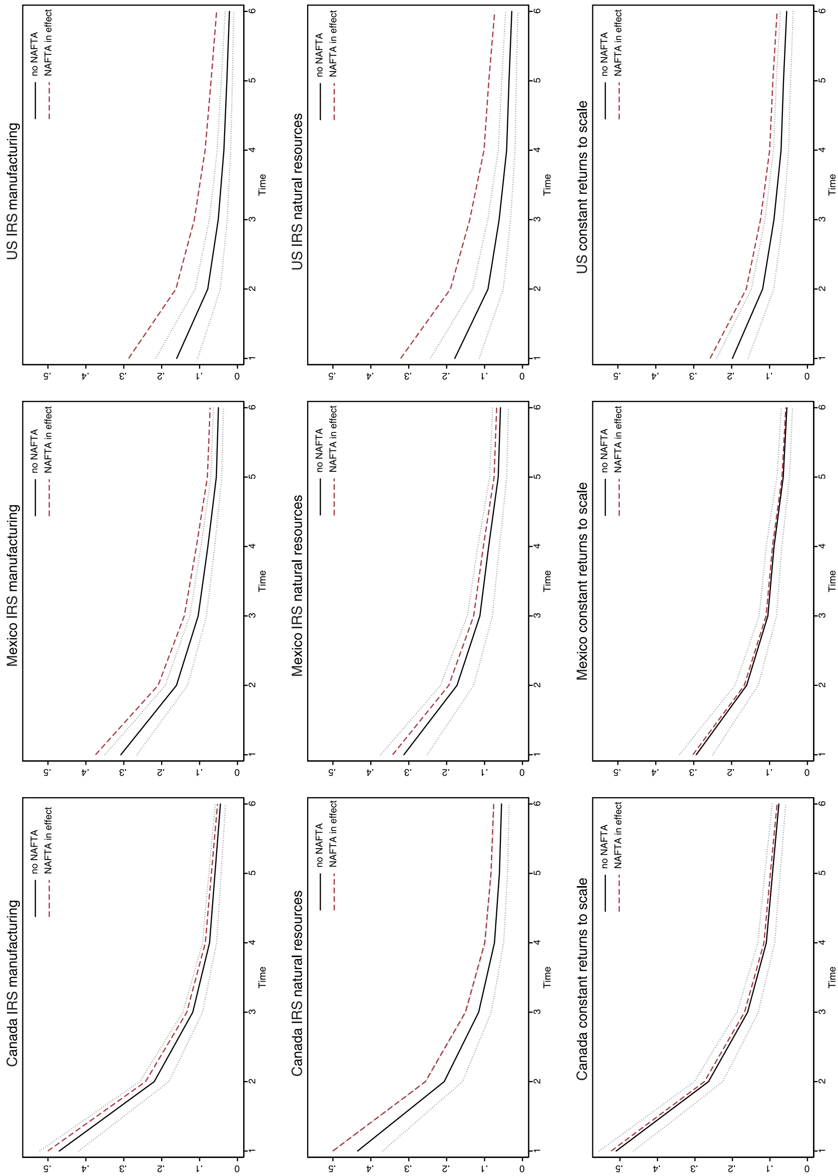

Figure 8 The effect of NAFTA and returns to scale on the estimated hazard.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 5 Time-dependent effect of NAFTA across returns to scale

	IRS manufacturing	IRS natural resources	CRS
	Canada		
NAFTA members	−0.260***	−0.271***	−0.173***
	(0.066)	(0.086)	(0.059)
NAFTA in effect 1994–1996	0.283***	0.371***	0.192**
	(0.086)	(0.112)	(0.077)
NAFTA in effect 1997–1998	0.006	0.110	0.022
	(0.088)	(0.114)	(0.079)
NAFTA in effect 2000–2002	−0.093	−0.065	−0.103
	(0.079)	(0.104)	(0.072)
NAFTA in effect 2003–2005	0.110	0.227**	0.048
	(0.072)	(0.095)	(0.066)
ρ	0.0318***	0.0259***	0.0306***
	Mexico		
NAFTA members	−0.703***	−0.626***	−0.774***
	(0.063)	(0.090)	(0.067)
NAFTA in effect 1994–1996	0.110*	−0.040	−0.136**
	(0.056)	(0.082)	(0.057)
NAFTA in effect 1997–1998	0.214***	0.088	0.044
	(0.056)	(0.080)	(0.056)
NAFTA in effect 2000–2002	0.325***	0.289***	0.196***
	(0.057)	(0.080)	(0.056)
NAFTA in effect 2003–2005	0.098*	0.004	0.021
	(0.057)	(0.082)	(0.057)
ρ	0.159***	0.168***	0.193***
	US		
NAFTA members	−1.092***	−1.041***	−1.035***
	(0.116)	(0.126)	(0.076)
NAFTA in effect 1994–1996	0.525***	0.463***	0.135
	(0.137)	(0.151)	(0.090)
NAFTA in effect 1997–1998	0.244*	0.500***	0.115
	(0.138)	(0.148)	(0.089)
NAFTA in effect 2000–2002	0.637***	0.610***	0.385***
	(0.134)	(0.147)	(0.087)
NAFTA in effect 2003–2005	0.270*	0.232	0.146
	(0.139)	(0.152)	(0.089)
ρ	0.118***	0.135***	0.238***

Robust standard errors clustered by relationships in parentheses with *, **, *** denoting significance at 10, 5, and 1%; year in spell and spell number fixed effects included.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

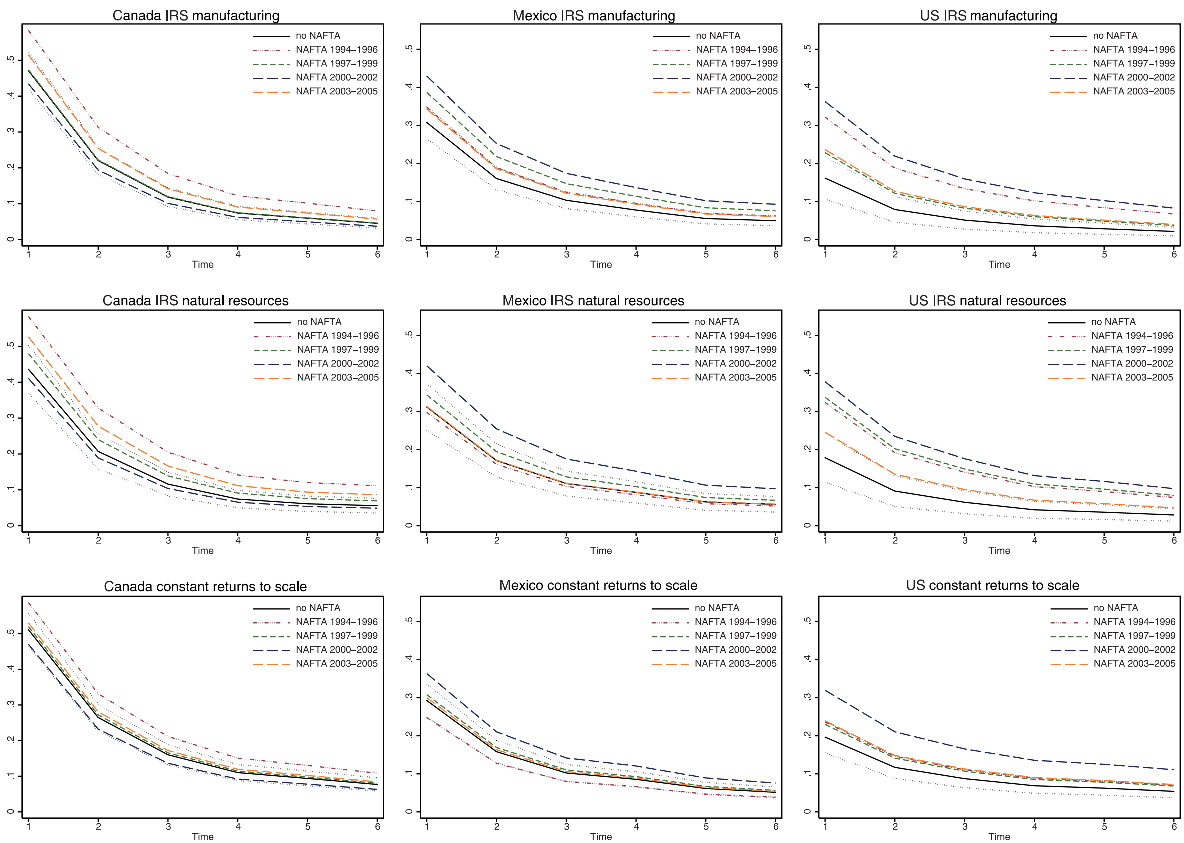

Figure 9 The effect of NAFTA and returns to scale differences.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

For Mexico, NAFTA has increased the hazard for IRS manufacturing products consistently ever since it was enacted, though the effect has declined in magnitude after 2002. The effect is statistically significant for the 1997–2002 period, and only marginally so for the other two periods (Figure 9). NAFTA has increased the hazard for IRS natural resource products only between 2000 and 2002, which is statistically significant as illustrated in Figure 9. NAFTA's effect on the hazard of Mexican exports of CRS products is the most varied. It has initially reduced it with the estimated hazard right on the lower bound of the 95% confidence interval, and then increased it between 1997 and 2002, a statistically significant effect. NAFTA has had no effect since 2003.

NAFTA's effect on the hazard of US exports to NAFTA members is the most consistent one, having increased the hazard for every type of product in almost every year. Exports of CRS products have a higher hazard of ceasing in a statistically meaningful manner only for the 2000–2002 period. Exports of IRS natural resource products have had a higher hazard in every period, with differences statistically significant, though only marginally for the 2003–2005 period. Exports of IRS manufacturing products had a statistically significantly higher hazard in every period, with the effect stronger immediately after NAFTA's onset and between 2000 and 2002.

5 Conclusion

In this article, I investigate how NAFTA has affected the members' hazard of exports ceasing and how differences in returns to scale manifest themselves in the hazard of exports ceasing for the three members: Canada, Mexico, and the US. NAFTA itself has not had a beneficial effect on the hazard of exports ceasing. Rather, the effect has been negative for the US and Mexico, with the hazard of their exports ceasing to NAFTA members increasing with the enactment of NAFTA. However, the said increase was not large enough to offset the much lower hazard of exports ceasing that NAFTA members enjoy due to the geography of the free-trade area and proximity of the members to each other. Canada, Mexico, and the US enjoy a significantly lower hazard on exports to each other without the presence of NAFTA. Given the particular geography of NAFTA, the effect of common borders between the members, a well-known positive force in international trade, is largely indistinguishable from the effect of NAFTA. The nature of the geography of NAFTA makes it difficult to broadly conclude that free trade agreements increase the hazard of exports between the members ceasing, calling for an investigation of the effect of

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

free trade agreements with less restrictive geographic characteristics, or a broader set of such agreements.

The effect of NAFTA is much stronger once one scratches below the surface, in terms of evaluating its effect on each returns to scale type. It has had the most consistent effect on US exports of all three types, particularly in the case of both increasing returns to scale product types. In addition, it has had different effects during different subperiods since its inception, likely reflecting the ability of firms to adjust to new conditions and the fact that some of its provisions were phased in over time.

I presented the first evidence of the effect of a free trade agreement on the hazard of exports ceasing. While NAFTA increases the hazard, further investigation is needed with free trade agreements among countries which are not as geographically clustered as the NAFTA members are. Mercosur and the European Union are two free trade areas which offer a different geography that could shed additional light on the role of a free trade agreement. In addition, I presented the first evidence on the effect of the returns to scale on the hazard of exports ceasing. Unlike differences along the product differentiation dimension, which are largely consistent across a number of countries, the identified effects of returns to scale are exporter specific. Since these results are based on three exporters only, additional investigation of other countries is warranted.

References

Antweiler, W. and D. Trefler (2002), ‘‘Increasing Returns and All That: A View from Trade’’, *American Economic Review* **92**, 93–119.

Baier, S. L. and J. H. Bergstrand (2007), ‘‘Do Free Trade Agreements Actually Increase Members’ International Trade?’’, *Journal of International Economics* **71**, 72–95.

Besedeš, T. (2008), ‘‘A Search Cost Perspective on Formation and Duration of Trade’’, *Review of International Economics* **16**, 835–849.

Besedeš, T. (2011), ‘‘Export Differentiation in Transition Economies’’, *Economic Systems* **35**, 25–44.

Besedeš, T. and T. J. Prusa (2006a), ‘‘Ins, Outs, and the Duration of Trade’’, *Canadian Journal of Economics* **39**, 266–295.

Besedeš, T. and T. J. Prusa (2006b), ‘‘Product Differentiation and Duration of U.S. Import Trade’’, *Journal of International Economics* **70**, 339–358.

Besedeš, T. and T. J. Prusa (2010), ‘‘The Duration of Trade Relationships’’, in B. Hoekman and G. G. Porto, eds., *Trade*

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Adjustment Costs in Developing Countries: Impacts, Determinants and Policy Responses, World Bank, Washington, DC, pp. 265–282.

Besedeš, T. and T. J. Prusa (2011), "The Role of Extensive and Intensive Margins and Export Growth", *Journal of Development Economics* **96**, 371–379.

Brenton, P., C. Saborowksi and E. von Uexkull (2010), "What Explains the Low Survival Rate of Developing Country Export Flows?", *World Bank Economic Review* **24**, 474–499.

Cadot, O., L. Iacovone, F. Rauch and M. Denisse Pierola (2011), "Success and Failure of African Exporters", World Bank Policy Research Working Paper No. 5657.

Clausing, K. (2001), "Trade Creation and Trade Diversion in the Canada–United States Free Trade Agreement", *Candian Journal of Economics* **92**, 889–904.

Debaere, P. and S. Mostashari (2010), "Do Tariffs Matter for the Extensive Margin of International Trade? An Empirical Analysis", *Journal of International Economics* **81**, 163–169.

Fugazza, M. and A. Cristina Molina (2009), "The Determinants of Trade Duration", HEI Working Papers No. 5.

Görg, H., R. Kneller and B. Muraközy (2012), "What Makes a Successful Export? Evidence from Firm–product Level Data", *Canadian Journal of Economics*, forthcoming.

Hess, W. and M. Persson (2011a), "Exploring the Duration of EU Imports", *Review of World Economics* **147**, 665–692.

Hess, W. and M. Persson (2011b), "The Duration of Trade Revisited: Continuous-Time vs. Discrete-Time Hazards", *Empirical Economics*, doi: 10.1007/s00181-011-0518-4.

Iacovone, L., F. Rauch and A. J. Winters (2010), "Trade as Engine of Creative Destruction: Mexican Experience with Canadian Competition", London School of Economics, CEP DP 999.

Jaud, M., M. Kukenova and M. Strieborny (2009), "Financial Dependence and Intensive Margin of Trade", Paris School of Economics Working paper 2009–35.

Kehoe, T. J. and K. J. Ruhl (2009), "How Important is the New Goods Margin in International Trade?", Federal Reserve Bank of Minneapolis Research Department Staff Report 324.

Melitz, M. (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity", *Econometrica* **71**, 1695–1725.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Nitsch, V. (2009), ''Die Another Day: Duration in German Import Trade'', *Review of World Economics* **145**, 133–154.

Rauch, J. E. and J. Watson (2003), ''Starting Small in an Unfamiliar Environment'', *International Journal of Industrial Organization* **21**, 1021–1042.

Romalis, J. (2007), ''NAFTA's and CUSFTA's Impact on International Trade'', *Review of Economics and Statistics* **89**, 416–435.

Trefler, D. (2004), ''The Long and Short of the Canada–U.S. Free Trade Agreement'', *American Economic Review* **94**, 870–895.

Appendix A

Data	Source
US exports	US census bureau, US imports CDs and DVDs
Canadian and Mexican exports	UN Comtrade
GDP	World Bank's World Development Indicators
Distance, contiguity, common language, and colonial history	CEPII http://www.cepii.fr/anglaisgraph/bdd/gravity.htm
Returns to scale classification	Antweiler and Trefler (2002)
Economic risk	International risk guide http://www.prsgroup.com/ICRG.aspx

A1. Data Appendix

Data used in this article are available from public sources.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

CESifo Economic Studies, Vol. 59, 2/2013, 337–359 doi:10.1093/cesifo/ifs029
Advance Access publication 18 July 2012

Investment Promotion and FDI Inflows: Quality Matters

Torfinn Harding*,† and Beata S. Javorcik‡

*Department of Economics, University of Oxford and CESifo, Manor Road Building, Manor Road, Oxford OX1 3UQ, UK
†Oxford Centre for the Analysis of Resource Rich Economies and Statistics Norway. e-mail: torfinn.harding@economics.ox.ac.uk
‡Department of Economics, University of Oxford, CEPR and CESifo, Manor Road Building, Manor Road, Oxford OX1 3UQ, UK. e-mail: beata.javorcik@economics.ox.ac.uk

Abstract

Information asymmetries constitute a significant obstacle to capital flows across international borders, and in particular to flows of foreign direct investment (FDI) to emerging markets. Many governments aim to reduce information barriers by engaging in investment promotion activities. Despite potentially large benefits of FDI and popularity of investment promotion intermediaries (IPIs), relatively little is known about their effectiveness. This study uses data collected through the Global Investment Promotion Benchmarking (GIPB) exercise to examine whether higher quality of IPI services translates into higher FDI inflows. The analysis, based on information on 156 countries, suggests that countries with IPIs able to handle investor inquiries in a more professional manner and IPIs possessing higher quality Web sites tend to attract greater volume of FDI. These results are robust to using sector-level data and instrumental variable approach. (JEL codes: F21, F23, H11)

Keywords: investment promotion, foreign direct investment, industrial policy, investment promotion intermediaries

1 Introduction

The benefits of global economic integration have become increasingly evident over the last decades. Increased movement of goods, services, people, and capital across international borders has helped many developing countries achieve fast and sustained economic growth. Many observers argue that foreign direct investment (FDI) has been a key ingredient in this process by facilitating transfer of productivity-enhancing techniques and knowledge from developed to developing countries (e.g., Hoekman and Javorcik 2006).

The theoretical and empirical literature suggests that information asymmetries constitute a significant obstacle to capital flows across international borders.[1] Informational asymmetries between domestic and

[1] Gordon and Bovenberg (1996, p. 1059) argue that 'Investors, by living and working in a particular country, know much more about the economic prospects of that country than they do about those in other countries. When foreigners try to acquire a firm in the country, they can easily end up being overcharged by domestic owners, who have access to better information not only about that specific firm, but also about future government policies affecting the firm. ... Foreigners' lack of knowledge can result also in a less efficient use of real resources, due for example to their poorer ability ... to deal with idiosyncratic aspects of the domestic contract law ... and local customs governing labor relations.'

© The Author 2012. Published by Oxford University Press on behalf of Ifo Institute, Munich. All rights reserved.
For permissions, please email: journals.permissions@oup.com

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

foreign investors have been hypothesized to be a possible explanation for home bias, the tendency of investors to invest less in foreign equities relative to the prediction of a portfolio choice model (Stulz 1981; Ahearne et al. 2005).[2] The negative effects of information asymmetries on capital flows have been documented in empirical studies (Portes et al. 2001; Portes and Rey 2005; Gelos and Wei 2005). Moreover, Daude and Fratzscher (2008) have shown that FDI flows are 'substantially more sensitive to information frictions than investment in portfolio equity and debt securities.' Information asymmetries are also the reason why the theoretical model by Gordon and Bovenberg (1996) suggests that a capital-importing country would raise welfare by subsidizing foreign direct investment and other capital inflows from abroad.

Being aware of the fact that lack of information constitutes a barrier for inflows of FDI, many governments engage in investment promotion activities. The purpose of such activities is to reduce transactions costs facing foreign investors by providing information on the host country, helping foreign investors deal with bureaucratic procedures, and offering fiscal or other incentives. There are more than 189 investment promotion intermediaries (IPIs) at the national level and over a thousand at the sub-national level.[3] Public funding of investment promotion activities is justified on the grounds that the presence of FDI generates knowledge externalities. This belief is confirmed by recent empirical evidence suggesting that FDI leads to positive productivity spillovers to local firms in the supplying industries.[4]

The small existing literature on investment promotion suggests that investment promotion is a cost effective way of attracting FDI to developing countries (Harding and Javorcik 2011). At the same time, the results for industrialized countries appear to be mixed. Although Bobonis and Shatz (2007) and Charlton and Davis (2006) provide evidence suggesting that investment promotion is associated with higher FDI inflows into developed countries, Head et al. (1999) and Harding and Javorcik (2011) do not find any significant effect of investment promotion efforts in developed countries.

The main shortcoming of the above literature is its reliance on crude measures of investment promotion, such as presence of IPIs in the investor's home country or information on sectors targeted by a particular host

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

[2] For evidence on home bias, see, for instance, French and Poterba (1991) and Tesar and Werner (1995).

[3] IPIs are also referred to as IPAs (investment promotion agencies).

[4] See studies by Javorcik (2004a); Blalock and Gertler (2008); Javorcik and Spatareanu (2008, 2009, 2011); and literature reviews by Görg and Strobl (2001) and Görg and Greenaway (2004).

country. These measures do not take into account the fact that IPIs vary widely in terms of the quality of services they provide. Simply put, not all IPIs are equally good at providing relevant business data to prospective investors.

This study aims to fill this gap in the literature by using data collected by the World Bank's Foreign Investment Advisory Services through the Global Investment Promotion Benchmarking (GIPB) series 2006–2012 to assess how much the quality of IPI work may affect FDI inflows. GIPB data are particularly suitable for the purpose of testing the importance of the quality of investment promotion. They capture how well IPIs perform in facilitating site selection by providing potential investors with information needed to determine the location for their project. GIPB assesses two aspects of IPI facilitation. The first one is the quality of the agency's Web site, which is rated based on its content, architecture, design, and promotional effectiveness. Web sites are judged on whether they contain relevant, clear, and credible information presented in an attractive and user-friendly way. The second rating focuses on the way IPIs handle direct project inquiries from investors. This rating captures competence and responsiveness of the agency's staff, including timeliness, quality, and credibility of informational content.

Our empirical analysis suggests that these differences in IPI quality translate into different levels of FDI inflows. In other words, countries with IPIs that appear to perform better at their core function attract more FDI. We find a positive and statistically significant relationship between the average inflow of FDI during years 2000–2010 and the average quality of the national IPI. In the analysis, the IPI quality is measured using the GIPB results obtained by GIPB 2006, GIPB 2009, and GIPB 2012.[5] The positive relationship between IPI performance and FDI inflow is obtained controlling for the average level of GDP per capita, GDP growth, population size, inflation, and political stability observed in the host country during the period considered. The analysis is based on 156 countries for which the necessary data are available. The effect of IPI performance is statistically significant at the 1 or the 5% level.

The magnitude of the estimated effect is economically meaningful. A country with the IPI quality score of 60% received on average 25% higher FDI inflows than a country with an IPI receiving the score of 45% (controlling for the country-specific characteristics mentioned above). In other words, a one unit increase in the GIPB score was associated

[5] For summary of these results, see GIPB reports at www.globalinvestmentpromotion.com. If an agency was not rated in all 3 years, the available information was used to compute the average performance score.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

with a 1.5% increase in FDI inflows. Thus, for example, holding everything else equal, countries with agencies with the average GIPB performance score of the Latin America and Caribbean region received 40% more FDI than countries where the GIPB score was equal to the average for sub-Saharan Africa.

A series of robustness checks confirms our main conclusions. First, we show that the results are robust to focusing on just developing countries. Second, we show that the findings hold when we control for various aspects of the business climate. Third, we find that using sector-specific information supports our findings. And finally, we demonstrate that our results hold when we instrument for the quality of IPIs.

Two conclusions emerge from this analysis. First, good investor facilitation matters. But it is not enough to set up an IPI and expect a huge boom in FDI inflows. Successful investment promotion requires professionalism, effort, and commitment to customer service. It requires maintaining an up-to-date, attractive, and user-friendly Web site that includes relevant and useful information that an investor requires during the site selection process. Providing the necessary data to support this decision process makes a difference. Second, the GIPB initiative and its results can be a valuable guiding tool for IPIs striving to achieve excellence and secure more FDI flows. The GIPB criteria specify what high quality of inquiry handling (IH) and Web sites mean, and its assessment process can provide useful feedback on what areas need improvement.

The article is structured as follows. The next section explains functions of investment promotion, reviews evidence on its effectiveness, and documents large differences in IPI quality across countries. Section 3 describes and presents the empirical results at the national level, country–sector level as well as the instrument variable approach. The last section presents the conclusions.

2 What Do We Know about Investment Promotion?

2.1 What is investment promotion?

In order to attract foreign investors, many countries have set up investment promotion intermediaries. There are more than 189 IPIs at the national level and over a thousand at the sub-national level. Their main role is to reduce the costs of undertaking FDI by providing information to potential investors and by alleviating the burden of bureaucratic procedures. Information provision takes the form of marketing campaigns, participation in international conferences and fairs, setting up informational Web sites and actively pursuing investors through phone, mail, or personal contacts. It might also take the form of assisting investors with site

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

visits and introducing them to potential joint venture partners, customers, and suppliers. Alleviating the burden of red tape takes place through assisting committed investors with the registration process, license applications, or any other formalities that may be required. For instance, Singapore's IPI is a one stop shop in terms of bureaucratic procedures. The agency will either deal directly with the investor's registration request or it will guide the investor in the next steps. This reduces the set up costs and hence the costs of investing in Singapore.

Investment promotion can affect the choice of a foreign investment site at all stages of the decision-making process. A process of a site selection usually starts with drawing a long list of potential locations. This list includes 8–20 economies and often contains (i) the most popular FDI destinations, (ii) countries located in proximity to the investor's existing operations, and (iii) emerging FDI destinations. The 3rd category presents an opportunity for countries which so far have had a limited success in attracting FDI. By means of advertising, presentations at trade shows or pro-active contacting of potential investors IPIs can increase the chances of their country being included in the long list. When the long list is then narrowed down to about five potential locations, IPIs again play an important role. Short listing is usually done without visiting sites under consideration; therefore, the accessibility of the information about potential host countries plays a crucial role. Agencies that provide up-to-date, detailed, and accurate data on their Web sites, and agencies that are willing to invest time in preparing detailed answers to investors' inquiries and customize these answers to the needs of an individual investor can increase the chances of their countries being included in the short list. The next step, investor's visit to the host country, also gives IPIs the opportunity to emphasize the advantages of locating in their country, answer questions, show off potential investment sites, and facilitate contacts with local business community. Finally, IPIs can play a role in the last stage of the process by providing information on investment incentives and offering help with the registration process.

2.2 Does investment promotion work?

The work of Harding and Javorcik (2011) provides evidence on the effectiveness of investment promotion efforts. The authors use data collected by the World Bank's 2005 Census of Investment Promotion Agencies covering 109 IPIs around the world. Particularly useful for their research purposes is the detailed information on sectors denoted as priority in national investment promotion efforts. The Census gathered information on which sectors were targeted by IPIs and when this targeting started and stopped. Sector targeting is viewed by investment promotion professionals as the

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

most effective way of attracting FDI. It is believed that efforts tailored to the needs of investors operating in a particular sector will work better than attempts to target all potential investors.

Harding and Javorcik combine the information on sector targeting with the Bureau of Economic Analysis (BEA) data on US FDI abroad. The focus on US FDI is driven by the fact that the BEA is the only comprehensive source of FDI data with country–sector–time variation and extensive coverage of developing countries. Considering developing countries is of interest since many of them struggle with attracting FDI. Moreover, not having a proper set of developing countries in the sample would threaten the external validity of the econometric approach. The final sample consists of 124 countries over the period 1990–2004. This is more than the number of countries covered by the Census because some economies are not engaged in investment promotion.

Using a difference-in-differences specification, the authors ask whether sectors explicitly targeted by IPIs received more FDI than non-priority sectors during the same time period. Their analysis relies on within country–sector variation over time as the empirical specification controls for country–sector, country–year, and sector–year fixed effects.

The results suggest that sectors targeted by IPIs receive on average more than twice as much FDI inflows than non-targeted sectors. The budget information collected in the IPI survey allows for a back-of-the-envelope comparison between the costs and benefits of investment promotion: 1 dollar spent on investment promotion raises FDI inflows by 189 dollars and that one extra job in an US affiliate requires 78 dollars spent on investment promotion.[6]

The results further suggest that investment promotion works better in countries where English is not an official language and in countries which are culturally distant from the USA. Investment promotion has also larger effects in countries with less effective or corrupt governments and in countries where it takes longer to start a business or obtain a construction permit. Finally, investment promotion is effective mainly in

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

6 As emphasized in the article, this cost-benefit calculation should be treated with caution. On the one hand, the calculations capture only the effect of targeting on flows of FDI from the USA. As investment promotion is likely to have a similar impact on investors from other source countries, the analysis underestimates the benefits of investment promotion activities. On the other hand, there may be other factors that contribute to the success of investment promotion and whose costs were not accounted for (for instance, access to accelerated bureaucratic procedures for targeted sectors). Furthermore, the analysis captures the average, not the marginal, effect. In other words, the results should not be interpreted as suggesting that a large increase in investment promotion spending in countries already engaged in such practice will lead to huge increases in FDI inflows. Instead, the results should be viewed as indicating that countries not involved in investment promotion may benefit from such activities.

developing countries. All these findings support the view that investment promotion reduces information asymmetries between foreign investors and host countries and that it alleviates the burden of red tape.

Harding and Javorcik rely on data measuring the mere presence of investment promotion efforts. The data used in this study allow us to take into account also the quality of the information provision of IPIs. As documented in the next subsection, there are large differences in the performance of IPIs in this respect.

2.3 Differences in IPI quality

The results of GIPB 2012 reveal huge differences in performance of various IPIs (Figure 1). The Web sites of only three IPIs have received close to perfect scores (95–97%). Although the top 50 Web sites received scores above 80%, 24 IPIs received a positive score below 30% and 13 agencies obtained a zero score. A zero score means that an agency has no online presence. Web sites received low scores if they were not available in English, which is generally recognized as the language of international business.

The quality of responses to project inquiries received much lower ratings. The top two scores were 81 and 88%. The majority of agencies received a rating below 50%, which means they are of limited assistance in providing the information that the foreign investor needs. In many cases, IPIs were not contactable by the foreign investor researching from overseas. Interestingly, IPIs with highly rated Web sites vary widely in how well they handle inquiries. There are quite a few agencies whose Web site obtained a score above 80% and who received a score below 40% (or even 20%) for IH.

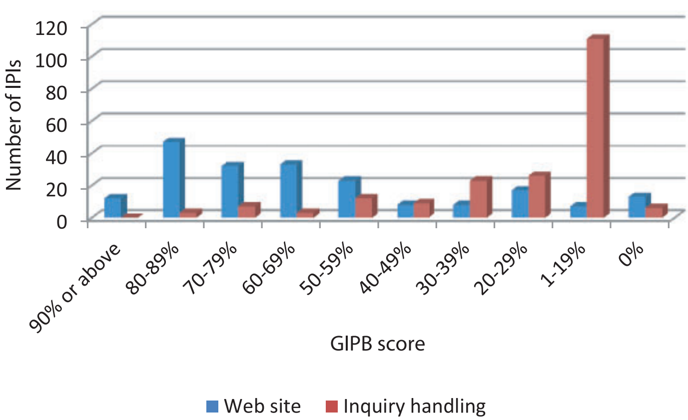

Figure 1 Distribution of IPI scores.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

3 Does IPI Quality Matter?

3.1 Analysis based on aggregate data

Our analysis of the link between the quality of IPI services and FDI inflows relies on the following data. Our dependent variable is the average FDI inflow received by country c during the 2000–2010 period as reported in the IMF's *International Financial Statistics* and expressed in the log form. The choice of time period is determined by the availability of data on IPI quality which is measured as the average GIPB score obtained in GIPB 2012, GIPB 2009, and GIPB 2006 exercises.[7] All the available scores are used, which means that if a particular agency was rated only once, we assume that this rating captured its performance during the 2000–2010 period. We distinguish between the total score, the Web site score, and the IH score. The IH score is the average of the two IH scores obtained in each exercise. We also include the difference between the two IH scores, as the lack of consistency in IH quality presumably deters foreign investors. Our sample covers 156 countries.[8]

Our empirical model controls for a set of host country characteristics: GDP per capita (average value for 2000–2010, expressed in logs), GDP growth (average value for 2000–2010), population size (average value for 2000–2010, expressed in logs), inflation (average value for 2000–2010), and political stability (average value for 2000–2009). The last variable comes from the Worldwide Governance Indicators project, while all other controls are from the World Bank's World Development Indicators database. For summary statistics, see Appendix Table A1 at the end of this article.

The results of the analysis, reported in Table 1, suggest that there exists a positive and statistically significant relationship between IPI quality and FDI inflows. This is true when we use the total GIPB score, the IH score, or the Web site score. The difference between the IH scores (proxy for lack of consistency in the service quality) does not appear to matter. The Web site score carries a higher significance level (1%) than the IH score (10%) and is more robust to different specifications.

The magnitude of the estimated effect is economically meaningful. A country with the IPI quality score of 60% received on average 25% higher FDI inflows than a country with an IPI receiving the score of 45%.[9] In other words, a one unit increase in the GIPB score was associated with a 1.5% increase in FDI inflows. As illustrated in Figure 2, GIPB scores vary

[7] These are the only waves of GIPB exercise conducted.

[8] Note that we excluded Suriname, which having experienced only FDI outflows, was an outlier in the sample. Including Suriname would not change the conclusions of the analysis, and it would increase the magnitudes of the estimated coefficients of interest thus strengthening our conclusions.

[9] This statement is based on Column 1 in Table 1.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

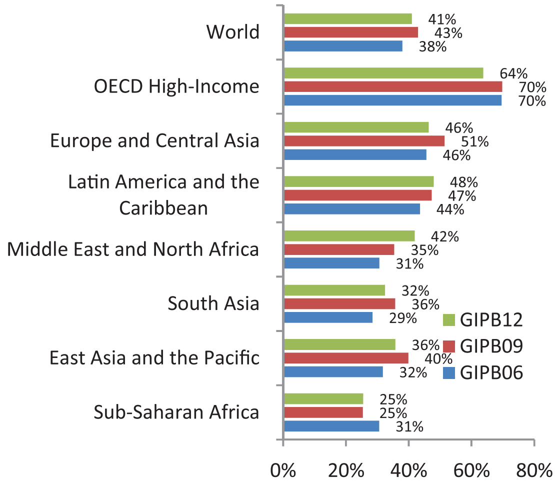

Figure 2 GIPB 2006–2012 scores by region.

widely by region. Thus, for instance, holding everything else equal, countries with agencies with the average GIPB performance score of the Latin America and Caribbean region received 40% more FDI than the flows attracted to countries where the GIPB score was at the level equal to the average for sub-Saharan Africa, 26% more than countries with an average GIPB score of South Asia, 22% relative to Middle East and North Africa, and 4% relative to Europe and Central Asia.

One may be concerned that the results in Table 1 are driven by differences in average score between developed and developing countries. This does not appear to be the case. We find a positive relationship between IPI quality in the full sample as well as in the sub-sample of developing countries. The same is true when we consider the IH score and the Web site score. Table 2 presents the estimation results for the sub-sample of developing countries. As before, we find that the overall IPI quality and the Web site quality are positively and significantly correlated with the amount of FDI inflows. Although the IH quality bears a positive sign in most specifications, it never appears to be statistically significant.

To examine whether our results are really capturing the quality of IPI rather than the general quality of the business environment, we conduct a series of robustness checks. We do so by adding to the baseline specification controls for various aspects of the business climate, which include the number of days needed to start a business, the number of days needed to obtain a construction permit, and the number of days needed to register a property. All of these variables come from the World Bank's *Doing*

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 1 Baseline specification. All countries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IPI quality	0.015*** (0.006)							
IH quality		0.010* (0.005)			0.005 (0.007)	0.012* (0.006)	0.003 (0.006)	
Difference between IH scores			0.001 (0.007)		−0.006 (0.007)	−0.005 (0.007)		−0.004 (0.007)
Web site quality				0.012*** (0.004)	0.011** (0.005)		0.011** (0.005)	0.013*** (0.005)
GDP per capita	0.983*** (0.082)	1.036*** (0.079)	1.097*** (0.073)	0.988*** (0.081)	0.976*** (0.083)	1.034*** (0.079)	0.980*** (0.082)	0.989*** (0.081)
GDP growth	12.720*** (3.781)	12.795*** (3.864)	11.650*** (3.866)	11.908*** (3.756)	12.200*** (3.831)	12.753*** (3.872)	12.272*** (3.826)	11.719*** (3.777)
Population	0.733*** (0.045)	0.742*** (0.045)	0.751*** (0.046)	0.732*** (0.045)	0.728*** (0.045)	0.740*** (0.046)	0.732*** (0.045)	0.730*** (0.045)
Inflation	−0.748 (1.033)	−0.731 (1.047)	−0.644 (1.059)	−0.706 (1.032)	−0.780 (1.038)	−0.773 (1.051)	−0.726 (1.035)	−0.735 (1.035)
Political stability	0.110 (0.133)	0.108 (0.136)	0.137 (0.136)	0.131 (0.133)	0.121 (0.134)	0.107 (0.136)	0.122 (0.134)	0.134 (0.133)
Constant	−0.146 (0.835)	−0.331 (0.843)	−0.648 (0.836)	−0.223 (0.828)	−0.064 (0.846)	−0.255 (0.852)	−0.167 (0.836)	−0.174 (0.833)
Observations	156	156	156	156	156	156	156	156
R^2	0.83	0.83	0.83	0.84	0.84	0.83	0.84	0.84

Notes. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% level, respectively. All variables are based on country averages over the period 2000–2010. The dependent variable is the log of the mean of aggregate inflow of foreign direct investment from all source countries. IPI quality, IH quality, and Web site quality are measured on a scale from 0 to 100. GDP per capita and population are included in the log form. Political stability ranges from −2.5 to 2.5, with a mean of 0 and standard deviation of 1. A higher value means more stability. The measure is available until 2009.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 2 Baseline specification. Developing countries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IPI quality	0.011* (0.006)							
IH quality		0.005 (0.006)			0.002 (0.008)	0.009 (0.008)	−0.002 (0.007)	
Difference between IH scores			−0.001 (0.008)		−0.010 (0.010)	−0.008 (0.010)		−0.009 (0.008)
Web site quality				0.009** (0.004)	0.011** (0.005)		0.010** (0.005)	0.011** (0.005)
GDP per capita	1.128*** (0.095)	1.177*** (0.093)	1.211*** (0.089)	1.124*** (0.093)	1.127*** (0.095)	1.176*** (0.094)	1.130*** (0.095)	1.131*** (0.093)
GDP growth	13.314*** (3.886)	13.338*** (3.988)	12.669*** (3.945)	12.663*** (3.850)	12.551*** (3.962)	13.510*** (3.998)	12.382*** (3.958)	12.365*** (3.859)
Population	0.771*** (0.052)	0.776*** (0.052)	0.777*** (0.053)	0.768*** (0.052)	0.764*** (0.052)	0.773*** (0.053)	0.768*** (0.052)	0.764*** (0.052)
Inflation	−1.164 (0.954)	−1.150 (0.967)	−1.121 (0.971)	−1.132 (0.948)	−1.235 (0.960)	−1.252 (0.976)	−1.116 (0.953)	−1.212 (0.951)
Political stability	0.156 (0.138)	0.152 (0.140)	0.157 (0.141)	0.161 (0.137)	0.172 (0.138)	0.159 (0.141)	0.163 (0.138)	0.172 (0.138)
Constant	−1.563 (1.004)	−1.715* (1.015)	−1.824* (1.011)	−1.558 (0.996)	−1.483 (1.007)	−1.635 (1.021)	−1.581 (1.002)	−1.508 (0.997)
Observations	114	114	114	114	114	114	114	114
R^2	0.82	0.82	0.82	0.82	0.83	0.82	0.82	0.83

Notes. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% level, respectively. All variables are based on country averages over the period 2000–2010. The dependent variable is the log of the mean of aggregate inflow of foreign direct investment from all source countries. IPI quality, IH quality, and Web site quality are measured on a scale from 0 to 100. GDP per capita and population are included in the log form. Political stability ranges from −2.5 to 2.5, with a mean of 0 and standard deviation of 1. A higher value means more stability. The measure is available until 2009. The definition of developing countries is based on the 2011 World Bank country classification.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Business Database and pertain to 2003–2010. The second set of controls captures government effectiveness, control of corruption, regulatory quality, rule of law, and voice and accountability. These measures were compiled at the World Bank by Kaufmann, Kraay, and Mastruzzi and are described in detail in their 2009 publication. Each measure is a composite index extracting information on governance from 35 different sources. The authors assume that the available individual governance ratings reflect both some true but unobserved level of governance and sampling variations and perception errors. The unobserved 'true' level of governance can be backed out statistically (assuming a linear unobserved component specification). The resulting estimates range from -2.5 to 2.5, with a mean of 0 and standard deviation of 1. The higher the estimate for each country, the better governed the country. The measures are available for 2000 and annually for 2002–2009. We use the average value in our model.

The measure of corruption captures 'perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.' The measure of government effectiveness captures 'perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.' Rule of law encapsulates 'perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.' Regulatory quality summarizes 'perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.' Voice and accountability measure reflects 'perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.'[10]

The various measures are highly correlated; therefore, they enter the specification one by one. In Table 3, we show that the link between FDI inflows and the overall IPI quality is robust to controlling for each of the eight aspects of business environment considered. In all regressions, the coefficient on IPI overall quality is positive and statistically significant (Panel A). Its magnitude remains pretty stable across different specifications. As illustrated in Panel B, the results for the IH quality are reasonably robust as well. Its coefficient is positive and statistically significant in five of eight specifications. In the case of Web site quality, the positive

[10] See http://info.worldbank.org/governance/wgi/resources.htm for more information.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 3 IPI quality with controls for different aspects of business climate. All countries

	Starting business (1)	Construction permits (2)	Registering property (3)	Voice and accountability (4)	Government effectiveness (5)	Regulatory quality (6)	Rule of law (7)	Control of corruption (8)
Panel A: IPI quality								
IPI quality	0.015** (0.006)	0.015*** (0.006)	0.015*** (0.006)	0.014** (0.006)	0.015** (0.006)	0.011* (0.006)	0.018*** (0.006)	0.017*** (0.006)
Business climate	−0.096 (0.118)	0.107 (0.155)	0.002 (0.081)	0.088 (0.149)	0.002 (0.215)	0.301 (0.205)	−0.299 (0.205)	−0.151 (0.178)
Observations	154	154	153	156	156	156	156	156
R^2	0.83	0.83	0.84	0.84	0.83	0.84	0.84	0.84
Panel B: IH quality								
IH quality	0.009* (0.006)	0.010* (0.006)	0.010* (0.006)	0.008 (0.006)	0.009 (0.006)	0.005 (0.006)	0.012** (0.006)	0.011* (0.006)
Business climate	−0.124 (0.119)	0.076 (0.157)	−0.005 (0.082)	0.148 (0.150)	0.093 (0.215)	0.397* (0.204)	−0.213 (0.205)	−0.109 (0.181)
Observations	154	154	153	156	156	156	156	156
R^2	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
Panel C: Web site quality								
Web site quality	0.012** (0.005)	0.013*** (0.004)	0.012*** (0.004)	0.011** (0.005)	0.012** (0.005)	0.009* (0.005)	0.014*** (0.005)	0.013*** (0.005)
Business climate	−0.089 (0.119)	0.130 (0.155)	−0.003 (0.080)	0.115 (0.144)	0.034 (0.210)	0.322 (0.195)	−0.277 (0.202)	−0.112 (0.175)
Observations	154	154	153	156	156	156	156	156
R^2	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84

Notes. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% level, respectively. All variables are based on country averages over the period 2000–2010. The dependent variable is the log of the mean of the aggregate inflow of foreign direct investment from all source countries. IPI quality, IH quality, and Web site quality are measured on a scale from 0 to 100. All specifications include the following controls: log GDP per capita, GDP growth, log population, inflation, and political stability. The set of business climate measures includes: the number of days required to start a business, obtain a construction permit or register a property (all in logs), an index of voice and accountability, government effectiveness, regulatory quality, rule of law, and control of corruption. The latter indices range from −2.5 to 2.5, with a mean of 0 and standard deviation of 1. They are available until 2009. Higher values imply greater stability, more voice and accountability, etc.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

relationship between FDI inflows and the variable of interest holds and is statistically significant in all eight specifications (Panel C).

3.2 Analysis based on country–sector data

In an alternative empirical strategy, we use data on inflows of US FDI disaggregated by host country and sector.[11] In this difference-in-differences approach, we compare flows to sectors treated as priority by IPIs in their investment promotion efforts (targeted sectors) to non-targeted sectors, which allows us to control for unobservable host country heterogeneity. We interact the dummy for targeted sectors with the measures of IPI quality. In this specification, we normalize the quality measures by subtracting the average value for the sample. More formally, we estimate the following model:

$$\ln(\text{FDI inflow}_{ci}) = \alpha_1 + \beta_1 \text{ Sector targeted}_{ci} + \beta_2 \text{ Sector targeted}_{ci} \times \text{IPI quality normalized}_{ci} + \gamma_c + \gamma_i + \varepsilon_{ci}.$$

The dependent variable is the natural log of the average inflow of US FDI into sector i in country c during the 2000–2010 period. The data come from the US BEA that publishes the stocks of US FDI abroad disaggregated into 15 sectors.[12] We use the first difference of the stocks to calculate flows. Sector targeted$_{ci}$ equals one if country c targeted sector i during this time period and zero otherwise. The information on targeted comes from the World Bank's Census of Investment Promotion Agencies conducted in 2005 and is described in detail in Harding and Javorcik (2011). We assume that sectors targeted at the time of the Census (2004) continued being targeted until 2010. If a sector was targeted during some (but not all) years during 2000–2004 (for which time-varying information is available) we define Sector targeted as the fraction of the time period when targeting was in place. γ_c and γ_i are country and industry fixed effects, respectively. The former makes the inclusion of country-level controls superfluous. The model is estimated on a sample of countries that have or have not practiced sector targeting.

The results of the analysis, presented in Table 4, indicate that the IPI quality matters for FDI inflows. The estimated coefficients should be

[11] We focus on US FDI because information on total FDI inflows is not available at the country-sector level.

[12] US direct investment abroad is defined as the ownership or control, directly or indirectly, by one US resident of 10% or more of the voting securities of an incorporated foreign business enterprise or the equivalent interest in an unincorporated foreign business enterprise. The data capture the cumulative value of parents' investments in their affiliates (source: http://www.bea.gov/bea/ai/0395iid/maintext.htm). Data points reported as values belonging to the range between −500 000 and 500 000 US dollars are treated as equal to 500 000 dollars. We interpolated missing information on stocks to increase the number of observations.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 4 Sector-level regressions. All countries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sector targeted	0.129 (0.589)	0.243 (0.578)	0.521 (0.572)	0.186 (0.594)	0.184 (0.594)	0.242 (0.579)	0.184 (0.593)	0.176 (0.594)
IPI quality × sector targeted	0.091*** (0.030)							
IH quality × sector targeted		0.087*** (0.029)			0.074* (0.043)	0.087*** (0.032)	0.074* (0.041)	
Difference between IH scores × sector targeted			0.057 (0.039)		−0.001 (0.045)	0.001 (0.044)		0.020 (0.043)
Web site quality × sector targeted				0.069** (0.028)	0.018 (0.040)		0.018 (0.040)	0.064** (0.030)
Constant	6.793*** (0.579)	6.792*** (0.579)	6.880*** (0.580)	6.821*** (0.580)	6.789*** (0.580)	6.793*** (0.579)	6.789*** (0.579)	6.824*** (0.580)
Observations	1167	1167	1167	1167	1167	1167	1167	1167
Countries	115	115	115	115	115	115	115	115
R^2	0.23	0.23	0.22	0.23	0.23	0.23	0.23	0.23

Notes. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% level, respectively. All variables are based on country–sector averages over the period 2000–2010. The dependent variable is the log of the mean of sectoral inflow of foreign direct investment from the US. Sector targeting captures the share of the years sector i was targeted by country c over the period, assuming that if a sector was targeted in 2004 it was kept targeted until 2010. IPI quality, IH quality, and Web site quality are measured as the deviation from their (global) mean, implying that the effect of the sector targeting variable is representative for the country with the average quality. Each specification includes country and sector fixed effects, the former making country-controls superfluous.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

interpreted as follows. The normalized IPI quality equals zero for the average country in the sample. Thus, in a country with the average IPI quality investment promotion does not seem to have made a difference to the inflows of FDI.[13] However, in countries with above average IPI quality, investment promotion efforts have paid off as priority sectors received more FDI than non-targeted industries. This is true when IPI quality is measured using the overall score, IH quality, or Web site quality. The results for the overall quality and the IH score also hold in the developing country sub-sample (Table 5).

3.3 Instrumental variable approach

Are countries successful at attracting FDI because of their good performing IPIs? Or do IPIs in popular FDI destinations perform better because of their frequent contacts with foreign investors? It is certainly possible that IPIs in some economies get better at their jobs as they build up experience interacting with foreign investors, though, it is also possible that high demands on IPI staff may result in lower quality of service provided to each individual investor.

To address this issue, we apply the instrumental variable approach to our aggregate regression. Our main instrument for IPI quality is the percentage of agency staff with private sector experience. Employees who have previously worked in the private sector are likely to better understand the needs of investors and their presence is likely to lead to a higher quality of IPI services. The second instrument is a dummy equal to one if the agency paid wages at or above the level offered in the private sector. Our next instrument takes advantage of various organizational forms of IPIs. Although some IPIs have the status of a ministry sub-unit, others are autonomous public bodies, semi-autonomous agencies reporting to a ministry, joint public–private entities, or fully private entities. We define a dummy equal to one if the agency is a quasi-government entity, i.e., if it is either an autonomous public body or semi-autonomous agency reporting to a ministry, and zero otherwise. Wells and Wint (2000) argue that IPIs set up as subunits of ministries or private entities are less effective than IPIs with a quasi-government status. The latter status gives an agency freedom from government staff recruiting procedures and pay scales thus allowing the agency to hire more motivated staff with better understanding of private sector needs. At the same time, the link to the government facilitates flow of feedback from foreign investors to relevant government

[13] This is based on the sum of the coefficients for Sector targeted and Sector targeted × IPI quality normalized. For the average country, the interaction term drops out (IPI quality normalized equals zero) and the coefficient on Sector targeted is not statistically significant.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 5 Sector-level regressions. Developing countries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sector targeted	0.017 (0.663)	0.194 (0.677)	−0.063 (0.660)	−0.115 (0.661)	0.114 (0.689)	0.124 (0.681)	0.159 (0.687)	−0.073 (0.661)
IPI quality × Sector targeted	0.079* (0.043)							
IH quality × sector targeted		0.091** (0.045)			0.059 (0.062)	0.062 (0.056)	0.081 (0.056)	
Difference between IH scores × sector targeted			0.128* (0.067)		0.072 (0.085)	0.074 (0.082)		0.106 (0.077)
Web site quality × sector targeted				0.047 (0.033)	0.004 (0.042)		0.012 (0.041)	0.022 (0.038)
Constant	5.502*** (0.644)	5.502*** (0.643)	5.540*** (0.643)	5.514*** (0.644)	5.514*** (0.644)	5.515*** (0.644)	5.500*** (0.644)	5.529*** (0.644)
Observations	864	864	864	864	864	864	864	864
Countries	84	84	84	84	84	84	84	84
R^2	0.21	0.21	0.21	0.21	0.22	0.22	0.21	0.21

Notes. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% level, respectively. All variables are based on country-sector averages over the period 2000–2010. The dependent variable is the log of the mean of sectoral inflow of foreign direct investment from the US. Sector targeting captures the share of the years sector i was targeted by country c over the period, assuming that if a sector was targeted in 2004 it was targeted until 2010. IPI quality, IH quality and Web site quality are measured as the deviation from their (global) mean, implying that the effect of the sector targeting variable is representative for the country with the average quality. Each specification includes country and sector fixed effects, the former making country-controls superfluous. The definition of developing countries is based on the 2011 World Bank country classification.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

agencies and aids agency's efforts to lobby on behalf of foreign investors. If the IPI changed its status during the period considered, we use the percentage of years it had the quasi-government status. The final two instruments are dummies capturing whether the agency's activities were evaluated by external entities and whether an evaluation of the agency took place at least once a year. It is likely that IPIs whose performance is rigorously evaluated are more likely to perform better. We use at the most two instruments at a time because otherwise missing values would severely reduce the size of the sample.

The results from the instrumental variable approach are consistent with a causal relationship between IPI quality and the magnitude of FDI inflows. Table 6 presents the results for the overall IPI quality based on various sets of instruments. As seen in the 1st stage, a higher percentage of agency employees with private sector experience translates into a higher quality of the agency. The higher quality in turn leads to more FDI inflows. These two relationships are statistically significant in all five specifications. *F*-statistics suggest that our instruments are reasonable predictors of the IPI quality. The Sargan test does not cast doubt on the validity of the instruments.

The results for IH quality (not reported to save space) are equally strong. Again, the 1st stage suggests that higher wages paid by the agency translate into higher quality of responses to investor inquiries. And as before we find that a higher IPI quality is positively related to FDI inflows. The results are weaker when the Web site quality is considered because the instruments do not do a good job at predicting the quality of the agency's Web site. Nevertheless, the coefficient on the Web site quality is statistically significant at the 5% level in one specification and at the 15% level in two specifications.

4 Conclusions

In response to global competition for FDI, most countries have set up IPIs as a key part of their strategy to attract foreign investors. Investment promotion is a relatively uncontroversial part of the industrial policy toolkit, and according to recent research (Harding and Javorcik 2011) it is an effective and cost-efficient way of increasing inflows of FDI, at least in developing countries.

However, not all IPIs are equally good at their core function, namely information provision. According to the GIPB initiative, IPIs vary widely in terms of the quality of information they provide on their Web sites and in response to direct inquiries from potential investors. These differences in quality of IPI services translate directly into differences in FDI inflows. Our study reaches this conclusion based on the analysis of aggregate FDI

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 6 Instrumental variable results (IPI quality)

	(1)	(2)	(3)	(4)	(5)
1st stage: dependent variable = IPI quality					
GDP per capita	6.975*** (1.214)	6.807*** (1.239)	7.167*** (1.242)	7.263*** (1.277)	7.057*** (1.335)
GDP growth	−133.997* (79.168)	−156.872* (79.758)	−127.957 (79.741)	−112.948 (81.243)	−139.186* (81.940)
Population	0.509 (0.859)	0.400 (0.857)	0.362 (0.881)	0.330 (0.884)	0.570 (0.894)
Inflation	19.754 (25.283)	21.204 (25.189)	20.485 (25.363)	16.922 (28.243)	13.731 (29.090)
Political stability	−0.948 (1.958)	−1.533 (1.988)	−1.148 (1.979)	−1.054 (1.975)	−0.732 (2.024)
% Private sector experience	0.110** (0.042)	0.110** (0.042)	0.111** (0.042)	0.110** (0.043)	0.088* (0.047)
Quasi-government status		0.982 (2.735)			
Annual evaluation			1.990 (2.555)		
External evaluation				3.611 (2.503)	
Well paid staff					3.364 (3.171)
Constant	−15.068 (15.479)	−11.480 (15.800)	−15.359 (15.522)	−17.245 (15.624)	−16.270 (16.186)

(continued)

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 6 Continued

	(1)	(2)	(3)	(4)	(5)
2nd stage: dependent variable = ln FDI flow					
IPI quality	0.086** (0.040)	0.087** (0.038)	0.076** (0.036)	0.064** (0.031)	0.091** (0.044)
GDP per capita	0.403 (0.320)	0.430 (0.300)	0.473 (0.289)	0.553** (0.249)	0.314 (0.357)
GDP growth	27.935** (10.884)	31.341*** (11.307)	26.163*** (10.012)	23.584*** (9.144)	27.407** (11.367)
Population	0.651*** (0.090)	0.661*** (0.088)	0.654*** (0.084)	0.664*** (0.079)	0.673*** (0.095)
Inflation	−2.864 (2.778)	−2.969 (2.729)	−2.642 (2.594)	−2.942 (2.544)	−4.316 (3.180)
Political stability	0.019 (0.205)	0.079 (0.206)	0.011 (0.192)	0.012 (0.178)	0.054 (0.217)
Constant	1.876 (1.635)	1.277 (1.596)	1.785 (1.530)	1.710 (1.407)	2.109 (1.762)
Observations	87	86	87	86	83
Part. R^2 Shea	0.08	0.08	0.08	0.10	0.07
F-statistics	6.76	3.60	3.66	4.12	2.94
P-value	0.01	0.03	0.02	0.02	0.05
Sargan	n.a.	0.284	0.77	1.28	0.24
P-value	n.a.	0.59	0.38	0.26	0.63

Notes. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% level, respectively. All variables are based on country averages over the period 2000–2010. The dependent variable is the log of the mean of aggregate inflow of foreign direct investment from all source countries. IPA quality, IH quality, and Web site quality are measured on a scale from 0 to 100. GDP per capita and population are included in the log form. Political stability ranges from −2.5 to 2.5, with a mean of 0 and standard deviation of 1. A higher value means more stability. The measure is available until 2009. The instruments are from the 2005 World Bank IPA survey. % Private sector experience measures the percentage of the IPI staff who have private sector experience. The other instruments are dummy variables taking the value of 1 if the IPI was a quasi-governmental agency, performed annual evaluation of its activities, external evaluations of its activities and paid their staff at or above the level of the private sector, respectively, and zero otherwise.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

flows to 156 countries during the 2000–2010 period. These conclusions are confirmed in the analysis of more disaggregated country-sector-level data. They are also robust to the instrumental variable approach.

Acknowledgements

The authors are indebted to Valeria Di Fiori and Robert Whyte, without whose help and inspiration this project would not have happened. The authors are also grateful to Leora Klapper, Kusi Hornberger, Volker Nitsch, and two anonymous reviewers for helpful comments.

References

Ahearne, A., W. Griever and F. E. Wanrock (2004), ‘‘Information Costs and Home Bias: An Analysis of U.S. Holdings of Foreign Equities’’, *Journal of International Economics* **62**, 313–36.

Blalock, G. and P. J. Gertler (2008), ‘‘Welfare Gains from Foreign Direct Investment through Technology Transfer to Local Suppliers’’, *Journal of International Economics* **74**, 402–42.

Bobonis, G. J. and H. J. Shatz (2007), ‘‘Agglomeration, Adjustment, and State Policies in the Location of Foreign Direct Investment in the United States’’, *Review of Economics and Statistics* **89**, 30–43.

Charlton, A. and N. Davis (2006), ‘‘Does Investment Promotion Work?’’, London School of Economics, Mimeo.

Daude, C. and M. Fratzscher (2008), ‘‘The Pecking Order of Cross-Border Investment’’, *Journal of International Economics* **74**, 94–119.

French, K. and J. Poterba (1991), ‘‘Investor Diversification and International Equity Markets’’, *American Economic Review Papers and Proceedings* **81**, 222–6.

Gelos, G. and S. J. Wei (2005), ‘‘Transparency and International Portfolio Holdings’’, *Journal of Finance* **60**, 2987–3020.

Gordon, R. and L. Bovenberg (1996), ‘‘Why Is Capital so Mobile Internationally? Possible Explanations and Implications for Capital Income Taxation’’, *American Economic Review* **86**, 1057–75.

Görg, H. and D. Greenaway (2004), ‘‘Much Ado about Nothing: Do Domestic Firms Really Benefit from Foreign Direct Investment’’, *World Bank Research Observer* **19**, 171–97.

Görg, H. and E. Strobl (2001), ‘‘Multinational Companies and Productivity Spillovers: A Meta-analysis’’, *Economic Journal* **111**, 723–39.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Harding, T. and B. S. Javorcik (2011), "Roll Out the Red Carpet and They Will Come: Investment Promotion and FDI Inflows", *Economic Journal* **121**, 1445–76.

Head, K., J. Ries and D. Swenson (1999), "Attracting Foreign Manufacturing: Investment Promotion and Agglomeration", *Regional Science and Urban Economics* **29**, 197–218.

Hoekman, B. and B. S. Javorcik (eds.) (2006), *Global Integration and Technology Transfer*, Palgrave Macmillan, Washington DC and New York.

Javorcik, B. S. (2004a), "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages", *American Economic Review* **94**, 605–27.

Javorcik, B. S. and M. Spatareanu (2008), "To Share or Not to Share: Does Local Participation Matter for Spillovers from FDI?", *Journal of Development Economics* **85**, 194–217.

Javorcik, B. S. and M. Spatareanu (2009), "Tough Love: Do Czech Suppliers Learn from their Relationships with Multinationals?", *Scandinavian Journal of Economics* **85**, 194–217.

Javorcik, B. S. and M. Spatareanu (2011), "Does it Matter Where you come from? Vertical Spillovers from Foreign Direct Investment and the Origin of Investors", *Journal of Development Economics* **96**, 126–138.

Portes, R. and H. Rey (2005), "The Determinants of Cross-Border Equity Flows", *Journal of International Economics* **65**, 269–96.

Portes, R., H. Rey and Y. Oh (2001), "Information and Capital Flows: the Determinants of Transactions in Financial Assets", *European Economic Review* **45**, 783–96.

Stulz, R. (1981), "On the Effects of Barriers to International Investment", *Journal of Finance* **36**, 923–34.

Tesar, L. and I. Werner (1995), "Home Bias and High Turnover", *Journal of International Money and Finance* **14**, 467–93.

Wells, L. T. and A. G. Wint (2000), "Marketing a Country", Foreign Investment Advisory Service Occasional Paper No. 13, IFC and the World Bank, Washington DC.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Appendix

Table A1 Descriptive statistics

	Mean	SD	Minimum	Maximum
FDI flow (level)	7.5 billion	22.1 billion	0.002 billion	187.9 billion
IPA quality	44.332	18.851	1.000	87.000
IH quality	28.125	18.869	1.750	79.500
Web site quality	60.551	22.060	0.000	95.000
GDP per capita	8910.700	13014.347	113.906	78322.508
GDP growth	0.042	0.023	−0.021	0.159
Population	38882160	139 million	47476	1299 million
Inflation	0.068	0.080	−0.003	0.803
Political stability no violence	−0.032	0.874	−2.215	1.467
Observations	156			
Starting business days	38.268	32.321	2.000	244.667
Construction permits days	215.622	139.831	37.667	1179.000
Registering property days	72.795	81.601	2.000	524.143
Voice and accountability	0.019	0.893	−1.616	1.613
Government effectiveness	0.041	0.912	−1.429	2.114
Regulatory quality	0.079	0.856	−1.572	1.859
Rule of law	−0.017	0.921	−1.607	1.880
Control of corruption	0.005	0.956	−1.476	2.348
Observations	153			

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

CESifo Economic Studies, Vol. 59, 2/2013, 360–391 doi:10.1093/cesifo/ifs003
Advance Access publication 7 March 2012

Learning about Financial Market Integration from Principal Components Analysis

Vadym Volosovych

Erasmus University Rotterdam, Tinbergen Institute and ERIM, 3062 PA Rotterdam, The Netherlands. e-mail: volosovych@ese.eur.nl.

Abstract

Using principal components analysis, I examine capital market integration of 15 industrialized economies from 1875 to 2009. The methodology accounts for several dimensions of integration (markets comovement and segmentation) and delivers more credible conclusions concerning the patterns of financial integration than conventional techniques (for example, simple correlations). Patterns of both nominal and real returns on long-term government bonds imply a higher level of integration by the end of the 20th century compared to earlier periods. Policy variables, common shocks, and the global market environment play a role in explaining the time variation in integration, while 'unexplained' changes in the overall level of country risk are also empirically important. (JEL codes: F02, F36, G15, N20)

Keywords: financial markets integration, nominal returns, real returns, principal components, sovereign bonds

1 Introduction

The extent of international financial integration has important implications for economic theory and policy. Specifically, the relative degree of financial integration during two capital market booms, before the First World War and after the collapse of the Bretton Woods system, remains subject to debate. Quinn (2003), Bordo and Flandreau (2003), and Bordo and Murshid (2006) argue that financial markets were more integrated during the pre-First World War era. In contrast, Bordo et al. (1999b), Bordo et al. (2001) and Mauro et al. (2002) find that markets became more integrated post-Bretton Woods. Others, including Obstfeld and Taylor (2003, 2004) and Goetzmann et al. (2005), argue that financial markets demonstrate a U-shape pattern, hence, an equal extent of integration before 1914 compared to after 1971. Typical measures of integration include bond or stock market correlations (Mauro et al. 2002; Goetzmann et al. 2005; Quinn and Voth 2008) or various parity conditions (Lothian 2002).[1]

[1] Others measure integration by the stocks of external liabilities relative to country output (Obstfeld and Taylor 2004), the degree of cross-country transmission of shocks and incidence to financial crises (Bordo et al. 2001; Bordo and Murshid 2006), or transportation costs, government barriers, and information asymmetries in commodities and financial markets (Bordo 1999b; Quinn 2003). Obstfeld and Taylor (2003, 2004) employ a combined approach with a variety of price and quantity criteria of integration.

© The Author 2012. Published by Oxford University Press
on behalf of Ifo Institute for Economic Research, Munich. All rights reserved.
For permissions, please email: journals.permissions@oup.com

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

I extend on Volosovych (2011) and further explore insights from the principal component analysis (PCA) to study financial market integration. My methodology extends the classic PCA, which was developed in 1901 by Karl Pearson (Pearson 1901), and measures several dimensions of integration (co-movement and segmentation). This method is in contrast to the conventional correlation measure, which simply measures co-movement, but fails to differentiate between global or country shocks. Using the proposed approach, I first take the perspective of investors in financial assets or *financial arbitrageurs*. Departing from the benchmark of the Uncovered Interest Parity (UIP) condition, I measure financial integration by the degree of co-movement of countries' nominal bond returns. Second, I consider the case of *real arbitrage*. For example, a domestic investor who speculates in foreign goods or simply considers investment in domestic financial assets or physical capital focuses on real, rather than nominal, returns. I follow Obstfeld and Taylor (2003, 2004) and approximate the real costs of capital as well as the expected marginal return on investments by real long-term bond returns. In theory, because of real arbitrage, the Real Interest Parity (RIP) condition holds by which the expected real interest rate differential is zero or constant. Co-movement of real interest rates is a broad measure of financial integration because, in an integrated economy, we would observe more co-movement, and, perhaps, eventual convergence of the real rates of return on both physical capital and financial assets. Further, it is possible that cross-country differences in inflation dynamics result in an entirely different pattern of integration compared to what is found when using nominal returns; this pattern could also be explained by a different set of variables. My goal is to establish the integration patterns in both markets, compare them, and explain possible differences.

I do not identify market integration with UIP or RIP conditions, thus, I do not focus on nominal or real interest rate convergence. Instead, I concentrate on a weaker notion of integration that is characterized by smaller and more stabile risk premia that results in a higher *co-movement* (but not necessarily equalization) of a country's financial returns. The advantage of the proposed empirical approach is the ability to measure the degree of integration, while tests of parity conditions assume full integration only. Loosely speaking, I employ PCA to construct a multi-variable analog of correlation and use this analog to measure co-movement of returns.[2] The PCA is a non-parametric empirical

2 As discussed in Volosovych (2011), the PCA-based measure of co-movement does not suffer from statistical and conceptual issues that plague the conventional co-movement measures such as correlation.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

methodology that is used to reduce the dimensionality of data and describe common features of a set of economic variables. Specifically, this method transforms observed data vectors into new variables, which are referred to as *components* and are linear combinations of the original data that maximize variance.[3] The goal of this method is to capture most of the observed variability in the data in a lower-dimensional object and, thereby, filter out noise. Often, a single component summarizes most of the variation of the original data. This feature of PCA is especially useful for studying market integration because when the observed economic variables have a high signal-to-noise ratio, which would be the case under economic integration, a single principal component with the largest variance can be associated with the unobserved 'world return.' In turn, this unobserved world return can capture the dynamics that is informative of the extent of market integration. In addition, I explore another aspect of integration, namely country- and group-specific effects (or *market segmentation*), which helps uncover possible reasons for the changing degree of integration. Finally, I verify how co-movement of returns has changed over time conditional on time-varying determinants of the risk premia.

My primary data include monthly series of sovereign bond yields from 1875 to 2009 that are available from the *Global Financial Database* (GFD) (Global Financial Data Inc. 2002). The sample includes 15 economies whose sovereign debt was continuously traded in a major international financial center (London) and was available in other locations as early as the mid-19th century.[4] Historically, sovereign bonds market has been the most actively traded segment of financial market. Additionally, characteristics of the underlying instruments (such as maturity, coupon payments, the identity of the issuer, etc.) are similar across countries and over time.[5] I carefully select early bond series that are most comparable with the subsequent series to minimize breaks and to ensure consistency of the long-term series. To measure nominal (monetary) returns, I use yields on long-term sovereign bonds that are payable in national currency; I do not convert the data into a single currency. The latter allows an analysis of all possible reasons for change in co-movement; for example,

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

3 See Jolliffe (2002) for a more detailed treatment of the PCA.
4 The largest sample includes Austria, Belgium, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland, the United Kingdom, and the United States. The motivation for this sample is described in Section 3.1.
5 In contrast to the stock market indices, the comparability of bond instruments makes these data attractive for long-term study of the dynamics of financial integration. Obstfeld and Taylor (2004) also stress that long-term bond yields are most appropriate for a study of the international capital mobility because they are most directly related to the financing costs for capital investments.

currency and country risk, cross-border frictions, and other limits to arbitrage. To compute the expected real returns I deflate the nominal yields by the national CPI (also from the GFD).

I estimate the index of integration from 1875 to 2009 using a relatively wide rolling window of 156 months (13 years), which leads to results that are relatively immune to short-term noise and conditional heteroskedasticity in returns. Over the very long-term, evidence points to higher financial market integration at the end of the 20th century compared to earlier periods for nominal and real returns. An analysis of market segmentation reveals that many countries frequently diverted from the group during the first half of the 20th century (at the turn of the 20th century for nominal returns or in the 1930s for real returns). Since the 1960s, 'crises' were caused by some individual divergences where, at most, two countries diverted from the group. Clearly, the divergence of domestic inflation rates is a major factor in the segmentation of markets for physical assets. These patterns are also confirmed by time-series regressions of indices of integration on linear and quadratic trends and controls that proxy for shocks, policies, and the global environment. These regressions show that policy variables (average inflation, government budget deficit, capital controls, and the exchange-rate regime) and the global market environment (approximated by average trade openness) are correlated with broad long-term integration trends in financial and physical asset markets. In particular, when inflation rates were relatively low, integration was higher; when the world was open to trade, finance generally followed (or went hand-in-hand). Interestingly, the signs of the coefficients for some variables differ depending on the use of nominal or real returns. I conjecture that the transformation of returns leads to somewhat different determinants of integration because of the differences in architecture of two markets that these transformations represent. Some forms of pegged exchange rate regime seem to be conducive to financial integration, while profligate fiscal policy hurts real arbitrage. I do not interpret this evidence literally making, for example, the case for fixed exchange rate regimes in order to promote integration. Rather, these results imply that the policies of low inflation, fiscal prudence, openness to trade, and adherence to rules-based monetary arrangements signal countries' commitments to protecting the interests of investors and to global cooperation. Because the pattern of integration resembles a J-shape, in combination with the finding that these particular policies have the strongest correlation with the integration index, indirectly shows that such policies have likely caused the upward trend in integration we have experienced since the 1930s. In addition, financial crises seem to act as a common shock in disrupting the integration of markets for physical assets. The significance of financial crises, the war, and hyperinflation stresses the importance of accounting

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

for global and large country shocks when measuring integration. Finally, the results show the existence of an unexplained variation in integration that is captured by the average level of country risk.

The present study makes several contributions. First, it extends Volosovych (2011) in demonstrating the merit of classic PCA to study historical financial integration in a dynamic context. A few existing studies limit the use of PCA to calculate the principal components over a particular exogenously-defined time period (see Nellis 1982; Gagnon and Unferth 1995; Mauro et al. 2002; Bordo and Murshid 2006). However, I demonstrate that PCA is not limited to its traditional use as a measure of co-movement. My approach allows exploration of market segmentation (country or group-specific effects) to investigate possible reasons for the changing degree of integration. The patterns seen in group-specific effects and their clustering during some time periods are not only interesting per se but also important for the motivation of regression analysis. Second, I stress important differences in the patterns of integration in markets for financial and physical assets and find empirical determinants of integration within these markets. Finally, my study relates to two distinct literatures that study financial integration using parity conditions with risk (see Alper et al. 2009 for a survey). One strand identifies the degree of integration with the risk premium and analyzes its time-series pattern and properties. The second approach consists primarily of country studies and relates the total risk premium or its components to macroeconomic fundamentals, transaction costs, or monetary policy.[6] Similar to the first strand of literature, I investigate the deviations from UIP, indirectly, by looking at co-movement of returns over time and instances of market segmentation. I also follow the second strand of literature and verify how the co-movement of returns changes over time, conditional on time-varying determinants of risk premia.

The remainder of this article is organized as follows. In the following section I lay out a conceptual framework that motivates the empirical analysis and helps interpret the results, as well as discuss the methodology used to quantify integration in various markets. Section 3 describes the data, establishes the pattern of integration, and offers some explanations of the observed patterns. Section 4 concludes.

6 Alternatively, the capital asset pricing models (CAPM) suggest that only non-diversifiable 'systematic' risk can be interpreted as a risk premium. According to the CAPM literature, if exchange and country risks cannot be (completely) diversified they are the part of the systematic risk and hence would explain return differentials.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

2 Conceptual Issues and Methodology

2.1 Theoretical motivation: deviations from the interest parity conditions

The financial literature has a long tradition of measuring financial markets integration (or market efficiency) by comparing returns on similar financial assets. The standard no-arbitrage theory predicts that free international capital flows (*financial* arbitrage) result in the UIP condition between expected rates of return of two countries, expressed in a common currency, such that $1 + i_{t,k} = (1 + i^*_{t,k})\frac{S^e_{t,k}}{S_t}$ or, as log-approximation, $i_{t,k} - i^*_{t,k} = \Delta^e_k s_t$, where $i_{t,k}$ and $i^*_{t,k}$ are the nominal interest rate on domestic asset (in domestic currency) and foreign asset (in foreign currency), S_t and $S^e_{t,k}$ are today's and expected future spot exchange rate k periods ahead, quoted as the domestic-currency price of one unit of foreign currency, and $\Delta^e_k s_t \equiv s^e_{t+k} - s_t$ (with the lower-case exchange rates being in logs). Under the UIP, a domestic investor's exchange risk exposure is uncovered.[7] By extension, in the multi-country case, perfect capital mobility would preclude local asset prices to deviate from global prices.

Subtracting expected inflation differential $\Pi^e_{t,k} - \Pi^{e*}_{t,k}$ from each side of the UIP condition and using the Fisher decomposition of the interest rate $r_{t,k} = i_{t,k} - \Pi^e_{t,k}$ we obtain the expression for the real interest differential $r_{t,k} - r^*_{t,k} = \Delta^e_k s_t - (\Pi^e_{t,k} - \Pi^{e*}_{t,k}) = \Delta^e_k q_t$, where the right-hand side represents the expected changes in real exchange rate (the deviation from the purchasing power parity). The left-hand side of this expression could be interpreted as the real domestic currency return to a domestic investor from buying foreign goods and holding them for one period. Notice that from the financing perspective, such domestic investor has to borrow funds at nominal interest rate $i_{t,k}$ and incur the real cost of borrowing $r_{t,k} = i_{t,k} - \Pi^e_{t,k}$. Therefore, domestic investors considering returns in their own country or speculating in foreign goods (the *real* arbitrage) would consider real rather than nominal returns. The RIP condition would imply that expected real return to such investment strategy would be zero in an efficient market, i.e., $\Delta^e_k q_t = 0$, which is equivalent to the

[7] The above UIP condition assumes that the number of international investors is sufficiently large, the investors are exchange-risk-neutral, transaction costs are negligible, and assets located in different countries are identical with respect to liquidity, maturity, political and default risk. UIP hypothesis can be empirically tested by estimating a regression $\Delta_k s_t = s_{t+k} - s_t = \alpha + \beta(i_{t,k} - i^*_{t,k}) + u_{t+k}$ and testing the joint hypothesis of $\alpha = 0$, $\beta = 1$, and u_{t+k} is orthogonal to the information available at t. Such test assumes the rational expectations by which $s_{t+k} = E(s_{t+k}|I_t) + u_{t+k}$ and the forecast error u_{t+k} is independent of the information at time t.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

PPP condition to hold.[8] Lothian (2002) explains that ex ante real rates converge if (i) UIP and PPP both hold perfectly, or (ii) deviations from the two conditions completely offset one another (if they are due to a common cause, such as errors in forecasts of exchange rate).

The international macroeconomics literature has established that 'frictions' associated with national borders are the main reason for having lower integration in the recent decades than what we would expect in theory.[9] Explicit government trade and capital controls, sovereign and default risk, information asymmetries, poor institutions, and high price of physical capital are examples of such cross-border frictions (see Wei 2000; Reinhart and Rogoff 2004; Caselli and Feyrer 2007; Alfaro et al. 2008). Obstfeld and Taylor (2004) conclude that the changes in quantity and price indicators of financial integration over the past 150 years have been caused by changes in barriers to international capital flows (or changes in arbitrage opportunities) over time, not by the structural changes within economies. Obstfeld and Taylor and also Eichengreen (1996) further emphasize the dramatic political, economic, and intellectual changes over the course of the century and stress the political economy considerations behind the changes in integration over time in accordance with the Macroeconomic Policy Trilemma. According to the Trilemma hypothesis, growing political tensions at home pushed national governments toward a greater macroeconomic activism in the 1920s–1930s compared with prior to the First World War, when government policies were subject to maintaining the 'rules of the game' of international Gold Standard arrangement. Consequently, the UIP or RIP with risk are natural benchmarks to study the relationship between the degree of international financial integration and economic policies, fundamentals, and international market environment, all of which determine the 'risk' for an international investor. Theoretically, the deviations from the above parity conditions could be attributed to the non-rationality of market expectations, risk aversion of investors (by which investors would demand a premium for holding assets they consider risky), existence of transaction costs, market frictions, government interventions, and limits to speculation (investors engage in arbitrage only if the excess return per unit of risk is large enough). Retaining the assumption of rational

[8] Alternatively we can argue that the real arbitrage results in the expected real interest rate differential to be constant (see Taylor and Sarno 2004).

[9] The manifestations of low international financial integration include home bias in equity holdings (French and Poterba 1991; Tesar and Werner 1995), high correlation between country savings and investment (Feldstein and Horioka 1980), lower cross-country consumption correlations than output correlations (Backus et al. 1992), lack of flows of capital from rich to poor economies (Lucas 1990).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

expectations, the deviation from the UIP can be represented as $i_{t,k} - i^*_{t,k} - \Delta^e_k s_t = \rho_t$, where ρ_t is the time-varying risk premium (broadly defined). Risk premium is positive if domestic interest rate is higher than the level predicted by the UIP. Depending on the identity of the issuer (home or foreign country), the currency of denomination, or the place where the asset is traded, or where returns are paid the risk premium may reflect the exchange risk ρ^E_t (when the assets are identical in terms of the issuer and the jurisdiction but different by the currency denomination), the default risk ρ^D_t (when the assets differ by the issuer country but both are in the foreign currency and floated in foreign market from the perspective of home investor), and the political risk of shifting the capital across borders ρ^P_t (when the assets differ by the jurisdiction but are the same in terms of currency and the issuer).

Combining the expressions for real interest differential with the formula for the deviation from the UIP we arrive at the following expression for the deviation from the RIP $r_{t,k} - r^*_{t,k} = \Delta^e_k q_t + \rho_t = (\Delta^e_k s_t - \Pi^e_{t,k} + \Pi^{e*}_{t,k}) + \rho_t$. Taylor and Sarno (2004) demonstrate that financial market efficiency, which implies that the basic (no-risk) UIP holds or even when the deviations from the UIP $i_{t,k} - i^*_{t,k} - \Delta^e_k s_t = \rho_t$ are stationary, does not preclude the long-run PPP condition from holding. This result assumes a plausible dynamic equilibrium correction relationship between prices and nominal exchange rates by which a real appreciation of a country's currency has a net long-run deflationary impact on the economy. One of the conditions for this mechanism stipulates that the expected real interest rate differential is not a constant value but a stationary process, which has empirical support (see references in Taylor and Sarno 2004). The expression for the deviation from the RIP demonstrates that in addition to the components of risk premium ρ we need to take into account the behavior of inflation rates to study co-movement of real returns.[10] Over the very long-run this analysis is complicated by the variability of development paths cross-sectionally and over time. The less-developed countries could exhibit higher inflation rates than more industrialized countries (the Balassa–Samuelson effect), and it is not obvious how such 'Balassa–Samuelson trends' evolve and how the convergence of real returns will be affected. According to Lothian (2002), the advantage of the UIP relative to the real-interest differential is that errors in expectations of inflation do not enter in.

One approach to the study of financial integration within the UIP-with-risk paradigm is to identify the degree of integration with the

[10] Frankel (1992) notes that we might consider the deviations from the PPP $\Delta^e q_t$ to be a part of the exchange risk premium ρ^E_t. I keep it separately to stress the additional factor associated with the expected inflation differential.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

risk premium and analyze its time-series pattern and properties (see Lothian 2002 for a group of industrial countries over the long term; Holtemöller 2005 for the EU accession countries). The second approach consists primarily of country studies and relates the total risk premium or its components to macroeconomic fundamentals, transaction costs, or monetary policy (see Alper et al. 2009 for the survey).[11] As Obstfeld and Taylor (2004) stress, we always face the problem that 'every test [for capital mobility] is usually a matter of degree', and the choice of benchmark to which today's integration should be compared is difficult. Furthermore, the data availability often poses a challenge to the direct studies of the UIP or RIP over the long term. Considering this, I do not test for the interest parity conditions directly but instead combine these distinct approaches.

Similarly to the first literature, I study the deviations from UIP/RIP indirectly by looking at *co-movement* of returns over time. To illustrate the intuition, let us focus on 2-country case with nominal returns and assume that the UK's nominal interest rate i^{UK} can be represented as $i^{w} + \rho^{UK}$, where i^{w} is the unobserved 'world' interest rate and ρ^{UK} is the country-specific risk premium (time subscripts are omitted for brevity). According to the UIP with risk, France's interest rate can be represented as $i^{FR} = i^{w} + \rho^{UK} + \Delta^{e}\ s^{FR/UK} + \rho^{FR}$. For the analysis of co-movement of the two returns what matters is the second movements of the items on the right-hand side of the last equation because the covariance $\mathrm{Cov}(i^{UK},\ i^{FR}) = \mathrm{Cov}(i^{w} + \rho^{UK},\ i^{w} + \rho^{UK} + \Delta^{e}\ s^{FR/UK} + \rho^{FR})$ consists of the variance of the world rate, variances and all the covariances of risk factors (ρ^{UK} and ρ^{FR}) and the expected exchange rate changes ($\Delta^{e}\ s^{FR/UK}$) with each other. If the majority of countries are integrated into the world financial markets, their interest rates move together with the world rate i^{w}, and thus the covariance of each pair of countries is dominated by the variance of world return $Var(i^{w})$.

I also follow the second strand of literature and verify how the co-movement of returns changed over time, *conditional* on time-varying determinants of risk premia. As in Volosovych (2011), I control for trade openness, the measures of domestic economic policy and macroeconomic fundamentals (average inflation, average government deficit, capital

[11] Alternatively, the capital asset pricing models (CAPM) suggest that only non--diversifiable 'systematic' risk can be interpreted as a risk premium. According to the CAPM literature, if exchange and country risks cannot be (completely) diversified they are the part of the systematic risk and hence would explain return differentials.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

controls, and exchange rate regime), and the proxies for economic shocks such as financial and economic crises or hyperinflation episodes.[12]

2.2 Empirical methodology

The conventional approach to measure co-movement of economic series is by correlation coefficient. To summarize co-movement in a group of markets, the usual practice is to compute the average of the correlation coefficients for each country-pairs (see Mauro et al. 2002 or Quinn and Voth 2008, for example). As argued by Volosovych (2011) interpreting a high correlation of economic series as evidence of substantial integration has several issues. In particular, the choice of the reference country when calculating return spreads is often ad hoc and may result in conflicting time patterns of co-movement; the sample correlation is not a robust statistics in the presence of outliers or heavy-tailed distributions; very often, the observed differences in correlations result from the changes in statistical properties of a sample rather than from actual economic links. The latter complication is related to the issue of conditional heteroskedasticity of market returns, or the hypothesis that cross-market correlations depend on market volatility (see details and references in Volosovych 2011).[13] Finally, one cannot distinguish high integration and a common shock as both may show up as a higher value of correlation.

Therefore, I follow Volosovych (2011) and apply the principal component analysis to capture the co-movement aspect of integration over time for nominal and real bond returns. This method is valid without needing specific assumptions regarding the particular distributions of the data except that it does require the data is interval-level. The goal is to transform the observed data vectors into unobserved orthogonal linear combinations, referred to as *components*, that maximize the variance of the components. The components are then ranked by their variance, with the first component having the largest variance. The lower-order components typically yield a larger variance compared to the original series. Additionally, each component of a higher order 'explains' most of the *residual* variation in the data that is not captured by the previous component, and so on. As a result, a smaller number of components, often just the first component, summarizes most of the observed variation in the data and filters out noise. Because the PCA has the data-reduction as a goal, the first component is essentially a muli-variate analog of correlation.

[12] See Volosovych (2011) for the theoretical motivation for these variables and Appendix A for details on their calculation.

[13] This issue is an inherent flaw of correlations and the literature seemingly has not reached a consensus on how to correct the problem of conditional heteroskedasticity in general.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

In empirical implementation, I follow the practical issues advocated in Volosovych (2011). First, I estimate principal components with rolling windows and interpret the proportion of the total variation in the level of individual returns explained by the first principal component as a dynamic measure of integration in a group of countries (the *index of integration*). The issue of conditional heteroskedasticity is not known to plague the results of PCA, which is a more outlier-resistent and distributionally robust method. Still, I choose a relatively wide window to guard against the short-term noise and possible issues of conditional heteroskedasticity in returns. Second, I apply PCA to the *levels* of returns. In this case, assuming some integration across markets, the first principal component may naturally be interpreted as the unobserved 'world' return (the variable i^w in Section 2.1) since this factor captures the most variation in individual returns. Spreads versus this estimate of world return are independent of a particular reference country, which is an advantage. Third, the index of integration can be plotted over time to provide a visual illustration of the time path of integration and be used to explain the discovered patterns in a regression framework.

3 Evidence from the Index of Integration

3.1 Historic sovereign bond data

I use monthly bond data over 1875–2009 from the *Global Financial Database* (GFD). This database contains monthly financial data from about 100 countries with bond and equity series for some countries beginning as early as the 18th century. The GFD reports bond yields for comparability. I conduct extensive data cleaning to achieve data consistency, referring to the descriptions of particular historical bond issues in the GFD manual, *Kimber's Record of Government Debts and other Securities* (Bartholomew and Kimber), published in 1920 and 1922, Bordo and Flandreau (2003), and other sources (Volosovych 2011 provides details on bond data). I use a fairly homogenous but representative sample of 15 relatively advanced economies whose sovereign debt was continuously traded in the major international financial center (London) as early as the mid-19th century. It is reasonable to assume that all these countries had similar structural or institutional conditions, at least in relation to the development of financial markets, if not the level of overall economic development. With a few exceptions, there were no major defaults on government debt by these countries that would create discontinuities in the time series. Countries include Austria, Belgium,

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

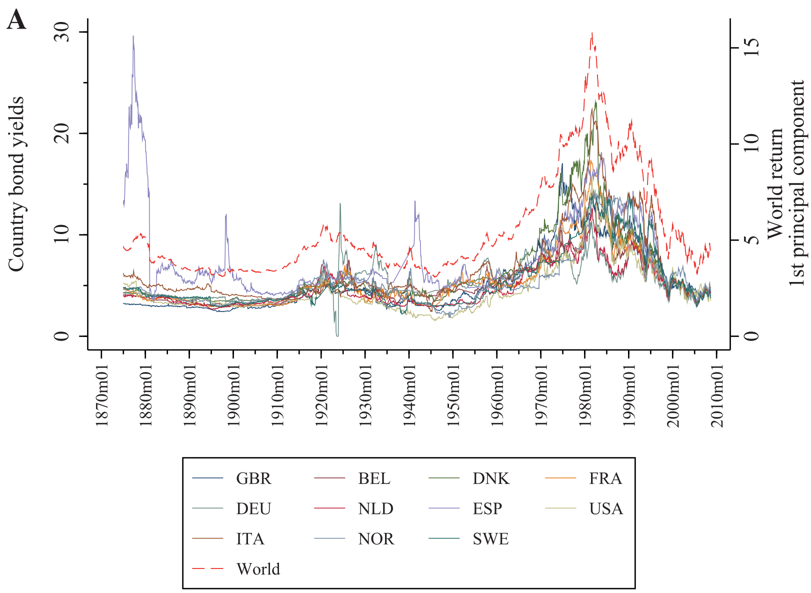

Figure 1. Returns on sovereign long-term bonds and the estimated "world" return, 1875–2009. *Notes:* The graph depicts historical monthly series for the returns on long-term government bonds issued by industrialized economies (thin solid lines, left axis) and the estimated 'world' return (thick dashed line, right axis). (A) reports series on nominal returns (bond yields); (B) reports series on real returns computed using the 'exact' Fisher formula $r_t = (1 + i_t)/(1 + \Pi^e_t) - 1$, where i_t is nominal bond yield (A) and Π^e_t is the expected inflation rate. For this graph, the individual real returns data is winsorized at top and bottom 1% of the distribution of the pooled real returns sample. The following abbreviations for the country names are used at the graph: AUT for Austria, BEL for Belgium, DNK for Denmark, FIN for Finland, FRA for France, DEU for Germany, ITA for Italy, JPN for Japan, NLD for The Netherlands, NOR for Norway, ESP for Spain, SWE for Sweden, SWI for Switzerland, GBR for UK, and USA for the USA. World is the estimate of the first principal component using all countries (the 'world' return). The estimation of the component is performed as the centered rolling window with the bandwidth of 156 months.

Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland, UK, and the USA.[14] For the longest time period, I use a sub-sample of 11 countries with appropriate yield

[14] Obstfeld and Taylor (2004) study similar set of countries and refute the idea that these countries faced different 'shocks to technology' over the century. Bordo and Schwartz (1996) classify Austria, Denmark, Italy, Japan, Spain and the USA to be emerging countries in pre-First World War period.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

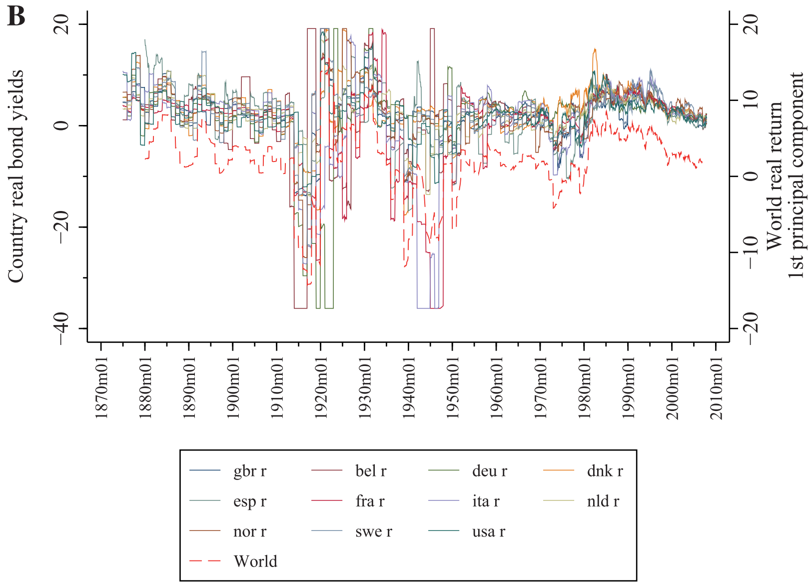

Figure 1. Continued.

data available for the entire time period.[15] I exclude British Empire countries since colonial ties with the UK influenced their spread behavior (i.e., over the 1880–1930, the Empire countries typically had interest rates within two percentage points of British rates; see Obstfeld and Taylor 2004).

For the analysis of integration based on *nominal returns*, I use the series of bonds payable in national currency, even when some issues were floated in London, because my goal is to analyze all possible reasons for changes in co-movement including exchange, political, default risk, cross-border frictions, and other limits to the arbitrage discussed in Section 2.1. Figure 1A illustrates the individual country returns used in this study with the values on the left axis and the estimated 'world' return (the first principal component) on the right axis. Upper panel of Table 1 presents the summary statistics for bond returns across exogenous periods according to prevailing international monetary arrangements as

[15] I exclude Austria, Finland, Japan, and Switzerland whose consistent bond yield series start in the GFD after the First World War. For example, for Switzerland the only data available before 1925 is the average of 12 state and federal railway bonds. In other cases, I only have the data for the sovereign bonds payable in gold until the late 1920s, which does not allow to investigate the currency risk.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

defined by Bordo and Schwartz (1999).[16] The first two columns report the average return and corresponding standard deviation for each country over the entire time period. The remainder of the table reports the average returns across five historical periods. For all the countries, bond returns remained low and stable throughout the first half of the 20th century. Following the Second World War, the returns rose continuously reaching double digits in the 1970s–80s, then turned down sharply in the 1990s. The 1971–1990 period was also exceptional as evidenced by the highest variability or returns. Volatility decreased somewhat in the 1990s; however, it did not reach the low pre-Second World War level. A similar pattern is observed in the behavior of cross-section variability of returns. Specifically, both unadjusted and mean-adjusted cross-section standard deviations imply a hump-shaped pattern of variability with the top during the Bretton Woods and Modern Float and the tendency for convergence during the last period. Interestingly, the entire Bretton Woods period was not exceptional (compared to, for example, the Gold Standard) in terms of low volatility of interest rates despite the common belief that that quiescent period did not see many shocks. The general ranking of countries in terms of levels of returns is approximately preserved over the sample period despite some important changes in individual yields over time.

I compute *real returns* using the 'exact' Fisher formula $r_t = (1 + i_t)/(1 + \Pi_t^e) - 1$, where Π_t^e is the expected inflation rate. As Obstfeld and Taylor (2004), I assume perfect foresight and calculate Π_t^e as 12-month-forward inflation rate from the realized consumer price indices as $P_{t+12}/P_t - 1$. For countries that have only annual CPI series before the 1930s I use annual inflation rate in the following year.[17] Lower panel of Table 1 demonstrates that real returns were much more volatile over time (judging from time standard deviations for individual countries or cross-section standard deviations and Figure 1B), especially in inter-War

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

[16] With slight modifications these periods include (i) Classical Gold Standard, 1875:01–1914:07; (ii) Interwar Period, 1919:01–1939:08; (ii) Bretton Woods System, 1945:06–1971:07; (iv) Modern Float, 1971:08–1990:12; and (v) Modern Globalization, 1991:01–2008:09. I also combine the latter two periods into a single post-Bretton Woods period. I omit the years of two world wars because then markets were usually inactive and reliability of data is questionable. I also define the sub-periods to obtain the maximum coverage across countries. Sometimes I have to fill the missing values with linear interpolation but preferred not to do interpolation where the data were missing in the beginning of the sub-sample for a particular country. In such case the series for this country is started from first available observation.

[17] I also tried estimating Π_t^e from interpolated monthly CPI series for such cases with similar results.

Table 1. Historic bond data. Descriptive statistics

	Full period 1875:01–2008:09 Mean (s.d.)	Gold Standard 1875:01–1914:07 Mean (s.d.)	Inter-War 1919:01–1939:08 Mean (s.d.)	Bretton Woods 1945:06–1971:07 Mean (s.d.)	Modern Float 1971:08–1990:12 Mean (s.d.)	Globalization 1991:01–2008:09 Mean (s.d.)	Post-Bretton Woods 1971:08–2008:09 Mean (s.d.)
	Nominal returns						
Austria	6.0 (1.7)	4.6 (1.0)	6.6 (0.9)	6.1 (1.2)	8.4 (1.1)	5.5 (1.5)	7.0 (1.9)
Belgium	5.3 (2.4)	3.3 (0.3)	4.6 (0.8)	5.4 (1.1)	9.6 (2.0)	5.7 (1.7)	7.8 (2.7)
Denmark	6.3 (3.8)	3.8 (0.3)	4.9 (0.6)	6.1 (1.9)	13.9 (3.6)	5.9 (1.8)	10.0 (4.9)
Finland[a]	7.9 (2.8)	4.5 (0.1)	6.9 (1.5)	8.0 (1.8)	10.9 (1.8)	6.1 (3.3)	8.6 (3.5)
France	5.5 (2.8)	3.4 (0.4)	4.7 (0.9)	5.3 (0.9)	11.0 (2.5)	5.5 (1.7)	8.4 (3.5)
Germany	5.4 (1.8)	3.8 (0.3)	5.9 (1.8)	6.0 (1.1)	7.7 (1.5)	5.2 (1.5)	6.5 (2.0)
Italy	6.6 (3.7)	4.5 (0.7)	4.8 (0.4)	5.9 (1.3)	13.4 (3.8)	7.2 (3.5)	10.4 (4.8)
Japan	5.7 (2.2)	5.7 (0.9)	5.1 (0.7)	7.5 (2.3)	7.1 (1.6)	2.2 (1.6)	4.7 (2.9)
The Netherlands	4.9 (2.0)	3.4 (0.4)	4.4 (1.2)	4.5 (1.5)	8.3 (1.5)	5.4 (1.5)	6.9 (2.1)
Norway	5.4 (2.7)	3.8 (0.5)	5.3 (0.8)	4.0 (1.2)	10.3 (2.8)	6.2 (1.8)	8.3 (3.1)
Spain	7.7 (4.3)	7.7 (5.6)	5.8 (0.5)	5.9 (1.2)	12.9 (2.4)	6.7 (3.1)	9.9 (4.2)
Sweden	5.4 (2.8)	3.7 (0.5)	4.1 (0.8)	4.6 (1.4)	10.7 (2.0)	6.4 (2.5)	8.6 (3.1)
Switzerland[a]	4.1 (1.2)	– (–)	4.6 (0.9)	3.5 (0.8)	4.9 (1.0)	3.6 (1.3)	4.3 (1.3)
UK	5.4 (3.1)	2.9 (0.2)	4.1 (0.7)	5.2 (1.8)	11.3 (2.0)	6.2 (1.9)	8.9 (3.2)
USA	4.7 (2.4)	3.4 (0.6)	3.6 (0.8)	3.7 (1.5)	9.2 (2.3)	5.5 (1.2)	7.5 (2.6)
All countries	5.8 (3.0)	4.2 (2.0)	5.0 (1.3)	5.4 (1.9)	10.0 (3.3)	5.5 (2.4)	7.9 (3.6)
Number of months	1483	475	248	314	233	213	446
Average σ_{CS}	2.14 (2.06)[b]	2.03	1.30	1.91	3.27	2.43	2.87
Average CV_{CS}	0.39 (0.37)[b]	0.49	0.26	0.35	0.33	0.44	0.38

(continued)

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 1. Continued

	Full period 1875:01–2008:09 Mean (s.d.)	Gold Standard 1875:01–1914:07 Mean (s.d.)	Inter-War 1919:01–1939:08 Mean (s.d.)	Bretton Woods 1945:06–1971:07 Mean (s.d.)	Modern Float 1971:08–1990:12 Mean (s.d.)	Globalization 1991:01–2008:09 Mean (s.d.)	Post-Bretton Woods 1971:08–2008:09 Mean (s.d.)
	Real returns						
Austria	−0.5 (16.0)	3.8 (3.6)	−12.2 (32.9)	−3.1 (12.7)	3.4 (2.1)	3.3 (1.3)	3.4 (1.8)
Belgium	2.7 (6.2)	2.3 (7.3)	1.3 (9.4)	3.1 (4.2)	3.6 (3.5)	3.6 (1.9)	3.6 (2.9)
Denmark	3.6 (4.6)	3.7 (3.5)	4.0 (8.3)	1.7 (2.7)	5.7 (3.6)	3.7 (2.0)	4.8 (3.1)
Finland[a]	2.4 (7.2)	−1.5 (7.2)	5.0 (6.6)	−0.0 (9.3)	2.0 (5.0)	4.5 (3.4)	3.1 (4.5)
France	1.1 (8.1)	2.9 (2.3)	−0.7 (12.5)	−3.3 (11.6)	2.9 (3.4)	3.8 (1.7)	3.3 (2.8)
Germany	4.0 (34.1)	2.7 (3.7)	8.8 (82.9)	3.0 (3.3)	4.0 (1.8)	3.0 (1.1)	3.6 (1.6)
Italy	2.3 (6.7)	4.3 (2.6)	0.5 (10.2)	0.6 (8.8)	1.2 (5.4)	4.1 (2.7)	2.6 (4.6)
Japan	1.7 (11.5)	3.4 (5.4)	5.5 (8.8)	−3.9 (20.9)	1.8 (4.4)	1.9 (1.3)	1.8 (3.3)
The Netherlands	3.2 (4.0)	3.5 (3.4)	5.9 (6.0)	0.2 (2.6)	3.6 (2.8)	3.0 (1.4)	3.4 (2.2)
Norway	3.2 (5.3)	3.3 (3.8)	7.0 (8.9)	−0.0 (3.3)	2.3 (4.0)	4.1 (1.9)	3.1 (3.3)
Spain	2.6 (5.0)	5.3 (3.6)	3.2 (6.5)	−0.4 (3.9)	0.7 (5.4)	3.1 (2.6)	1.8 (4.5)
Sweden	3.1 (4.9)	3.4 (4.2)	6.0 (7.6)	0.3 (3.1)	2.0 (3.2)	4.7 (2.5)	3.3 (3.2)
Switzerland[a]	2.7 (4.5)	– (–)	6.8 (7.1)	1.1 (1.8)	0.8 (2.3)	2.3 (1.0)	1.5 (1.9)
UK	2.9 (4.6)	3.4 (3.3)	5.3 (8.0)	1.2 (2.4)	1.2 (4.0)	3.3 (2.1)	2.2 (3.4)
USA	2.7 (4.1)	3.0 (3.5)	4.8 (6.1)	0.6 (3.4)	2.7 (4.0)	2.8 (1.4)	2.8 (3.1)
All countries	2.5 (11.5)	3.4 (4.2)	3.5 (24.7)	0.1 (8.4)	2.5 (4.0)	3.4 (2.1)	2.9 (3.3)
Number of months	1474	475	248	314	233	204	437
Average σ_{CS}	92.6 (96.6)[b]	96.1	92.6	90.3	89.7	91.7	90.6
Average CV_{CS}	4.88 (4.71)[b]	4.49	4.84	5.64	4.93	4.64	4.79

Notes: For individual countries the mean and standard deviation represent the time statistics for the sub-period. World War periods (1914:08-1918m11 and 1939:09-1945:08) are excluded. [a]Data for Finland starts from 1910:09; data for Switzerland starts from 1915:02. Average σ_{CS} is the time average of the cross-sectional standard deviation across in-sample countries. Average CV_{CS} is the time average of the cross-sectional standard deviation divided by the cross-sectional mean across in-sample countries. [b] without Finland, and Switzerland with missing data in Gold Standard period. See Section 3.1 for details on return series.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

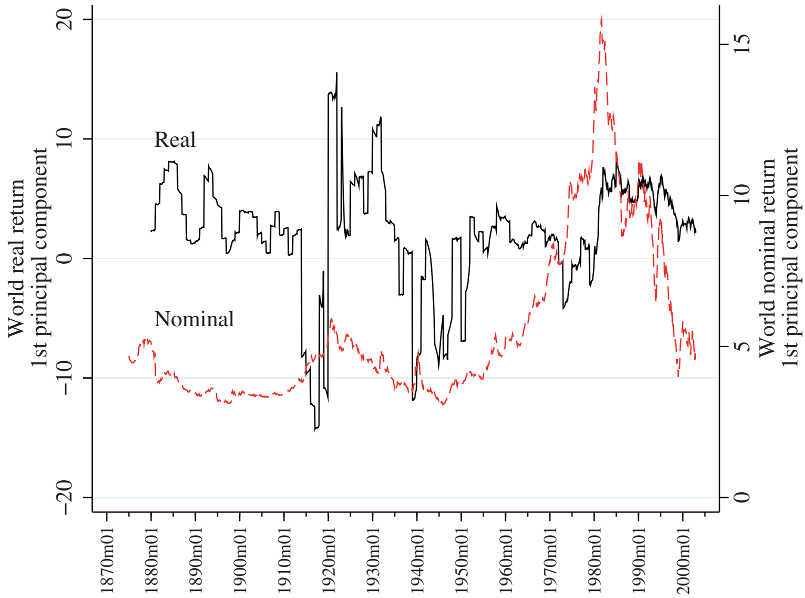

Figure 2 The estimated "world" return based on nominal and real bond returns, 1875–2009. *Notes:* The graph depicts the estimated 'world' return based on the nominal (right axis) and real (left axis) returns on long-term government bonds. The world return is equal to the first principal component using all countries. The estimation of the component is performed as the centered rolling window with the bandwidth of 156 months. The real returns are computed using the 'exact' Fisher formula $r_t = (1 + i_t)/(1 + \Pi_t^e) - 1$, where i_t is nominal bond yield (as in Panel A) and Π_t^e is the expected inflation rate. See Section 3.1 for details on return series.

and early Bretton Woods periods.[18] Among the general trends, one can see relatively 'calm' periods of Gold Standard, the 1960s and from the late 1980s and a period of negative real returns in the high-inflation 1970s. As noted earlier, the PCA method is less prone to biases due to extreme outliers, and hence should be especially useful for the analysis of integration based on real returns. Figure 2 plots the estimated 'world' return based on the nominal (right axis) and real (left axis) returns for better visual comparison of these patterns. As seen, the first principal component captures the features of the data quite well.

---

[18] Sometimes I observe extreme values of estimated returns. Hence I winsorize the data for the empirical analysis on top 0.05% in the distribution of real returns. For Figure 1B, the individual real returns data is winsorized at top and bottom 1% of the distribution of the pooled real returns sample.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

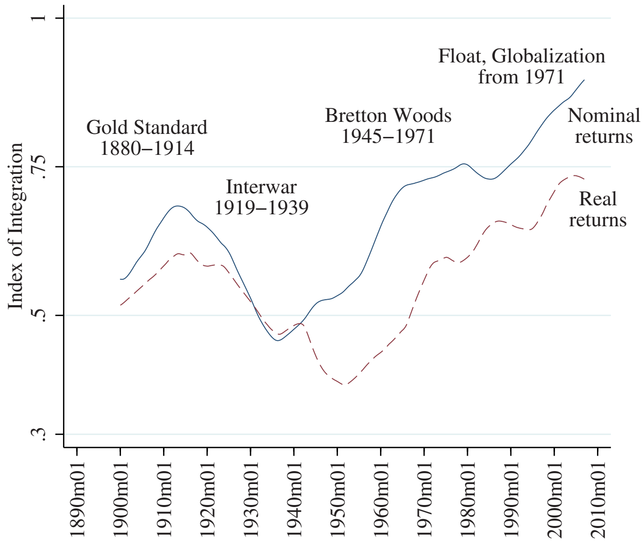

Figure 3 Index of integration in bond markets, 1875–2008. Proportion of variation in bond returns explained by the 1st principal component (smoothed series). *Notes:* The figure reports the estimates of the proportion of variation in bond returns explained by the first principal component smoothed using the uniformly weighted moving average smoother. Government bond returns are in levels. The real returns are computed using the 'exact' Fisher formula $r_t = (1 + i_t)/(1 + \Pi_t^e) - 1$, where i_t is nominal bond yield and Π_t^e is the expected inflation rate. See Section 3.1 for details on return series. The estimation of the component is performed by rolling window with the bandwidth of 156 months. In-sample countries are Belgium, Denmark, France, Germany, Italy, The Netherlands, Norway, Spain, Sweden, UK, and USA.

3.2 Trends in bond markets integration

I estimate principal components for 11 countries with longest data using 156 month rolling windows, separately for nominal and real bond return series.[19] The graph of the total variation in returns explained by the first principal component, the index of integration, is presented in Figure 3.

[19] Estimations with the bandwidth of 120 and 180 months produced very similar pattern. Widening the window makes the trend line representing the 'integration index' less jiggered. In addition, I used the historic bond data from 1880 to 1914 kindly shared by Marc Weidenmier. The resulting pattern of index of integration is very similar. Finally, I experimented with altering my country sample by (i) estimating the indices of integration using larger (15-country) sample including Austria, Finland, Japan, and Switzerland but over shorter time period; and (ii) dropping the USA (the remaining 'emerging market' in the early 20th century) from the baseline 11-country sample with just the European countries remaining. These experiments did not change the pattern much.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

The monthly index of integration is quite volatile hence I present the smoothed trend lines for nominal returns (solid line) and real returns (dashed line). Based on this figure, we can draw the following conclusions. There are many similarities observed for nominal and real returns. First, consistent with the literature, the dynamics of integration did not follow a simple linear trend over 130 years. Integration grew from the late-19th century up to 1914, when the First World War broke out. Following this, the trend in integration turned negative and reached a historic low around the time of the Great Depression during the 1930s. There was a partial recovery of international financial linkages in the 1920s; however, it was very short-lived. Second, despite the common view, it is clear that in the present group of countries integration in sovereign bond market reached the levels comparable to the Gold Standard era as early as the late 1960s, on the verge of the collapse of the Bretton Woods system. Third, the evidence points to a higher financial market integration at the end of the 20th century compared to the earlier periods. The integration thus followed a *J*-shaped trend with a trough as early as the 1920s, rather than a U-shape, as documented by Obstfeld and Taylor (2004).

Comparing the pattern of co-movement of the nominal and inflation-adjusted returns, one can see some interesting differences. Specifically, most of the time the first principal component explains a lower proportion of variation in real returns than in nominal returns (the index based on real returns has a lower value than the index from nominal yields). This implies that inflation dynamics creates an additional reason for divergences of real returns. It should be noted, however, that the key in describing the pattern of integration is to focus on the long-run *trend* and not the level of an index per se. The main difference is in the timing of the all-time minimum of two trend lines. The trend in integration based on nominal returns has a low in the 1930s, turns positive after the Second World War and continues its upward crawl almost non-interrupted to the present time. The historic minimum in the index computed from real returns is observed in the late 1940s–early 1950s, following the Second World War. The possible reason for this pattern is a fairly large difference in inflation performance of various countries after the war[20] I do not discuss this issue further because the primarily focus of this article is the financial integration patterns based on nominal versus real returns and not developments of inflation.

[20] From the mid-1950s inflation rates moved more in-sync, and the trend in the index based on real returns turns positive and largely follows the index based on nominal returns. The graph of the country inflation rates in post-World War II period is excluded for brevity and is available upon request.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

3.3 Toward explaining integration patterns

First step in explaining the time variation in the co-movement-based index of integration involves the analysis of country or group-specific effects. Time-varying country weights (called 'loadings' in PCA) on the first principal component help identify the periods of sudden drops in individual country co-movement with the world, or group divergencies. Countries with larger loadings contribute the most to the unobserved world return approximated by the first principal component; low or negative loadings reveal those countries whose bond returns move independently. My methodology delivers two *indeices of segmentation* that supplement analysis of markets co-movement and give a more complete picture about the dynamics of integration. Figure 4 presents two indices of segmentation computed from nominal returns (Figure 4A) and real returns (Figure 4B). The line is the standard deviation of the individual country loadings associated with the first principal component and the bars represent the number of countries out of 11 (the sample size) with negative loadings. The indices of segmentation in both Figure 4A and B show that the 'crises' picked up by the index of integration in various time periods are in fact brought about by very different causes.[21] There was little integration in the entire sample in the first half of the 20th century: all countries frequently diverted from the group and their weights in the world return varied. In contrast, the late and post-Bretton Woods era does not observe this kind of divergences—then, at most two countries diverted from the group. There is some instability at the end of the 1970s, otherwise, integration of bond markets of in-sample countries is remarkable. What is interesting, for the nominal returns the most divergencies are observed at the turn of the 20th century, while the real returns show most 'dis-integration' in the 1930s. This pattern makes sense: before inflation expectations were anchored by the gold standard (fully established in all the countries in sample by the 1910s) the variation in inflation trends across countries contributed to divergencies in nominal returns; while the mid-1920s (for Germany and Austria) and the time of Great Depression showed the divergence across countries in terms of national economic policy priorities, including control of inflation. In addition, the real asset markets show much more segmentation than the markets for financial assets in the

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

---

[21] Appendix D in a working paper version of Volosovych (2011) [Volosovych V. 2011. Measuring Financial Market Integration over the Long Run: Is there a U-Shape? Tinbergen Institute Discussion Paper Series, Discussion Paper no. 11-018/2] names a number of global and country events that could have caused the time pattern of integration discovered by indices of integration and segmentation seen in Figures 3 and 4. Overall, the methodology advocated in these papers matches country events and global crises in financial integration remarkably well.

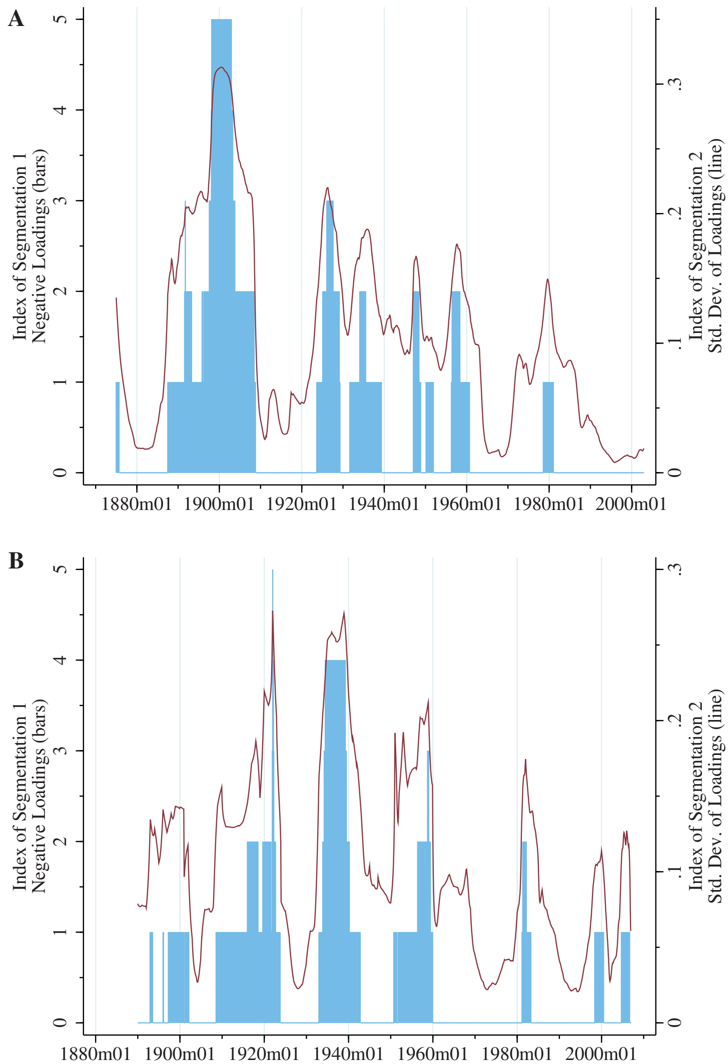

Figure 4 Segmentation in bond markets, 1875–2002. Variability of the loadings of the bond returns on the 1st principal component (Indices of segmentation). *Notes:* In each panel, the line is the standard deviation of the individual countries' component loadings associated with the 1st principal component (right scale). Bars represent number of countries out of 11 with negative loadings (left scale). The extraction of the 1st principal component is performed for the sample of 11 countries using centered moving window time sub-sample with 156 months bandwidth. (A) the nominal returns are used; (B) relies on real (inflation-adjusted) returns. In-sample countries are Belgium, Denmark, France, Germany, Italy, The Netherlands, Norway, Spain, Sweden, UK, and USA.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

decade following the Second World War given the variation in inflation experience.

Next, I run conditional regressions of the index of integration on linear and quadratic time trends and proxies for market frictions, policies, and institutional arrangements over the period 1975–2002 (see Appendix A for details on variables calculation).[22] I assume the error structure is heteroskedastic and autocorrelated up to 12 months, and also control for the periods of two world wars. The regression analysis helps verify whether the graphical evidence for the J-shaped pattern holds true in a statistical sense and study which factors were associated with the observed pattern of integration more formally. The first explanatory variable is a proxy for pro-globalization market environment, measured by the *average trade openness*. The measures of domestic economic policy and macroeconomic fundamentals characterize the extent of debt burden or laxity of lending standards in a given country. I use the cross-sectional *average inflation rate* as a proxy of overall laxity of government policy, such as the degree of commitment to a fixed exchange rate regime or inclination to finance excessive government expenditures. To control for fiscal policy, I compute the cross-sectional *average government deficit to GDP*. The variable *Prevalence of capital controls* is the fraction of countries with restrictions on capital flows in a given time period. The variable *Prevalence of pegged exchange rate regimes* is constructed similarly using monthly exchange rate regime dummies. I also include proxies for economic shocks such as financial and economic crises or hyperinflation episodes. Finally, I include the *average country risk*, computed as the cross-sectional average of individual bond spreads versus the estimated 'world' return. The definitions of the variables imply the following interpretation of results. If a particular explanatory variable is positive significant then this factor may be consistent with market integration or reflect a common shock affecting all or most of the countries, which may be interpreted as evidence of market integration too (see Bordo et al. 2001). Negative significance would be consistent with declines in integration. For country risk, the sign is ambiguous: if the variable captures all unaccounted country characteristics that may discourage foreign investment the sign would be negative; if it controls for unaccounted common shocks the effect is positive. I must stress that it is not clear a priori whether the above theoretical relationships work similarly for the measure of integration based on nominal returns (financial arbitrage) or real returns (real arbitrage).

[22] Since the index of integration is estimated by a centered moving window of 156 months (13 years), the ends of the periods would use less observations. In the regressions I prefer using the index estimated with all the data. The results over longer period of time, up to 2008:09 are very similar.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 2 shows the regression for the index of integration from nominal returns, thereby replicating results from Volosovych (2011). As discussed in Section 2.1 this measure is likely to reflect financial arbitrage and the perspective of an investor who considers monetary returns. The regression in column (1) tests for the overall unconditional trend in integration. The coefficients of linear and quadratic trends are highly significant and point to a non-linear trend line. The values of the coefficients imply that the trough in the trend in integration is around September 1928, which is remarkably consistent with the historical evidence pointing at the Great Depression as a 'watershed' of financial integration (Obstfeld and Taylor 2003, 2004). The fact that the minimum in integration corresponds to the late 1920s, closer to the beginning of the sample period used for this estimation, is evidence that the trend line follows a J-shape over the period 1875–2009.[23] Column (2) combines the explanatory variables in one multiple regression to determine which of them are 'preferred' by the data. Trade openness was, on average, complementary to financial openness as the literature shows. Out of policy variables, the results indicate that high-inflation and high-government deficit policies generally were associated with a divergence of returns or lower integration; while the coefficient of budget deficit is not significant. Pegged exchange rate regime is positive significant in support of the argument that these policies mostly served as a credible commitment device in this group of countries. Capital controls did not seem to tame financial returns from moving together. The remainder of the variables proxy for various economic shocks. The majority of crises happened during the turbulent 1920s and post-Second World War, which also saw the remarkable hyperinflation spells in central European countries. Consistent with the theory, the hyperinflation years are negative and very significant. Apparently, hyperinflations were detrimental to the overall integration but were not global shocks that would force most of the yields to move together and result in a positive coefficient. Financial or macroeconomic crises seem to work their way through the overall country risk. The Average Country Risk, which can be interpreted as a broad proxy for the 'unexplained' factors priced into the country yields, is significant and negatively correlated with the index. I interpret this result as evidence of the overall backlash against integration when the overall (actual or perceived) level of risk rose. With these controls, the linear time trend lost its significance while the quadratic trend remains positive and significant. One might argue that 6-lags (half-year) order assumed for the error structure might be too short to account for

[23] Volosovych (2011) reports that the results using quarterly or yearly averages of the monthly data are very similar, with the minima around the time of the Great Depression.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 2. Determinants of bond markets integration. Nominal returns, 1875–2002

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: index of integration based on nominal returns				
Type of data	Monthly			Quarterly	Yearly
Time trend	–0.607***	–0.237	–0.237	–0.702	–2.69
	(0.083)	(0.152)	(0.204)	(0.659)	(3.41)
Time trend^2	0.475***	0.215**	0.215	1.919	29.84
	(0.050)	(0.104)	(0.140)	(1.353)	(27.84)
Average trade openness		0.326***	0.326**	0.318*	0.278
		(0.122)	(0.163)	(0.176)	(0.227)
Average annual inflation rate		–0.020***	–0.020**	–0.021**	–0.025*
		(0.007)	(0.009)	(0.009)	(0.013)
Average government deficit		–0.024	–0.024	–0.084	–0.365
		(0.346)	(0.434)	(0.478)	(0.702)
Prevalence of pegged exchange rate regimes		0.147***	0.147***	0.146***	0.136**
		(0.031)	(0.040)	(0.044)	(0.058)
Prevalence of capital controls		0.037	0.037	0.035	0.025
		(0.034)	(0.045)	(0.049)	(0.065)
Hyperinflation years		–0.064***	–0.064**	–0.063**	–0.058
		(0.024)	(0.028)	(0.030)	(0.042)
Prevalence of financial crises		0.046	0.046	0.047	0.051
		(0.044)	(0.054)	(0.056)	(0.068)
Prevalence of consumption disasters		0.023	0.023	0.014	–0.026
		(0.064)	(0.083)	(0.091)	(0.122)
Average country risk		–0.234***	–0.234***	–0.239***	–0.270**
		(0.061)	(0.081)	(0.088)	(0.119)
World Wars	0.012	0.096**	0.096	0.110†	0.174
	(0.043)	(0.049)	(0.060)	(0.070)	(0.119)
Observations	1416	1416	1416	472	118
Lags included	6	6	12	4	1
F-stat	53.21	44.23	39.60	37.91	30.07

Notes: Newey–West standard errors in parentheses. *** , **, *, and † denote significance at 1%, 5%, 10% and 15% levels. The error structure is assumed to be heteroskedastic and autocorrelated up to the lag order shown in the table. Index of Integration is the estimate of the proportion of variation in the group of 11 bond returns explained by the first principal component. The estimation of the component is performed as the centered rolling window with the bandwidth of 156 months. Time period is chosen to always have 156 month of data to estimate the dependent variable. In-sample countries are Belgium, Denmark, France, Germany, Italy, The Netherlands, Norway, Spain, Sweden, UK, and USA. The estimation of the component is performed as the centered rolling window with the bandwidth of 156 months. Average Country Risk is the average across in-sample countries of the bond spread versus the estimate 'world' return. The 'world' return is the first principal component of country bond returns. 'Prevalence of X' denotes a fraction of countries where X occurs in a given time period. See Section 3.3 for detailed definitions of the variables. Quarterly and yearly data uses the corresponding averages of the monthly data.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

autocorrelation and to produce consistent standard errors. In column (3), I allow for 12 lags (1 year) in errors with similar results except that now the trend terms and world war dummies are insignificant at conventional levels. This result is intuitive since usually there is a great deal of persistence in monthly financial data and there is some short-term volatility in the index of integration or the monthly explanatory variables. Even so, short-term noise in the dependent variable would end-up in the error term and would not bias the coefficient estimates. It is also unlikely that the noise would affect the regression results given the long-run horizon. Still, in the remainder of the table I smooth the data by taking quarterly and annual averages of the monthly series and preserve 1 year-long lag structure in the Newey–West errors. The results in columns (4) and (5) match the results of monthly data well. The significance of the coefficients with annual averages decreases, especially for trade, perhaps due to over-smoothing or a smaller sample size.

Table 3 reports the results for index of integration from real returns. Overall, the results emerging for this measure of integration are qualitative similar to those in Table 2. In an unconditional relation in column (1) we observe the significant quadratic relationship. In multiple regression, the index of integration is positively correlated with trade openness and negatively with average inflation rate. Apparently, monetary policy is important even when the effect of inflationary expectation is removed from returns. However, there are a number of differences seen in columns (2)–(5). Government deficit now is significantly negative showing the role for fiscal policy in addition to monetary policy in explaining the variation in integration. The pegged exchange rates do not significantly correlate with it. Capital controls seem to matter by reducing co-movement of real returns, possibly because it is easier for the government to interfere with the movement of capital goods and the returns on real investments than with purely financial flows. Interestingly, the hyperinflation years, financial crises, and country risk have significant positive coefficients. I interpret this result as capturing various types of common shocks affecting returns to capital investments through expected inflation dynamics or real effects. In particular, the positive coefficient of the financial crises variable might imply that they typically affect many countries at the same time and disrupt their capital investments.

The evidence in these regressions should not be interpreted as causal. It is possible that governments would respond to changes in integration or arbitrage opportunities with certain policies. As such, establishing a causal relationship is an important but difficult task because of simultaneity in the degree of integration, policies, institutional changes, market frictions, shocks, and so on. In addition, my measure of integration does not tell what parties, private or public, dominated the sovereign debt markets in a

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 3. Determinants of bond markets integration. Real returns, 1875–2002

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: index of integration based on real returns				
Type of data	Monthly			Quarterly	Yearly
Time trend	–0.340***	0.562***	0.562***	1.712***	7.291***
	(0.086)	(0.104)	(0.138)	(0.357)	(2.317)
Time trend^2	0.295***	–0.263***	–0.263**	–2.406***	–41.045*
	(0.065)	(0.080)	(0.106)	(0.824)	(21.271)
Average trade openness		0.408***	0.408***	0.400***	0.371*
		(0.114)	(0.151)	(0.129)	(0.206)
Average annual inflation rate		–0.083***	–0.083***	–0.085***	–0.089***
		(0.014)	(0.018)	(0.016)	(0.024)
Average government deficit		–1.183**	–1.183*	–1.273**	–1.647*
		(0.531)	(0.644)	(0.599)	(0.928)
Prevalence of pegged exchange rate regimes		–0.008	–0.008	–0.010	–0.020
		(0.037)	(0.048)	(0.042)	(0.065)
Prevalence of capital controls		–0.286***	–0.286***	–0.292***	–0.314***
		(0.047)	(0.062)	(0.053)	(0.086)
Hyperinflation years		0.133***	0.133***	0.137***	0.149*
		(0.040)	(0.049)	(0.047)	(0.081)
Prevalence of financial crises		0.102**	0.102**	0.103**	0.110*
		(0.041)	(0.049)	(0.046)	(0.063)
Prevalence of consumption disasters		–0.017	–0.017	–0.034	–0.102
		(0.053)	(0.066)	(0.062)	(0.110)
Average country risk		0.010***	0.010***	0.010***	0.011**
		(0.003)	(0.004)	(0.003)	(0.005)
World Wars	–0.047	0.211***	0.211***	0.236***	0.339**
	(0.034)	(0.064)	(0.075)	(0.074)	(0.140)
Observations	1380	1380	1380	460	115
Lags included	6	6	12	2	1
F-stat	10.93	21.74	15.16	18.27	9.91

Notes: Newey–West standard errors in parentheses. ***, ** and * denote significance at 1%, 5% and 10% levels. The error structure is assumed to be heteroskedastic and autocorrelated up to the lag order shown in the table. Index of Integration is the estimate of the proportion of variation in the group of 11 real bond returns explained by the first principal component. Real returns are computed using the 'exact' Fisher formula $r_t = (1 + i_t)/(1 + \Pi_t^e) - 1$, where i_t is nominal bond yield (as in Panel A of Figure 1) and Π_t^e is the expected inflation rate. The estimation of the component is performed as the centered rolling window with the bandwidth of 156 months. Time period is chosen to always have 156 month of data to estimate the dependent variable. In-sample countries are Belgium, Denmark, France, Germany, Italy, The Netherlands, Norway, Spain, Sweden, UK, and USA. The estimation of the component is performed as the centered rolling window with the bandwidth of 156 months. Average Country Risk is the average across in-sample countries of the bond spread versus the estimate 'world' return. The 'world' return is the first principal component of country bond returns. 'Prevalence of X' denotes a fraction of countries where X occurs in a given time period. See Section 3.3 for detailed definitions of the variables. Quarterly and yearly data uses the corresponding averages of the monthly data.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

particular time period. The task of this article was more modest and included a search for broad patterns of integration that are common across countries and over a very long period. Uncovering explicable factors that are correlated with the degree of integration could point to policies and local or global institutional arrangements that are conducive to financial and real globalization.

4 Conclusion

Using a systematic methodology based on the method of principal components, I quantify economic integration in financial and physical asset markets and explore potential determinants of its long-run dynamics. This method overcomes the limitations of conventional approaches. Financial markets integration is motivated by the UIP condition with risk and is measured by co-movement of nominal returns for long-term government bonds. The integration of markets for physical assets is motivated by real arbitrage and measured the co-movement of real interest rates on the same instruments. Based on the suggested methodology, I find clear evidence of higher financial integration at the end of the 20th century compared to earlier periods. Additionally, time-series regressions show that policy variables (average inflation, government budget deficit, capital controls, and the exchange-rate regime) and the global market environment (approximated by average trade openness) played a role in explaining the time variation in the index of integration. Overall, variation in the co-movement of real bond returns is associated with a broader array of policy variables and proxies for economic shocks. I also find that 'unexplained' changes in overall level of country risk are empirically important and warrant further research concerning possible factors behind this unexplained country risk.

Acknowledgements

I thank Volker Nitsch, the participants of the 2011 CESifo Economic Studies Conference on Measuring Economic Integration in Munich, Germany, and anonymous referees for valuable suggestions and the CESifo Group for financial support.

References

Alfaro, L., S. Kalemli-Ozcan and V. Volosovych (2008), "Why Doesn't Capital Flow From Rich To Poor Countries: An Empirical Investigation", *Review of Economics and Statistics* **90**, 347–368.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Alper, C.E., O.P. Ardic and S. Fendoglu (2009), "The Economics Of The Uncovered Interest Parity Condition For Emerging Markets", *Journal of Economic Surveys* **23**, 115–138.

Backus, D.K., P.J. Kehoe and F.E. Kydland (1992), "International Real Business Cycles", *Journal of Political Economy* **100**, 745–775.

Barro, R.J. and J.F. Ursua (2008), "Macroeconomic Crises since 1870", NBER Working Paper 13940, National Bureau of Economic Research, Cambridge, MA.

Bartholomew, J.R. and A.W. Kimber (eds.) (2008), *Kimber's Record of Government Debts and other Securities*, A.W. Kimber & Company, New York, available at http://www.archive.org/details/kimbersrecordgo00unkngoog, accessed on 20 January 2009.

Bordo, M.D. (1999), *The Gold Standard and Related Regimes: Collected Essays (Studies in Macroeconomic History)*, Cambridge University Press, New York.

Bordo, M.D., M. Edelstein and H. Rockoff (1999a), "Was Adherence to the Gold Standard a 'Good Housekeeping Seal of Approval' During the Interwar Period?", NBER Working Paper 7186, National Bureau of Economic Research, Cambridge, MA.

Bordo, M.D., B. Eichengreen and D. Irwin (1999b), "Is Globalization Today Really Different from Globalization a Hundred Years Ago?", in S.M. Collins and R.Z. Lawrence, eds., *Brookings Trade Forum 1999*, Brookings Institution Press, Washington, pp. 1–72.

Bordo, M.D., B. Eichengreen, D. Klingebiel and M.S. Martinez-Peria (2001), "Is the Crisis Problem Growing More Severe?", *Economic Policy* **16**, 53–82.

Bordo, M.D. and M. Flandreau (2003), "Core, Periphery, Exchange Rate Regimes, and Globalization (National Bureau of Economic Research Conference Report)", in M.D. Bordo, A.M. Taylor and J.G. Williamson, eds., *Globalization in Historical Perspective*, University of Chicago Press, Chicago, pp. 417–472.

Bordo, M.D. and A.P. Murshid (2006), "Globalization and Changing Patterns in the International Transmission in Financial Markets", *Journal of International Money and Finance* **25**, 655–674.

Bordo, M.D. and H. Rockoff (1996), "The Gold Standard as a 'Good Housekeeping Seal of Approval' ", *The Journal of Economic History* **56**, 389–428.

Bordo, M.D. and A. Schwartz (1996), "The Specie Standard as a Contingent Rule: Some Evidence for Core and Peripheral Countries,

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

1880–1990", in B. Eichengreen and J. de Macedo, eds., *Historical Perspectives on the Gold Standard*, Routledge, New York.

Bordo, M.D. and A. Schwartz (1999), "Monetary Policy Regimes and Economic Performance: The Historical Record", in J. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*, North-Holland, Amsterdam.

Caselli, F. and J. Feyrer (2007), "The Marginal Product of Capital", *The Quarterly Journal of Economics* **122**, 535–568.

Eichengreen, B. (1994), *International Monetary Arrangements for the 21st Century*, Brookings Institution, Washington, D.C.

Eichengreen, B. (1996), *Golden Fetters. The Gold Standard and the Great Depression, 1919–1939*, Oxford University Press, New York.

Feldstein, M. and C. Horioka (1980), "Domestic Saving and International Capital Flows", *Economic Journal* **90**, 314–329.

Flandreau, M., J. Le Cacheux and F. Zúmer (1998), "Stability Without a pact? Lessons from the European gold Standard, 1880–1914", *Economic Policy* **13**, 115–162.

Flandreau, M. and F. Zúmer (2004), *The Making of Global Finance*, OECD, Paris.

Frankel, J.A. (1992), "Measuring International Capital Mobility: A Review Export", *The American Economic Review* **82**, 197–202.

French, K.R. and J.M. Poterba (1991), "Investor Diversification and International Equity Markets", *The American Economic Review Papers and Proceedings* **81**, 222–226.

Gagnon, J.E. and M.D. Unferth (1995), "Is there a world interest rate?", *Journal of International Money and Finance* **14**, 846–855.

Global Financial Data, Inc. (2002), *GFD Encyclopedia of Global Financial Markets*, Global Financial Data, Inc., Los Angeles.

Goetzmann, W.N., L. Li and G.K. Rouwenhorst (2005), "Long-Term Global Market Correlations", *Journal of Business* **71**, 1–38.

Holtemöller, O. (2005), "Uncovered Interest Rate Parity and Analysis of Monetary Convergence of Potential EMU Accession Countries", *International Economics and Economic Policy* **2**, 33–63.

Jolliffe, I.T. (2002), *Principal Component Analysis Series: Springer Series in Statistics,* second edn, Springer, New York.

Lothian, J.R. (2002), "The Internationalization of Money and Finance and the Globalization of Financial Markets", *Journal of International Money and Finance* **21**, 699–724.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Lucas, R. (1990), "Why doesn't Capital Flow from Rich to Poor Countries?", *American Economic Review* **80**, 92–96.

Mauro, P., N. Sussman and Y. Yafeh (2002), "Emerging Market Spreads: Then Versus Now", *Quarterly Journal of Economics* **117**, 695–733.

Nellis, J.G. (1982), "A principal components analysis of international financial integration under fixed and floating exchange rate regimes", *Applied Economics* **14**, 339–354.

Obstfeld, M. and K. Rogoff (2000), "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?", in B.S. Bernanke and K. Rogoff, eds., *NBER Macroeconomic Annual 2000*. MIT Press, Cambridge, MA.

Obstfeld, M. and A.M. Taylor (2003), "Globalization and Capital Markets", in M.D. Bordo, A.M. Taylor and J.G. Williamson, eds., *Globalization in Historical Perspective*, University of Chicago Press, Chicago.

Obstfeld, M. and A.M. Taylor (2004), *Global Capital Markets: Integration, Crisis, And Growth*, Cambridge University Press, New York/Cambridge.

Pearson, K. (1901), "On Lines and Planes of Closest Fit to Systems of Points in Space", *Philosophical Magazine* **2**, 559–572.

Quinn, D. P. (2003), "Capital Account Liberalization and Financial Globalization, 1890–1999: A Synoptic View", *International Journal of Finance and Economics* **8**, 189–204.

Quinn, D. P. and H.-J. Voth (2008), "A Century of Global Equity Market Correlations", *American Economic Review* **98**, 535–540.

Reinhart, C.M. and K.S. Rogoff (2004), "Serial Default and the 'Paradox' of Rich to Poor Capital Flows", *American Economic Review Papers and Proceedings* **94**, 52–58.

Reinhart, C.M. and K.S. Rogoff (2009), *This Time Is Different: Eight Centuries of Financial Folly*, Princeton University Press, Princeton NJ.

Taylor, M.P. and L. Sarno (2004), "International Real Interest Rate Differentials, Purchasing Power Parity and the Behaviour of Real Exchange Rates: The Resolution of a Conundrum", *International Journal of Finance and Economics* **9**, 15–23.

Tesar, L.L. and I.M. Werner (1995), "Home Bias and High Turnover", *Journal of International Money and Finance* **14**, 467–492.

Volosovych, V. (2011), "Financial Market Integration Over the Long Run: Is there a U-shape?", *Journal of International Money and Finance* **30**, 1535–1561.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Wei, S.-J. (2000), "Local Corruption and Global Capital Flows", *Brookings Papers in Economic Activity* **2**, 303–346.

Appendix A

A1 Construction of explanatory variables

The first variable is a proxy for pro-globalization market environment, measured by the trade openness. The *average trade openness* is defined as average over in-sample countries of exports plus imports over GDP. I supplement the historic trade data on the merchandize trade from Brian Mitchell's International Histological Statistics with the modern data on trade in goods and services by splicing the series from approximately 1960 so that the break is minimized.

The second set of variables includes measures of domestic economic policy and macroeconomic fundamentals. I use the cross-sectional *average inflation rate* as a proxy of overall laxity of government policy, such as the degree of commitment to a fixed exchange rate regime or inclination to finance excessive government expenditures. I calculate inflation rate as an ex post year-on-year change in monthly CPI, based on GFD data supplemented by data from the International Historical Statistics volumes and the IMF's International Financial Statistics database starting from 2003. In the regressions I use annual averages of these series to smooth volatility and because earlier price data is often available only at the annual frequency. In addition, to control for fiscal policy, I compute the cross-sectional *average government deficit to GDP* using annual data from Bordo et al. (2001), supplemented with the negative of the overall budget balance from the World Bank's World Development Indicators database in the 1990s and 2000s.[24] For each country in my sample, I construct the monthly binary variable equal to 1 if the country pursued an exchange rate regime other than a free flat in a given time period. I also construct the binary variable equal to 1 for periods of capital controls. Similarly to Bordo (1999), I treat the 'capital controls' broadly to include various restrictions on capital (in/out)flows or foreign currency transactions, foreign exchange controls, and other frictions related to currency convertibility. I use annual dummies from Bordo et al. (2001) and adjust them to

[24] The series is the Cash surplus/deficit, %GDP. Such flow measure was used by Bordo and Rockoff (1996) and Bordo et al. (1999a) while Flandreau et al. (1998) and Obstfeld and Rogoff (2004) advocate the stock of public debt to GDP as a better measure of overall country solvency. Besides the difficulty to construct a consistent series of debt/GDP ratio over 100+ years, Flandreau and Zúmer (2004) argue that the nominal debt is a poor measure of true indebtedness because the burden depends on the interest rate at which the debt is issued, not on its nominal amount.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

the exact months of changes in the regimes and capital controls using qualitative descriptions in this article, Bordo and Schwartz (1996), Bordo and Rockoff (1996), Eichengreen (1994, 1996), Bordo (1999), and other sources (the details of these adjustments are available from the author). The variable Prevalence of Capital Controls is the fraction of countries with restrictions on capital flows in a given time period. The variable Prevalence of Pegged Exchange Rate Regimes is constructed similarly using monthly exchange rate regime dummies.

My third set of variables includes proxies for economic shocks such as financial and economic crises or hyperinflation episodes. In order to control for the episodes of *financial crises* I refer to the chronology described in Bordo et al. (2001) and define a binary variable taking the value of 1 in the first and on-going years of banking, currency, or twin crisis, excluding the recovery period. The variable Prevalence of Financial Crises is the fraction of in-sample countries in the crisis state. I also control for '*economic disasters*', defined by Barro and Ursua (2008) as cumulative declines in consumption by at least 10% and shown to significantly affect the rates of return for stocks, bills, and bonds. Typically, GDP and consumption fall concurrently; however, I prefer using consumption disasters since these might cause more pressure on the government to change policies because of social unrest. The variable Prevalence of Consumption Disasters is the fraction of countries that have experienced extreme declines in consumption during a given time period. I also control for the *hyperination years* including the incidents covering my larger sample of 15 countries: Germany (1923), Italy (1944), Greece (1946), and Japan (1946-47) based to Bordo et al. (2001). Despite being a country phenomenon, hyperinflation could have international effects. Finally, in this group, I include the *average country risk*, computed as the cross-sectional average of individual bond spreads versus the estimated 'world' return, to capture all unaccounted country characteristics that may discourage foreign investment and thus negatively affect integration.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

CESifo Economic Studies, Vol. 59, 2/2013, 392–411 doi:10.1093/cesifo/ifs016
Advance Access publication 3 May 2012

China in the World Economy: Dynamic Correlation Analysis of Business Cycles

Jarko Fidrmuc*, Iikka Korhonen[†] and Ivana Bátorová[‡]

*Zeppelin University Friedrichshafen; CESifo Munich, Germany;
Mendel University Brno, Czech Republic. e-mail: jarko.fidrmuc@zeppelin-university.de
[†]Institute for Economies in Transition, Bank of Finland, PO Box 160, 00101 Helsinki, Finland. e-mail: iikka.korhonen@bof.fi
[‡]Comenius University, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics, Bratislava, Slovakia.
e-mail: ivana.batorova@gmail.com

Abstract

We analyze globalization and business cycles in China and selected OECD countries using dynamic correlation analysis. We show that dynamic correlations of business cycles of OECD countries and China are low at business-cycle frequencies and positive for short-run developments. Furthermore, trade of OECD countries and China lowers the degree of business-cycle synchronization within the OECD area, especially at business-cycle frequencies. Thus, different degrees of participation in globalization can explain the differences between the business cycles of OECD countries. (JEL codes: E32, F15, F41)

Keywords: Globalization, business cycles, synchronization, trade, FDI, dynamic correlation

1 Introduction

Few events in the world economy match the emergence of China in recent decades. Predominantly agrarian before 1980, China today boasts an extensive modern industrial economy with booming urban regions. The country's rapid trade growth is supported by large inflows of foreign direct investment. Not surprisingly, growth in the world's most populous country has changed the distribution of economic activities across the world. Between 1990 and 2006, the share of Chinese GDP in the world economy, valued at purchasing-power-adjusted prices, increased from 3.6% to 11.5%.

The international distribution of economic activities has important implications for business cycles. Emerging countries, particularly China, contribute significantly to global growth. Thus, global economic prospects may be less dependent than earlier on the performance of large developed economies such as the USA and Germany. This situation may make countries in a particular region less vulnerable to demand shocks.

The literature on business cycle synchronization stresses the importance of foreign trade and capital flows. Thus, the emergence of China as a large

© The Authors 2012. Published by Oxford University Press
on behalf of Ifo Institute, Munich. All rights reserved.
For permissions, please email: journals.permissions@oup.com

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

trading nation and a target for international investment may have significant effects on the business cycles of its partner countries.

Even as China has opened up to the world economy, recent business cycle trends may reflect differences among countries in their intensity of trade and financial relations with China. This seems especially important in the case of European countries. We observe a joint EU cycle up to the 1980s (Artis and Zhang 1997), which essentially vanishes in the 1990s (Artis 2003). Moreover, the intensity of trade and financial links with China differs among individual EU countries. For example, the UK, Germany, Finland, and the Netherlands have extensive links with China, while many other EU countries have quite modest economic ties.

Foreign trade and foreign direct investment (FDI) are generally seen as important drivers of business cycles. However, their effects on correlations across international business cycles are ambiguous. Frankel and Rose (1998) find a positive relationship between integration (supporting intra-industry trade according to their view) and correlation of business cycles between OECD countries. Krugman (1993) in contrast, argues that integration should induce trade specialization and thus business-cycle divergence between countries. Given China's specific position in international trade, the impact of globalization on business-cycle synchronization in China and developed countries is ambiguous.

Two major findings in our study stand out. First, the business cycle in China is quite different from OECD countries (with the exception of Korea). Second, trade with China has reduced the degree of business-cycle synchronization between OECD countries. This stands in sharp contrast to the positive relationship between trade and business cycles, which is extensively documented in the earlier literature (and confirmed here for OECD countries). To our knowledge, this result is new to the literature.

The article is structured as follows. The following section discusses the determinants of international business cycles. Section 3 introduces the concept of dynamic correlation and discusses the stylized facts of business cycles in selected developed countries and China. Section 4 describes the business cycle of China and Section 5 investigates the impact of China on the degree of business cycles synchronization between OECD countries. The last section concludes with suggestions for future research.

2 Determinants of International Business Cycles

Economic development is determined by domestic factors (e.g., aggregate demand shocks and economic policies) and international factors (e.g., external demand and international prices of traded goods), as well as their interaction. In open economies, international factors play an

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

important role, often driving the formulation of domestic policies so as to insulate the economy from adverse external economic shocks. Frankel and Rose (1998) argue that trade, and more generally economic integration among countries, results in increased synchronization of individual business cycles. They contend that trade links provide a channel for transmission of shocks across countries. Kose and Yi (2006) analyze this issue using an international real-business-cycle model. Although their model suggests a positive relation between trade and output movements, only modest qualitative effects are obtained.

The hypothesis of a positive relationship between trade and business cycles is not universally accepted. Krugman (1993), for example, argues that countries should be expected to become increasingly specialized as they become more integrated. Thus, the importance of asymmetric or sector-specific shocks should increase with the degree of economic integration—a pattern perhaps more appropriate here for explaining Chinese business cycles.

The role of trade links for international transmission of business cycles has been studied extensively in the empirical literature. Despite theoretical ambiguities, the authors generally find that countries that trade more extensively with each other exhibit a higher degree of output co-movement. The Frankel–Rose hypothesis underscores the fact that bilateral trade is mainly intra-industry trade. Given China's tendency to specialize vertically as documented by Dean Lovely and Jesse (2009), intra-industry trade may not be highly relevant to the Chinese business cycle. Instead, the specialization forces discussed by Krugman (1993) appear to dominate and drive differences in the business cycles of China and its trading partners.

Financial integration between countries could also play an important role in synchronization of business cycles, but again the impact of financial integration on business cycles is ambiguous. On the one hand, the impacts of financial markets are similar to those of trade links. Thus, the business cycle in one country is likely to affect investment decisions and asset prices in other countries via financial flows. Conversely, FDI enables countries to specialize (Imbs 2004; de Haan Inklaar and Jong-A-Pin 2008b) such that a high degree of financial integration may reduce the degree of co-movement. Here, empirical analysis seems to indicate a less robust impact of financial integration on business cycle synchronization.

The literature on business cycle correlation has focused mainly on developed economies. Few, if any, papers directly examine the correlation of Chinese business cycles versus other emerging Asian economies and OECD countries. Kose Otrok and Prasad (2008) compare business cycles of industrial countries and emerging economies, showing

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

convergence within both groups, but divergence (decoupling) between the groups of industrial and emerging economies. This decoupling of business cycles between China and developed economies has been confirmed by here Akin and Kose (2008) and Kose Otrok and Prasad (2008). Fidrmuc and Korhonen (2010) and Kim Lee and Park (2011) also show that correlation of business cycles between Asian economies and developed countries increased after the financial crisis of 2008. He and Liao (2011) use a structural factor model to assess business-cycle correlation between emerging Asian economies, including China, and the G7 countries. They find that role of global factors increased between 1995 and 2008, while Asian countries as a group remained somewhat disconnected from the G7 business cycle. Global factors mattered less for China than for Asian countries on average, but regional factors were more important.

3 Spectral Analysis and Dynamic Correlation

While analysis in the time dimension is a standard tool of business-cycle analysis, the application of spectral analysis may offer new and more robust insights. Business-cycle analysis is usually sensitive to the choice of detrending techniques (Canova 1998). Statistical filters, especially the Hodrick–Prescott filter, may generate artificial cycles (Harvey and Jaerger 1993). Moreover, the Hodrick–Prescott filter suffers from end-point bias. The band-pass filter, recommended in the more recent literature, results in a loss of observations for 3 years at the beginning and end of a time series.[1] In contrast, first differences of equal quality are available for the whole sample, but they include all frequencies. For relatively short samples (as is often case for emerging economies), static correlation may be artificially high if co-movements of cycles of different frequencies coincide in the sample. Subsamples may also display the periods of high and low business-cycle synchronization (decoupling and recoupling) commonly observed among countries (Fidrmuc and Korhonen 2010).

Spectral analysis can provide a way to avoid some of the caveats of standard business-cycle analysis. Spectral techniques enable decomposition of aggregate fluctuations into a sum of cycles of different frequencies that provides detailed information on the underlying cyclical structure of

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

[1] This would shorten our observation period from 1992:Q1-2011:Q2 to 1995:Q1-2008:Q2. As a result, the band-pass filter excludes completely the effects of the financial crisis. Therefore, we do not use the band-pass filter for comparisons with other measures of business-cycle synchronization in this article.

an economic series, while obviating both end-point bias and loss of observations. Information on short-run and long-run cycles can also be made available for economic analysis.

The first application of spectral analysis in macroeconomics occurred in the 1960s. Granger (1966) paved the way for the use of spectral analysis in economics. Currently, spectral analysis represents a promising stream of business cycle analysis (de Haan Inklaar and Jong-A-Pin 2008a), although applications are still rare. A'Hearn and Woitek (2001) discuss historical business cycles by means of spectral analysis. Hughes Hallett and Richter (2009, 2011) present spectral analyses of business cycles of Chinese regions and European emerging countries.

The spectrum can be estimated by parametric or non-parametric methods. Non-parametric methods assume that spectra for similar frequencies are also similar. Therefore, a spectrum can be estimated as a weighted average of the value of a sample periodogram, $S(\lambda)$, for frequencies λ_i and λ_j, where the weights depend on the distance between λ_i and λ_j. Thus, the non-parametric spectrum estimator can be written as

$$\hat{S}^{\mathrm{NP}}(\lambda_j) = \sum_{m=-h}^{h} \kappa(\lambda_{j+m},\lambda_j)\hat{S}(\lambda_{j+m}), \quad \text{where} \quad \lambda_j = \frac{2\pi j}{T}, \tag{1}$$

where κ denotes the kernel function (e.g., Bartlett kernel) that attributes weights to included frequencies, and h is a smoothing parameter (bandwidth).

Alternatively, the spectrum can be estimated parametrically as

$$\hat{S}^{P}(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{\left(1 - \sum_{j=1}^{p} \phi_j e^{-i\lambda j}\right)\left(1 - \sum_{j=1}^{p} \phi_j e^{i\lambda j}\right)}, \tag{2}$$

where the ϕ_j are parameters of an AR(p) process specified for autocorrelations of the variable y_t.

The most commonly used metric for co-movement between time series is classical correlation. Unfortunately, it does not enable the separation of idiosyncratic components from common co-movements and is basically a static analysis unable to capture the dynamics of co-movement. Spectral methods can also be used to analyze business-cycle synchronization between countries in the manner of correlation analysis. Granger (1969) first introduced cross-spectral techniques to economics by describing pairs of time series in frequency domain via decomposition of their covariance into frequency components. In this vein, we apply dynamic correlations as

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

proposed by Croux Forni and Reichlin (2001).[2] For two variables y_i and y_j with spectral density functions S_i and S_j and co-spectrum C_{ij} defined for the frequency λ over the interval $-\pi \leq \lambda \leq \pi$, the dynamic correlation, ρ_{ij}, is

$$\rho_{ij}(\lambda) = \frac{C_{ij}(\lambda)}{\sqrt{S_i(\lambda) S_j(\lambda)}}. \tag{3}$$

The dynamic correlation lies between -1 and 1. As it is also interesting to analyze the average dynamic correlations over a given interval of frequencies, we define an interval as $\Lambda = [\lambda_1, \lambda_2)$. The dynamic correlation within the frequency band Λ is then defined as

$$\rho_{ij}(\Lambda) = \frac{\int_\Lambda C_{ij}(\lambda) d\lambda}{\sqrt{\int_\Lambda S_i(\lambda) d\lambda \int_\Lambda S_j(\lambda) d\lambda}}. \tag{4}$$

In particular, if $\lambda_1 = 0$ and $\lambda_2 = \pi$, $\rho_{xy}(\Lambda)$ is reduced to the static correlation between y_i and y_j, i.e., corr(y_i, y_j). The dynamic correlation within the frequency band, defined in (4), can be used e.g. to measure the co-movement of business cycles of two countries, since we can select the frequency band of interest (business-cycle frequencies, or short-run and long-run frequencies) and evaluate the dynamic correlation within this frequency band. Croux Forni and Reichlin (2001) estimate the spectra and cross-spectra of analyzed time series by non-parametric methods.

4 Stylized Facts of the Business Cycle in China and Selected Countries

We use quarterly data on gross domestic production (GDP) in constant prices from IMF International Financial Statistics between 1992:Q1 and 2011:Q2. Where seasonal adjustment is required, we perform the US Census Bureau's X12 ARIMA procedure for the entire available period. All variables are taken in logarithms and first differences.

As official data for China are unavailable for a sufficiently long period, we chain data from different national and international sources, including the National Bureau of Statistics of China, the Hong Kong Monetary Authority, and the Bank of Finland database. We use national quarterly

[2] Messina Strozzi and Turunen (2009) discuss dynamic correlation in a discussion of wage developments over the business cycle. de Haan Inklaar and Jong-A-Pin (2008a) discuss alternative measures of synchronization of business cycles.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

GDP data in current prices deflated by the CPI.[3] We performed seasonal adjustment using the same procedure as for other countries. Thus, our data are available from 1992:Q1 to 2011:Q2. This allows us to assess the effect the recent global financial crisis has had on correlation of business cycles.

We test all variables for unit roots by the Dickey–Fuller GLS test, as proposed by Elliott Rothenberg and Stock (1996). This improves the power of the ADF test by detrending (Appendix Table A1). The test clearly rejects the null of unit root in outputs for all included countries. Similarly, the Kwiatkowski et al. (1992) tests fail to reject the null of stationarity for all countries. Panel versions of both tests (according to Im Pesaran and Shin 2003 and Hadri 2000) confirm these results.

As in most cited studies, we distinguish among three components of the aggregate correlation. First, the long-run movements (over 8 years) correspond to the low frequency band, below $\pi/16$. Second, the traditional business cycles (with periods between 1.5 and 8 years) belong to the medium part of the figure between $\pi/16$ and $\pi/3$. Finally, the short-run movements are defined by frequencies over $\pi/3$. Although it is usual to neglect these developments in the literature, we look at them here as the short-run dependences of economic development. This may be important in the case of China.

Figure 1 presents estimated spectra for the Bartlett kernel and the parametric estimator of autoregressive processes AR(2). Both methods yield largely similar spectra, although parametric estimators result in relatively smooth spectra. In general, the long-run and business-cycle frequencies dominate the spectra of nearly all countries including China. In contrast, the spectra for a few small open economies (Australia, Denmark, Israel, Norway, and Turkey) put more weight on the relatively short-run frequencies.

Figure 2 presents dynamic correlations of business cycles in China versus selected developed economies over the period studied. We see that cycles in China and selected economies vary significantly over the frequencies. Only a handful of countries show positive correlations with the long-run cycles of China. These countries include the non-European OECD countries (USA, Korea, Japan, Australia, and New Zealand). To a lesser degree, we see positive correlations among the long-run developments for Denmark, Norway, Italy, Israel, the Netherlands, and the UK. In general, the non-European OECD countries trade more intensively with China than the other countries in our sample, and this may help explain

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

[3] The GDP deflator is not available for the first years of the sample.

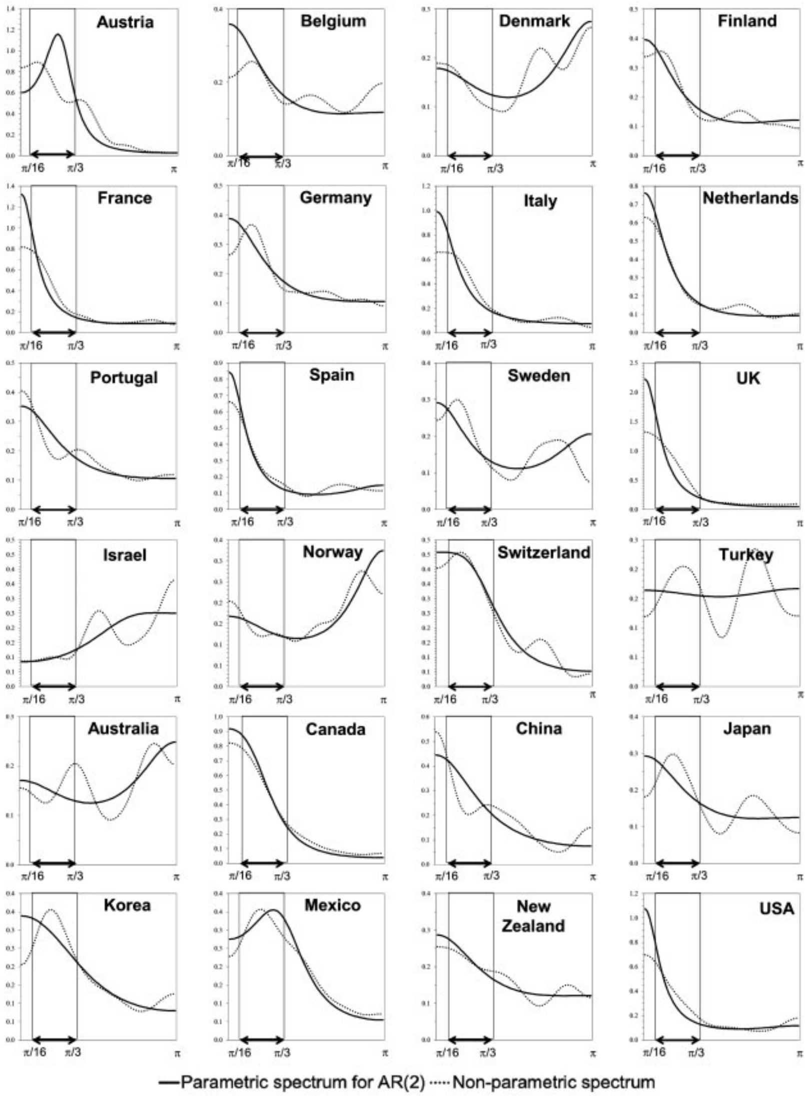

Figure 1 Estimated spectra for selected countries. *Note*: Box area and arrows denote business-cycle frequencies ($\pi/16$ to $\pi/3$). Dynamic correlations estimated using quarterly data between 1992:Q1 and 2011:Q2. *Source*: Own estimations.

the extent of dynamic correlation at long-run frequencies. For some European countries, this explanation is less credible.

We find a more homogeneous picture for the traditional business-cycle frequencies (between $\pi/16 \approx 0.2$ and $\pi/3 \approx 1$). In general, correlations of

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

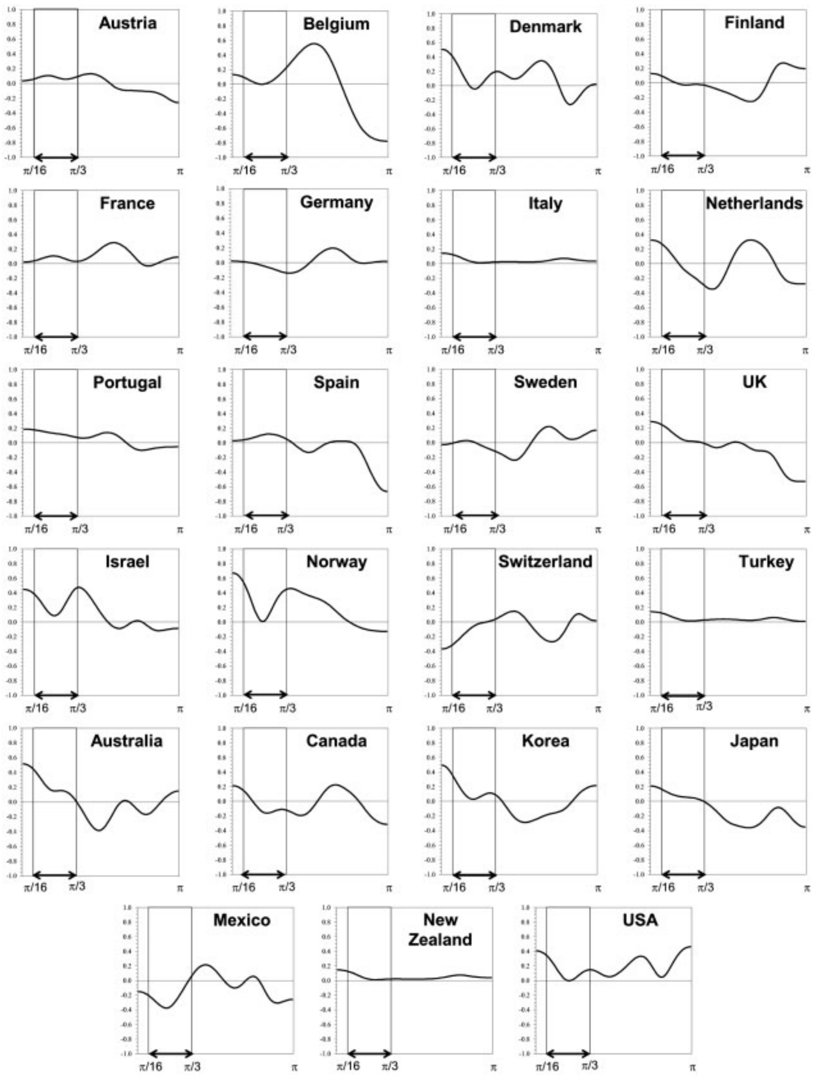

Figure 2 Dynamic correlations between China and selected countries. *Note*: Box area and arrows denote business-cycle frequencies ($\pi/16$ to $\pi/3$). Dynamic correlations estimated using quarterly data between 1992:Q1 and 2011:Q2. *Source*: Own estimations.

business cycles between China and OECD countries are low. Only Korea, Australia, Austria, Italy, Portugal, Turkey, and Israel show positive correlation over almost the whole interval of business-cycle frequencies. The positive correlation between business cycles in China and Korea

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

confirms the earlier findings of Shin and Sohn (2006) and Sato and Zhang (2006). As before, many non-European OECD countries show positive correlation at the lower range of the interval (close to 8 years). Only a few countries show positive correlation at business-cycle frequencies close to 1.5 years.

Finally, we see large differences in short-run frequencies. In general, the dynamic correlations tend to increase in this area (Figure 2). This would correspond to strong business linkages between suppliers from China and final producers in developed countries. Among the European countries, short-term correlation appears to be high for Finland and Sweden. Some short-run correlations are also high for the USA and Korea. All these countries can be characterized as having intensive trade with China over a longer period.

Figure 3 compares average dynamic correlations at business-cycle and short-run frequencies, as well as the static correlations for the sample. We see that average dynamic correlations are low for nearly all countries. At the same time, several countries show positive dynamic correlation at short-run frequencies. This is especially the case for France, Germany, Norway, and the USA.

Finally, the application of dynamic correlations confirms the evidence of decoupling of Chinese business cycles from those of the other countries.

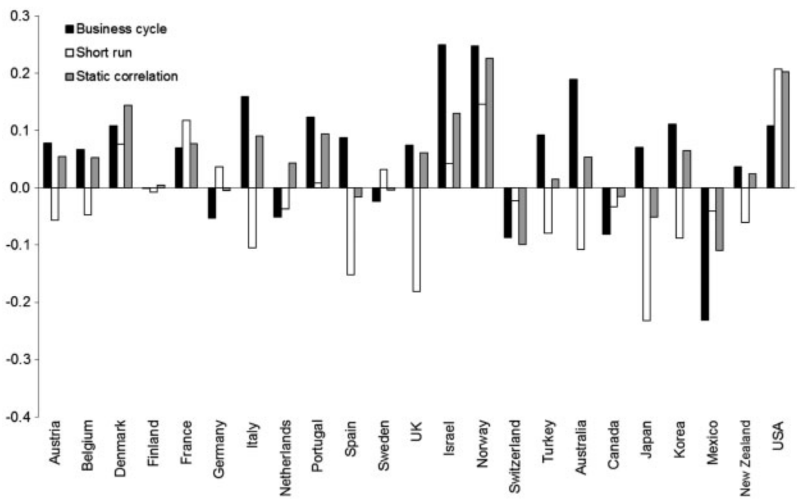

Figure 3 Average dynamic correlations in China and selected countries. *Note*: Business-cycle frequencies are the average of dynamic correlations for frequencies $\pi/16$ to $\pi/3$. Short-run frequencies are the frequencies over $\pi/3$ (cycle period less than 1.5 years). Dynamic correlations are estimated using quarterly data between 1992:Q1 and 2011:Q2. *Source*: Own estimations.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Average dynamic correlations are low (below 0.3) for business-cycle frequencies. It can be noted here that in an earlier version of this article (Fidrmuc Korhonen and Bátorová 2008) we obtained quite similar results with slightly lower dynamic correlations for data ending in 2007. Thus, the recent economic and financial crisis has had only a marginal effect on the correlation between the business cycles of China and the OECD countries.

5 Exposure to a 'Globalization' Shock and Business Cycles of OECD Countries

The findings of the previous sections show that business cycles in China and in the OECD countries are decoupled. Furthermore, the intensity of economic links with China differs substantially across OECD countries (Bussière and Mehl 2008), which can influence the business cycles of individual OECD countries. The synchronization between OECD countries may decline as a result of differing exposures to 'globalization' or 'China' shock. Alternatively, differing specialization patterns during the globalization period may also lead to increasing dissimilarities between business cycles in the OECD countries, despite similar exposure to trade and financial integration with China and other emerging markets.

Therefore, we focus our analysis on the business-cycle correlations between the OECD countries.[4] We start with estimation of the traditional equation of business-cycle synchronization, following Frankel and Rose (1998) for individual frequencies,

$$\rho_{ij}(\lambda) = \gamma_1(\lambda) + \gamma_2(\lambda) b_{ij} + \nu_i, \tag{5}$$

where ρ is the bilateral dynamic correlation at frequency λ and b_{ij} denotes the bilateral trade-to-GDP ratio for countries i and j. Using again IMF data (International Financial Statistics and Direction of Trade Statistics), we compute average trade intensity over the first decade in the data sample 1993–2003, which proxies the initial level of internationalization of OECD countries. Because estimating (5) by OLS may be inappropriate (see Imbs 2004), we use two-stage least squares.[5] This reflects the possibility that bilateral trade flows are influenced by exchange rate policies. Therefore, trade intensities have to be instrumented by exogenous determinants of bilateral trade and financial flows. Such instruments are provided by a 'gravity model' that includes the log of GDP and GDP per capita, log of distance between trading partners, and dummies for geographic adjacency,

4 We exclude the new OECD countries (Korea and Mexico) here, because they may be different than other OECD countries. Data on FDI with China is unavailable for Mexico.

5 OLS results are available from the authors upon request.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

common language, and whether the country was among the 15 earlier member states of EU or NAFTA.

Usually, equations similar to (5) are estimated for static correlation between OECD countries, the starting point of our analysis. The results are presented in the first column of Table 1. In addition, Table 1 presents results for average dynamic correlations (ADCs) for selected frequency intervals. As expected, we see that the trade coefficients estimated for the average dynamic correlations over all frequencies are nearly equal to the results for the static correlation. The same is true for the average of dynamic correlations over business-cycle frequencies. We also see that the trade coefficient is lower for the average dynamic correlation over the short-run frequencies. This means that trade mainly impacts the business-cycle and long-run frequencies. This is an interesting extension of the Frankel and Rose (1998) result.

The detailed results for the individual frequencies are reported in block A of Figure 4. We see that the highest relationship between business cycle similarities and degree of trade integration is found for the business-cycle frequencies, followed by the long-run frequencies in OECD countries. The relationship is positive, but the coefficients are lower for nearly all short-run frequencies.

In the next step, we extend equation (5) to

$$\rho_{ij}(\lambda) = \gamma_1(\lambda) + \gamma_2(\lambda)b_{ij} + \delta(\lambda)x_i + \delta(\lambda)x_j + \omega_i, \tag{6}$$

where x is a measure of economic and financial integration with China, which enters for both countries i and j. In particular, we examine the ratios of bilateral trade and FDI stocks and flows (between 2001 and 2005) recorded between OECD countries i and j to the GDP figures for the OECD countries studied. We take bilateral data on FDI in China from various issues of the China Statistical Yearbook. This shows the importance of economic and financial links from the perspective of the OECD countries. We restrict the coefficients for economic and financial integration with China, δ, to be the same for both countries, as the differences are caused by different ordering of the countries in the data matrix. This reflects the fact that we use only half of all possible combinations of n countries, because the indicators are the same (except for possible errors in trade and FDI statistics) for the country pair i and j and for the pair j and i.

The previous results for bilateral trade intensities of OECD countries remain unchanged (Table 1), if we include data for trade and financial links of OECD countries with China. Furthermore, we see that the adjusted coefficients of determination improve as well. In fact, trade flows between OECD countries explain only 11% of the variance of our measure of similarity of co-movements at the business-cycle frequencies.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 1 Estimation results for static correlation and average dynamic correlation over selected frequency intervals

	Static correlation	Average dynamic correlation	ADC: bus.-cycle frequencies	ADC: short-run frequencies	ADC: long-run frequencies
A. Basic Regression (only OECD bilateral data)					
OECD trade	1.318***	1.077***	1.409***	0.919***	1.202***
	(0.017)	(0.013)	(0.018)	(0.013)	(0.019)
Intercept	0.319***	0.242***	0.481***	0.097***	0.593***
	(0.017)	(0.013)	(0.018)	(0.013)	(0.019)
N	171	171	171	171	171
Adjusted R^2	0.132	0.164	0.110	0.143	0.056
B. Augmented Regression 1 (incl. OECD countries' trade with China)					
OECD trade	1.274***	1.055***	1.375***	0.905***	1.163***
	(0.042)	(0.033)	(0.045)	(0.035)	(0.046)
Trade with China	−1.104***	−0.621***	−0.973***	−0.413**	−1.101***
	(0.216)	(0.171)	(0.237)	(0.181)	(0.242)
Intercept	0.516***	0.351***	0.653***	0.170***	0.788***
	(0.042)	(0.033)	(0.045)	(0.035)	(0.046)
N	171	171	171	171	171
Adjusted R^2	0.259	0.225	0.192	0.166	0.160

(continued)

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 1 Continued

	Static correlation	Average dynamic correlation	ADC: bus.-cycle frequencies	ADC: short-run frequencies	ADC: long-run frequencies
C. Augmented Regression 2 (incl. OECD countries' FDI stocks in China)					
OECD trade	1.527***	1.193***	1.634***	0.985***	1.352***
	(0.022)	(0.016)	(0.023)	(0.017)	(0.024)
FDI stocks in China	0.046	0.028	0.087**	0.009	−0.024
	(0.039)	(0.029)	(0.041)	(0.030)	(0.042)
Intercept	0.293***	0.226***	0.443***	0.090***	0.591***
	(0.022)	(0.016)	(0.023)	(0.017)	(0.024)
N	171	171	171	171	171
Adjusted R^2	0.083	0.133	0.077	0.128	0.021
D. Augmented Regression 3 (incl. OECD countries' FDI flows to China)					
OECD trade	1.493***	1.162***	1.557***	0.979***	1.279***
	(0.030)	(0.022)	(0.032)	(0.023)	(0.031)
FDI flows to China	−0.699	−0.522	−0.705	−0.255	−2.111***
	(0.523)	(0.394)	(0.554)	(0.406)	(0.546)
Intercept	0.341***	0.261***	0.505***	0.105***	0.687***
	(0.030)	(0.022)	(0.032)	(0.023)	(0.031)
N	171	171	171	171	171
Adjusted R^2	0.095	0.146	0.081	0.130	0.114

Note: ADC (avg. dynamic correlation) over selected frequencies. Standard errors in parentheses. Business-cycle frequencies are the average of dynamic correlations for frequencies $\pi/16$ to $\pi/3$. Short-run frequencies are the frequencies over $\pi/3$ (cycle period less than 1.5 years). Estimations are performed for 171 country pairs for OECD countries. Dynamic correlations are estimated using quarterly data between 1992:Q1 and 2011:Q2. ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

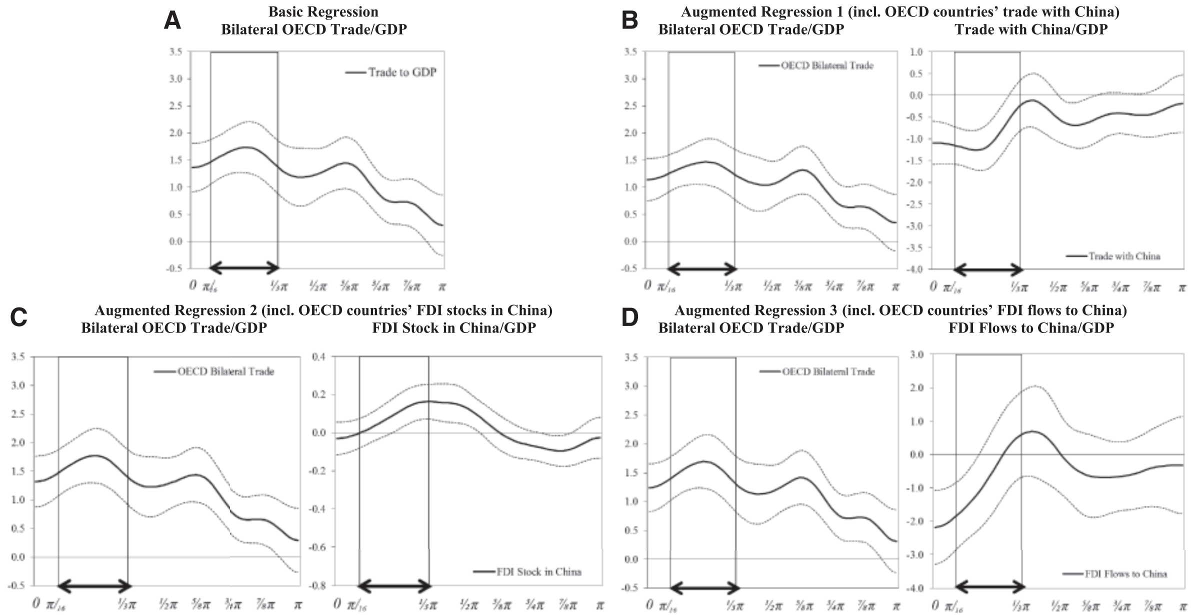

Figure 4 Regression results by frequencies, determinants of business cycle of OECD countries. *Note*: Each block of the table corresponds to a regression set, which includes the bilateral OECD trade and a proxy for links of countries to China (except basic regression). Confidence bands are for 1.96 standard errors. Box area and arrows denote business-cycle frequencies ($\pi/16$ to $\pi/3$). Estimations are performed for 171 country pairs for OECD countries. Dynamic correlations are estimated using quarterly data between 1992:Q1 and 2011:Q2. For better comparison, explanatory variables are rescaled to yield coefficients of the same magnitude.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

The inclusion of trade intensity with China explains an additional 8% of the variance in business-cycle similarities for the average of dynamic correlations for business-cycle frequencies. The share of explained variance is even higher for static correlations and average dynamic correlations for the long-run frequencies.

In contrast to trade integration between OECD countries, Table 1 and Figure 4 show that the coefficient of trade integration with China has a negative sign and is highly significant, particularly at the business-cycle frequencies. This seems to confirm our hypothesis that the high intensity of trade links with China has a negative effect on a country's synchronization with business cycles of other OECD countries. For short-run frequencies, the estimated coefficients are insignificant.

Unlike trade integration, the effects of financial integration (measured by FDI stocks and flows) on business cycle correlation are not uniform. For long-run and business-cycle frequencies, the effect of FDI flows is negative, although statistically insignificant. Figure 4 shows that the effect is practically zero for the short-run frequencies except for a narrow positive (but insignificant) interval.

The effects of FDI stocks in China on bilateral OECD country business cycle correlation are either positive[6] or zero. Notably, our results in Fidrmuc Korhonen and Bátorová (2008) were somewhat different regarding financial integration; financial integration with China also decreased bilateral business-cycle correlation between OECD countries before the financial crisis. Apparently, financial integration increased bilateral business-cycle correlation during the recent financial crisis.

In all estimations, the effects of bilateral OECD trade intensity remains positive and significant for business-cycle frequencies (especially those at the right-hand spectrum). However, the coefficients are slightly smaller in all estimations where trade with China is included.

6 Conclusions

One of the most significant economic events of recent decades has been the emergence of China as an important trading nation, and its evolution into a global heavyweight. While China has undoubtedly become an important factor in the growth of the global economy, we were specifically interested here in the extent of China's influence on business cycles in developed OECD countries.

[6] The effect is even statistically significant for average business cycle frequencies. However, Figure 4 shows that the estimated coefficients are significant only for some frequencies with cycle period close to 1.5 years.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

We show that the interdependence between business cycles in China and in developed economies is generally modest. However, many countries show a relatively high correlation for some short-run frequencies. Many transnational companies use China as a significant part of their production chain (Dean Lovely and Jesse 2009), and this is especially true for the other Asian countries. In turn, most countries show a negative correlation with China for the traditional business cycles (cycle periods between 1.5 and 8 years). This confirms the decoupling of business cycles between industrial countries and emerging economies discussed recently in the literature (Kose Otrok and Prasad 2008).

Overall, our results confirm the special position of China in the world economy, although countries with established intensive trading relationships with China (e.g., Korea and the US) have more similar cycles with China over all frequencies. Despite the increased trade links between the countries, the Chinese business cycle remains somewhat distinct from the rest of the world.

Finally, we show that countries engaged intensively in trade with China tend to have a lesser degree of synchronization of business cycles with the other OECD countries. At the same time, trade between the OECD countries increases the similarity of business cycles in the OECD countries. Both effects are less important for the short-run co-movements. Although these findings are somewhat subject to data problems, our results confirm the business-cycle dissynchronization effects of trade specialization between China and OECD countries as described by Krugman (1993), while synchronization effects prevail between the OECD countries (Frankel and Rose 1998).

Acknowledgements

We appreciate the research assistance of Yin Xia, as well as helpful comments from Volker Nitsch, Nikolaus Wolf, Gerhard Illing, Tomasz Kozluk, Michael Funke, Juraj Zeman, Pavol Brunovský, Eiji Ogawa, and seminar participants at Hitotsubashi University in Tokyo. Further thanks go to Tuuli Koivu, Aaron Mehrotra, Ayhanan Kose, and the conference participants at the Allied Social Science Meeting in San Francisco in January 2009, the CESifo Economic Studies Conference on Measuring Economic Integration in Munich in 2011, and two anonymous referees.

References

A'Hearn, B. and U. Woitek (2001), "More International Evidence on the Historical Properties of Business Cycles", *Journal of Monetary Economics* **47**, 321–46.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Akin, C. and M. A. Kose (2008), "Changing Nature of North-South Linkages: Stylized Facts and Explanations", *Journal of Asian Economics* **19**, 1–28.

Artis, M. J. (2003), "Is there a European Business Cycle?", Working Paper 1053, CESifo, Munich, http://www.cesifo-group.de/portal/pls/portal/-docs/1/1189526.PDF (last accessed 25 April 2012).

Artis, M. J. and W. Zhang (1997), "International Business Cycles and the ERM: Is there a European Business Cycle?", *International Journal of Finance and Economics* **2**, 1099–158.

Bussière, M. and A. Mehl (2008), "China's and India's Roles in Global Trade and Finance: Twin Titans for the New Millennium", *Occasional Paper No*. **80**, ECB, Frankfurt.

Canova, F. (1998), "De-trending and Business Cycle Facts", *Journal of Monetary Economic* **41**, 533–40.

Croux, C., M. Forni and L. Reichlin (2001), "A Measure of Comovement for Economic Variables: Theory and Empirics", *Review of Economics and Statistics* **83**, 232–41.

de Haan, J., R. Inklaar and R. Jong-A-Pin (2008a), "Will Business Cycles in the Euro Area Converge? A Critical Survey of Empirical Research", *Journal of Economic Surveys* **22**, 234–73.

de Haan, J., R. Inklaar and R. Jong-A-Pin (2008b), "Trade and Business Cycle Synchronisation in OECD Countries – A Re-examination", *European Economic Review* **52**, 646–66.

Dean, J., M. E. Lovely and M. Jesse (2009), "Decomposing China-Japan-U.S. Trade: Vertical Specialization, Ownership, and Organizational Form", *Journal of Asian Economics* **20**, 596–610.

Elliott, G., T. J Rothenberg and J. H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root", *Econometrica* **64**, 813–36.

Fidrmuc, J. and I. Korhonen (2010), "The Impact of the Global Financial Crisis on Business Cycles in the Emerging Economies in Asia", *Journal of Asian Economics* **21**, 293–303.

Fidrmuc, J., I. Korhonen and I. Bátorová (2008), "China in the World Economy: Dynamic Correlation Analysis of Business Cycles", Bank of Finland Institute for Economies in Transition (BOFIT) Discussion Paper 7/2008. http://ideas.repec.org/p/hhs/bofitp/2008_007.html (last accessed 25 April 2012).

Frankel, J. A. and A. K. Rose (1998), "The Endogeneity of the Optimum Currency Area Criteria", *Economic Journal* **108**, 1009–25.

Granger, C. W. J. (1966), "The Typical Spectral Shape of an Economic Variable", *Econometrica* **34**, 150–61.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Granger, C. W. J. (1969), ‘‘Investigating Casual Relations by Econometric Models and Cross-spectral Methods’’, *Econometrica* **37**, 424–38.

Hadri, K. (2000), ‘‘Testing for Stationarity in Heterogeneous Panel Data’’, *Econometrics Journal* **3**, 148–61.

Harvey, A. C. and A. Jaeger (1993), ‘‘De-trending, Stylized Facts and the Business Cycle’’, *Journal of Applied Econometrics* **8**, 231–47.

He, D. and W. Liao (2011), ‘‘Asian Business Cycle Synchronisation’’, Working Paper 06/2011, Hong Kong Institute for Monetary Research, http://www.hkimr.org/view_attachment.asp?type=22&id=270 (last accessed 25 April 2012).

Hughes Hallett, A. and C. Richter (2009), ‘‘Economics in the Backyard: How much Convergence is there Between China and her Special Regions?’’, *World Economy* **32**, 819–61.

Hughes Hallett, A. and C. Richter (2011), ‘‘Is there Clustering among the Eurozone Economies? Evidence from how the EU’s New Member States are Converging’’, *Journal of Policy Reform* **14**, 127–50.

Im, K. S., M. H. Pesaran and Y. Shin (2003), ‘‘Testing for Unit Root in Heterogeneous Panels’’, *Journal of Econometrics* **115**, 53–74.

Imbs, J. (2004), ‘‘Trade, Finance, Specialization, and Synchronization’’, *Review of Economics and Statistics* **86**, 723–34.

Kim, S, J.-W. Lee and C.-Y. Park (2011), ‘‘Emerging Asia: Decoupling or Recoupling’’, *World Economy* **34**, 23–53.

Kose, M. A. and K.-M. Yi (2006), ‘‘Can the Standard International Business Cycle Model Explain the Relation Between Trade and Comovement?’’, *Journal of International Economics* **68**, 267–95.

Kose, M. A., C. Otrok and E. Prasad (2008), ‘‘Global Business Cycles: Convergence or Decoupling?’’, Working Paper No. 14292, National Bureau of Economic Research, Cambridge, http://www.nber.org/papers/w14292 (last accessed 25 April 2012).

Krugman, P. R. (1993), ‘‘Lessons of Massachusetts for EMU’’, in F. Torres and F. Giavazzi, eds., *Adjustment and Growth in the European Monetary Union*, Cambridge University Press and CEPR, Cambridge, pp. 241–61.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992), ‘‘Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure are we that Economic Time Series have a Unit Root?’’, *Journal of Econometrics* **54**, 159–78.

Messina, J., C. Strozzi and J. Turunen (2009), ‘‘Real Wages over the Business Cycle: OECD Evidence from the Time and Frequency Domains’’, *Journal of Economic Dynamics and Control* **33**, 1183–200.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Sato, K. and Z. Zhang (2006), "Real Output Co-movements in East Asia: Any Evidence for a Monetary Union?", *World Economy* **29**, 1671–89.

Shin, K. and C.-H. Sohn (2006), "Trade and Financial Integration in East Asia: Effects on Co-movements", *World Economy* **29**, 1649–69.

Appendix A

Table A1 Selected unit root tests, first differences, 1992–2011

	DF GLS	Lags	KPSS
Australia	−9.597***	0	0.090
Austria	−3.363**	2	0.075
Belgium	−4.403***	0	0.091
Canada	−4.018***	0	0.085
China	−3.658**	0	0.199*
Denmark	−9.027***	0	0.049
Finland	−5.659***	0	0.094
France	−4.137***	0	0.118
Germany	−5.646***	0	0.044
Israel	−10.70***	0	0.128*
Italy	−4.156***	0	0.083
Japan	−6.973***	0	0.044
Korea	−6.126***	0	0.031
Mexico	−5.879***	1	0.046
The Netherlands	−3.995***	0	0.079
Norway	−10.51***	0	0.040
New Zealand	−7.355***	0	0.071
Portugal	−5.219***	0	0.143*
Spain	−4.208***	0	0.197**
Sweden	−3.649**	1	0.075
Switzerland	−5.048***	0	0.071
Turkey	−8.561***	0	0.041
UK	−3.386**	2	0.114
USA	−5.559***	0	0.062
Panel	-29.269^{IPS}***	0–1	-0.273^{PKPSS}

Note: DF GLS—Dickey–Fuller GLS test (incl. trend) of Elliott et al. (1996); KPSS, Kwiatkowski et al. (1992) test; IPS, Im Pesaran and Shin (2003) test (incl. trend); PKPSS, Panel version of KPSS tests (period 1992–2007) according to Hadri (2000). Lag structure determined according to Schwarz information criterion. ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

CESifo Economic Studies, Vol. 59, 2/2013, 412–442 doi:10.1093/cesifo/ifs009
Advance Access publication 15 April 2012

What Drives Commodity Market Integration? Evidence from the 1800s

Martin Uebele

Westfälische Wilhelms-Universität Münster. e-mail: 29maue@wiwi.uni-muenster.de

Abstract

This article provides empirical evidence from the 'first wave of globalization' in the 19th century for the question as to how commodity markets integrated domestically and internationally. I apply a dynamic factor model borrowed from business cycle analysis that for the first time allows me to fully exploit the cross-sectional and time-series dimensions of my large wheat price data set. It treats national and international market integration as conditional, and provides unique evidence on the integration of single cities as well as of countries and country groups. Three main results emerge from this: (i) The strongest push toward globalization happened in the first half of the century, not the second. This contradicts conventional wisdom emphasizing a transport revolution after 1850. (ii) After 1880, protectionist countries experienced a globalization backlash despite their well-developed transportation networks. (iii) National differences matter even when controlling for geography and trade policy. Some countries integrated domestically *after* some single cities, while others first developed a well-functioning domestic market and then globalized as a nation. The latter coincide with countries that have a long history as a unified nation. (JEL codes: N70, N71, N73, C32, F15, E32)

Keywords: market integration, 19th century, dynamic factor analysis, wheat prices, transport costs, trade policy

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

1 Introduction

Politicians have long understood the importance of trade for state finance and economic welfare. Thus, they always tried to create and tax trade. Since the Second World War, these attempts are most visible in free trade arrangements and multinational organizations such as the World Trade Organization. They aim at aligning national interests and allowing all countries to reap the fruits of trade, and are witness to the central role that tariff and non-tariff barriers play in international trade.

Given the enormous weight that political organizations have carried in international trade since the Second World War, the relevance that transport technology has in large parts of the economic literature may therefore raise some eyebrows. Perhaps the most prominent example is the 'first wave of globalization' before the First World War. The classical view emphasizes the last quarter of the 19th century as the defining period, and accredits this to reduced transport costs through the 'transport revolution' (Harley 1980; Harley 1988; O'Rourke 1997); i.e. steam-related transport infrastructure. This view is widely accepted, although it is

© The Author 2012. Published by Oxford University Press
on behalf of Ifo Institute, Munich. All rights reserved.
For permissions, please email: journals.permissions@oup.com

based on a rather small sample of price observations covering only the second half of the 19th century.

This study goes back in time until the early 1800s and analyzes 67 annual series of wheat prices in Europe and the USA. Since even the latest methods have problems capturing the full dynamics and cross-sectional variance of large data sets, I borrow a multilevel dynamic factor model from international business cycle analysis (Kose et al. 2003). It measures the degree to which the Atlantic economy, its member countries, and even single cities are integrated internationally and nationally.

Applying such a method produces three findings. First, a picture different from the well-known story of a steam-driven transport revolution emerges. While market integration in the second half of the 19th century improved impressively, earlier developments seem to have been even more notable. Confirming earlier results by Jacks (2005) and Federico and Persson (2007), this raises the question of how important the transport revolution truly was if similar developments occurred earlier in its absence.

The second finding is that after 1880 the importance of national borders increased exactly in those countries that reintroduced protectionist policies, which contrasts Jacks (2005), and once again confirms Federico and Persson (2007).

Both findings taken together allow for drawing a thought-provoking conclusion: improvements in commodity market integration can be obtained even in the absence of major infrastructure investments, while policy is able to raise national barriers despite well-developed railroads and sea ports. Revolutionary technological changes are neither sufficient nor necessary for market integration, while in the period under study political initiatives always have played an important role.

While Federico and Persson (2007) finds comparable results for the role of protectionism after 1870, I show that even within the group of protectionist nations, considerably different paths of market integration were followed in the 19th century. This leads to a partial reevaluation of Federico and Persson (2007).

Federico and Persson (2007) classify the USA as a 'free trader'. However, I find that the USA did not reach a higher international market integration than protectionist France up to 1890, and was much less well integrated than Germany another protectionist nation.

The third set of findings rests on the model's unique ability to zoom in on single cities. Some countries integrated domestically *after* some single cities, while others first developed a well-functioning domestic market and then globalized as a nation. The latter coincide with countries that have a long history as a unified nation. The evidence on single cities therefore reveals that, in addition to transport infrastructure and trade policy, more explanations are needed.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Concluding from this new evidence, it seems that the scholarly debate tends to be too focused on transport cost reductions as improving market integration. Once the focus is adjusted toward a more holistic angle, policy-makers may feel encouraged by the historical fact that trade policy matters even in the presence of different levels of transport costs. However, the interplay with additional factors such as a country's history as a unified nation seems to matter and needs further research.

In the next section, I show what this study adds to the related literature. In Section 3, I explain the empirical model. The data set presented afterwards builds on recent work, but has been extended and corrected for an important currency conversion error that plagues the data sets in Jacks (2005) and Federico and Persson (2007). I supplement the results section with an extensive discussion of the literature on early 19th century improvements in commodity market integration. The article concludes with a short summary and proposals for future research.

2 Comparison with Related Methods

This section compares the empirical setups of Jacks (2005) and Federico and Persson (2007) with mine.[1] The extensions and corrections I made to the data set by Jacks (2005) are documented in the data section.

Jacks (2005) analyzes monthly wheat prices of cities on both sides of the Atlantic. It exploits the unit root properties of bilateral price differences and reports trade cost and speed-of-adjustment estimates averaged across pairs and within 11-year overlapping time windows. The paper finds that internationally, trade costs decline quicker in the first than in the second half of the 19th century for most countries under observation, while adjustment speeds show no clear trend. Domestic trade costs declined similar to international trade costs, while domestic adjustment speeds show some upward trend in the late 19th century. The paper finds no conclusive effects of protectionism in the last quarter of the 19th century (Jacks 2005, p. 399, fn. 10).

Federico and Persson (2007) and Federico (2011) analyze price dispersion between cities year-by-year and the coefficient of variation (CoV). Due to the multilevel setup the CoV differentiates between both intranational- and international price dispersion, and reveals the difference of international price dispersion between free trading nations and protectionist countries. The findings are that international and domestic price dispersion declined quickly, especially between 1830 and 1870, and that in the

[1] I thank the editor and two anonymous referees for suggesting this.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

late 19th century price dispersion between free trading nations declined more than among protectionists.

Both studies exploit data sets that are large both in the time and in the cross-sectional dimension. However, the methods used exploit the data only insufficiently. Jacks (2005) relies on bilateral comparisons, and therefore the complexity of the empirical model grows exponentially in the number of price series, which limits the feasible number of cities to be analyzed. This may well be the reason why Jacks (2005, p. 389) does not compare all possible international city pairs. Instead, the paper analyzes only price differences between the respective national markets and Lwow, Bruges, London, New York, and Marseilles, which may bias the results.

In contrast, the dynamic factor model compares price changes multilaterally, and therefore grows only proportionally in the number of price series. If estimated using the Gibbs sampler (Section 3), the number of cities (i.e., the cross-section) may even be larger than the number of observations per city (i.e., the time dimension).

Moreover, Jacks (2005) estimates co-integration between domestic city pairs unconditionally on international price changes. To see what that means, consider the following example. Cities such as London and Liverpool are likely to have had low trade costs and high speeds of adjustments for two possible reasons: first, because they were part of the UK national market and second, because they both profited from developments in international market integration. However, the empirical setup in Jacks (2005) does not provide a gauge for how much each phenomenon mattered. Maybe as a consequence of this setup, the estimates of international trade costs bear no relationship to the respective countries' trade policies, but show the same increase for free traders and protectionists alike (Jacks 2005, Fig. 7, p. 397).

Jacks' (2005) results on single countries are largely comparable to mine, but because the units of observation are city pairs and not cities, nothing can be said as to how single cities integrated internationally or domestically, which, as I show below, differed substantially across countries.

In contrast to Jacks' (2005) time series approach, Federico and Persson (2007) and Federico (2011) exploit only the cross-sectional information in the data set, as the coefficient of variation is estimated for each single year. This is paralleled by the fact that price *levels* are analyzed, while the dynamic factor model compares price *changes*.[2]

What makes the analysis of prices in second moments more attractive than in first moments is that price changes can be decomposed into

[2] Both equal price levels and equal price changes may be interpreted as necessary conditions for market integration according to the definition by Cournot (quoted in Federico (2011, p. 95)).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

components explained by international and *additional* domestic price comovement exploiting orthogonality conditions. This yields a wholly new set of evidence and reveals that differences between the countries are not explainable by transport infrastructure and trade policy alone. The exact working of the model will be explained in the next section.

3 Empirical Setup

In this article, I borrow a technique from international business cycle analysis, a dynamic factor model, to study market integration. When compared with the methods discussed in the previous section, it has three advantages:

(1) It makes use of the time series properties of the price data.
(2) Its parameter space grows proportionally with the number of markets, not exponentially.
(3) It accounts for both the intranational and international market integration in a unified setup. This allows for investigating the development of national markets conditionally on changes of globalization.

Dynamic factor models have been used in business cycle analysis to study common cycles at different geographical levels: e.g. worldwide, among countries on the same continent, and within countries (Kose et al. 2003). When using prices of the same commodity instead of aggregate output, the model can assess market integration in the same multilevel setup. The similarity of price changes is then interpreted as a manifestation of the law of one price in second moments. The multilevel setup delivers degrees of integration among all markets, and among specified subsets.

The dynamic specification of the model implies that not only coincident correlation between time series but also correlation at different leads and lags is incorporated. This makes the model particularly suitable if leading or lagging relationships across series cannot be excluded, which would arguably be a strong assumption for 19th century commodity markets. This is its main advantage over a more traditional static factor model, which takes only contemporaneous correlations into account (Sánchez-Albornoz 1974).

3.1 Model

Similar to correlation, comovement measures linear dependence, but comovement is also defined for $N > 2$ series. A comparison of more than two series is performed by checking all series against a benchmark.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

The benchmark is found as follows. Decompose $p_{i,t}$ ($i = 1, \ldots, N$ are places, and $t = 1, \ldots, T$ are time units):

$$p_{i,t} = a_i + \lambda_i c_t + u_{i,t}, \tag{1}$$

where c_t represents the common component, which is identical for all markets. a_i is a constant and λ_i links the common component to the i-th variable. $u_{i,t}$, the idiosyncratic component, accounts for market-specific influences, e.g. local crop failures or temporary demand fluctuations.

The idiosyncratic components may experience their individual dynamics, and are expressed as AR(p)-processes:

$$u_{i,t} = \theta_{i,1} u_{i,t-1} + \ldots + \theta_{i,p} u_{i,t-p} + \chi_{i,t} \tag{2}$$

Equation 1 resembles a linear regression, only that we do not observe the regressor c_t. We can instead describe c_t's dynamics by an AR(q)-process and treat it together with Equation 2 as the transition equation in a state-space model:

$$c_t = \varphi_1 c_{t-1} + \ldots + \varphi_q c_{t-q} + \nu_t \tag{3}$$

In this basic model, each price series is explained by its comovement with all prices observed in all other places in the sample. Additional comovement may exist with cities in a subsample such as a nation. The multilevel model is formulated by adding national components to Equation 1:

$$p_{i,t} = a_i + \lambda_{i,w} c_{t,w} + \sum_{k=1}^{K} \lambda_{i,k} c_{t,k} + u_{i,t}, \tag{4}$$

where $c_{t,w}$ is the international common component, and $c_{t,k}$, $k = 1, \ldots, K$ are the national components. Zeros set by the researcher in $\lambda_{i,k}$ make sure only one $c_{t,k}$ loads on each series.[3]

The national common components are orthogonal to the international common component, which ensures that they are linearly independent. Likewise, the idiosyncratic component is orthogonal to the sum of the international and the national component.

[3] Identifying national components *ex ante* is opposed to obtaining multiple orthogonal common components endogenously and identifying them *ex post*. This is for example possible using principle components such as in Sánchez-Albornoz (1974), but does not allow for the multilevel setup. I experimented with the zero restrictions, however, and included Munich counterfactually in Austria-Hungary and Toulouse and Pau in Spain. This did not change the results for Munich, but increased the national components for Toulouse and Pau, suggesting an integrated market between southern France and northern Spain.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

The assumptions about the error terms are that the local market shocks $u_{i,t}$ are assumed to be normal with mean 0 and variance $\sigma^2_{u_i}$, and uncorrelated in the cross-section. The error term $\chi_{i,t}$ in the local market shock's process is likewise normal with mean 0 and variance $\sigma^2_{\chi_i}$, and serially and cross-sectionally uncorrelated. The common component's error term ν_t is normal with mean 0 and variance σ^2_ν and serially uncorrelated. Finally, the error of the common component ν_t is uncorrelated with the error of the local component $\chi_{i,t}$.

In estimating the model, I follow Kose et al. (2003) who use a Bayesian technique called Gibbs sampling.[4] An advantage of this estimation method is that large cross-sections can be estimated even if the time series are short. Also, the uncertainty of the results can be exactly quantified by stating the standard deviations of the estimated parameter distributions.

Gibbs sampling is based on a decomposition of the joint distribution of the common components and the parameters into conditional marginal distributions. The results are obtained by making iterative random draws from the posterior distributions derived from the model. First, a vector of arbitrary starting values is chosen for the common component. The distribution of the parameters conditional on that value is then determined and a vector of values for the parameters is sampled, which finishes the first iteration. In the second iteration, a new value for the common component is drawn conditional on the draw for the parameters from the previous iteration. Then, new values for the parameters are sampled conditional on the new common component draw. The procedure is repeated until convergence is achieved.[5]

Estimation of the multiple common components happens in a sequence; i.e., $c_{t,\mathrm{k}}$ (for $\mathrm{k} = 1, \ldots, \mathrm{K}$) are estimated for the variance unexplained by $c_{t,w}$. In each step, Gibbs sampling is applied.

The AR-order of the common components is chosen to be $q = 8$, which reflects business cycle frequency in annual data (Burns and Mitchell 1946). For the idiosyncratic processes, $p = 1$ is chosen following Kose et al. (2003). I have estimated several variations of this setup and found that the results are robust to the choice of the AR-orders.

4 GAUSS code for the model is available from Chris Otrok's website at http://people.virginia.edu/~cmo3h/research/wfac3b.prg, and code with the necessary modifications is available from the author.

5 The number of draws is 24 000 of which I use 20 000 for inference, and discard the first 4000 to minimize the impact of the random starting values. As a convergence check, I repeat the procedure several times with different starting values where the draws of different runs must not deliver significantly other results than the draws from the run before. Differences between medians are considered significant if they are larger than one standard deviation.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

There are two identification problems: first, the following two cases are observationally equivalent: $\lambda_i c_t$ and $(-\lambda_i)(-c_t)$. This problem can be solved by pinning down an arbitrary λ_i to be positive (Kose et al. 2003, p. 1219). I choose London's weight to be positively correlated with the 'international' common component, which is not a strong restriction. The cities whose prices are assumed to be positively correlated with their respective national common component are the respective national capitals if available, New York for the USA, and Santander for Spain.[6]

Second, the scale of the common component is undetermined. This is due to the fact that the variance of the common components' error term $\sigma^2_{\nu^k}$ is not identified. Following, among others, Sargent and Sims (1977) it is set to one.

The variance of the local shock's error term has an inverted gamma prior distribution with scale 6 and shape 0.001, which is fairly loose. The AR-parameters of both the common component and the local shocks have normally distributed prior distributions with zero mean, implying they are serially uncorrelated. The variance around zero decreases exponentially; i.e., the more distant the lag is, the more accurate the assumption of no serial correlation becomes. The prior distribution of the factor loadings λ is standard normal (Kose et al. 2003, p. 1221).

3.2 Interpreting the model output

The common components estimated above are used to decompose each price variance σ^2_i in the following way (with $k = 1, \ldots, K$; and suppressing the i-subscript):

$$\sigma^2 = \lambda^2_w \sigma^2_w + \lambda^2_k \sigma^2_k + \sigma^2_u \tag{5}$$

The fraction of series i's volatility explained by the international component of series i is $\widehat{\sigma^2_w} = \frac{\lambda^2_w \sigma^2_w}{\sigma^2}$.

These numbers are calculated at each iteration of the Markov chain, and from the resulting distributions, medians and standard deviations are saved. I calculate national arithmetic averages of these medians to present a better overview over the results:

$$\overline{\sigma^2_k} = \frac{1}{L_k} \sum_{i=1}^{L_k} \widehat{\sigma^2_{i,k}},$$

where $\widehat{\sigma^2_{i,k}} = \frac{\lambda^2_{i,k} \sigma^2_{i,k}}{\sigma^2_i}$ (including the subscript i), and L_k is the number of cities in country k. The variance share explained by the international component $\overline{\sigma^2_w}$ is then computed by averaging over all individual variance shares.

6 I tried changing the anchor city to Burgos but did not receive different results for Spain.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Since drawing randomly from conditional distributions yields sampling error, the orthogonality of the common components is not automatically given, although they are uncorrelated. Thus, at each step of the Markov chain the national components are made orthogonal relative to the international component by regression and proceeding with the residual. This ensures that the volatility shares add up to one (Kose et al. 2003, p. 1226).

The model is estimated separately four times in the (roughly even) subperiods 1806–1830, 1831–1855, 1856–1880, and 1881–1907. This choice was made to divide the 102-year period starting in 1806 into roughly equal subperiods, and because they can be expected to characterize major phases of commodity market formation in 19th century Atlantic economy. The first quarter captures a period of a fragmented Atlantic market, as a result of the Napoleonic Wars and the British Corn Laws. The next period up to the mid-1850s may exhibit increasing market integration as fewer wars were fought on the European continent, and liberal trade politics spread. Steam ships proliferated, and the railroad was introduced, but both were initially not used for low-value goods such as grains. The following period, 1856–1880, should continue that development although it includes the American Civil War, which is likely to have had a severe impact on world wheat trade. In the same subperiod, tariffs were reduced due to the treaties induced by Cobden–Chevalier, which however had little effect on wheat trade, which had already been liberalized (Lampe 2009). The last subperiod starting in 1881 is likely to exhibit a strong drive toward Atlantic market integration according to O'Rourke and Williamson (1999). On the other hand, some countries increased tariffs that had been lowered or abolished earlier in the century.[7]

Note that results for each period were obtained with a different number of cities, adding price series as they became available. To ensure robustness, all calculations were also performed holding the number of cities constant. The major results are independent from the variation of the number of cities, see Appendix A, Table A2.

Finally, when discussing the estimated results, the reported standard deviations can be used to calculate error bands by adding them to or subtracting them from the medians. Two medians are considered to be different (similar to a significance test in classical statistics) when the difference equals at least one standard deviation.

[7] Experimenting with different subperiods did not change the qualitative results. See Table 3.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 1 Countries, cities and total number of cities per country

Sample composition							
Austr.-H. 5	Belgium 3	Britain 12	France 12	Germany 4	Spain 12	Sweden[a] 11	USA 8
Vienna	Brussels	London	Paris	Berlin	Santander^{2}	Stockholm	New York
Ljubljana	Ghent	Manchester	Bayeaux	Hamburg	Cordoba^{2}	Uppsala	Philad.
Lwow	Bruges	Liverpool	Saint-Brieuc	Königsb.	Gerona^{2}	Södermanland	Alexandr.
Krakow		Exeter	Toulouse	Munich	Granada^{2}	Östergötland	Cincinnati^{2}
Budapest^{4}		Carmarthen	Bordeaux		Lerida^{2}	Kalmar	Indianap.^{3}
		Dover	Chateauroux		Segovia^{2}	Halland	Ithaca^{3}
		Gloucester	Mende		Oviedo^{2}	Skaraborg	San Franc.^{3}
		Worcester	Barleduc		Zaragoza^{2}	Örebro	Chicago^{3}
		Cambridge	Arras		Burgos^{2}	Vöstmanland	
		Norwich	Pau		Coruna^{3}	Gästrikland	
		Leeds	Lyon		Toledo^{3}	Hälsingland	
		Newcastle	Marseille		Leon^{3}		

[a]Names refer to historical administrational regions.

Superscripts refer to the period the city enters the data set, where ‘2’ = 1831–1855, ‘3’ = 1856–1880, ‘4’ = 1881–1907.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

3.3 Data

The data set is composed of prices provided in Jacks (2005), Jörberg (1972, Sweden), and Jacobs and Richter (1935, Germany). The data set contains between 48 and 67 annual wheat price series ranging from 1806 to 1907 and includes prices from the markets as shown in Table 1.

My data do not cover Italy and Russia because data for Italy end in the 1890s and start very late for Russia. For Germany, there are other series available apart from the four used here, but not for the years after the 1860s (Oberschelp 1986). See Uebele (2010) for a discussion.

Since this empirical model only measures relative price changes, it is not necessary to express them in the same units as long as the units do not change. However, this matters for other methods such as the CoV and for cointegration based trade cost estimates. For completeness, I corrected a conversion error in the data set by Jacks (2005) (also used by Federico and Persson 2007). Some prices are erroneously expressed in greenbacks, not in gold dollars. This causes spurious international exchange rate fluctuations for US-American, Belgian, and French prices. I deflated these by the greenback-dollar exchange rate 'XRUSGLDD' from the Global Financial Database. See Uebele (2009, pp. 156–160) for details.

All series were detrended with a Hodrick–Prescott filter ($\lambda = 6.25$). First differencing, Baxter–King and Christiano–Fitzgerald filters yield comparable results (Baxter and King 1999; Christiano and Fitzgerald 2003). See also Table A1 in the Appendix A.

4 Results

This section starts by discussing price comovement across all cities in the sample, and how my results are reflected in the literature. It then describes the anti-global effect of protectionist policies at the end of the 19th century, before elaborating upon how market integration differed at the city level between countries.

4.1 The timing of the first wave of globalization

4.1.1 *Evidence*

Consider Table 2, which contains the full sample averages of the variance shares in the first row. While in the first period only 35% of price fluctuations are shared internationally, this number increases to 48% in the period 1831–1855 (Column 4). After 1855, growth of international price

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 2 Results (full sample averages)

	1806–1830			1831–1855			1856–1880			1881–1907		
	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC
Nr of markets		48			58			66			67	
World	0.35	0.48	0.17	0.48	0.38	0.15	0.61	0.27	0.12	0.63	0.23	0.14
	(0.11)	(0.12)	(0.04)	(0.03)	(0.04)	(0.02)	(0.04)	(0.04)	(0.01)	(0.02)	(0.02)	(0.01)
Austria	0.15	0.41	0.44	0.55	0.12	0.32	0.77	0.12	0.12	0.57	0.38	0.05
	(0.04)	(0.08)	(0.08)	(0.05)	(0.08)	(0.07)	(0.04)	(0.05)	(0.04)	(0.02)	(0.02)	(0.01)
Belgium	0.56	0.40	0.05	0.94	0.04	0.03	0.67	0.32	0.01	0.90	0.08	0.02
	(0.15)	(0.15)	(0.02)	(0.04)	(0.04)	(0.01)	(0.06)	(0.06)	(0.01)	(0.01)	(0.01)	(0.01)
France	0.55	0.31	0.14	0.73	0.18	0.09	0.67	0.27	0.06	0.71	0.18	0.10
	(0.17)	(0.17)	(0.03)	(0.04)	(0.04)	(0.01)	(0.05)	(0.05)	(0.00)	(0.01)	(0.02)	(0.01)
Germany	0.21	0.57	0.22	0.81	0.06	0.13	0.89	0.01	0.10	0.82	0.11	0.08
	(0.07)	(0.08)	(0.05)	(0.03)	(0.04)	(0.03)	(0.05)	(0.06)	(0.03)	(0.01)	(0.02)	(0.02)
UK	0.48	0.44	0.07	0.64	0.33	0.03	0.90	0.09	0.01	0.92	0.04	0.04
	(0.17)	(0.16)	(0.02)	(0.03)	(0.03)	(0.00)	(0.04)	(0.04)	(0.00)	(0.02)	(0.02)	(0.00)
USA	0.39	0.50	0.11	0.24	0.66	0.10	0.39	0.44	0.17	0.72	0.21	0.07
	(0.07)	(0.10)	(0.08)	(0.03)	(0.03)	(0.02)	(0.05)	(0.05)	(0.01)	(0.02)	(0.02)	(0.01)
Sweden	0.06	0.68	0.26	0.11	0.74	0.15	0.71	0.11	0.18	0.45	0.29	0.26
	(0.04)	(0.05)	(0.04)	(0.02)	(0.03)	(0.01)	(0.04)	(0.04)	(0.01)	(0.02)	(0.03)	(0.02)
Spain	–	–	–	0.17	0.48	0.35	0.18	0.60	0.22	0.24	0.44	0.32
	–	–	–	(0.03)	(0.04)	(0.02)	(0.04)	(0.04)	(0.01)	(0.01)	(0.03)	(0.03)

Logged and Hodrick-Prescott (6.25) filtered prices.
Median values of Bayesian parameter distributions. Standard deviations in parentheses.
Results may vary and not add up to one due to sampling error and rounding.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

comovement does not increase anymore (Column 7).[8] Additionally, the uncertainty around the international variance share (expressed as standard deviation in brackets) declines from 0.11 to 0.03 between the first and the second period, which shows that after 1830 the price series lend themselves easily to the underlying single factor structure and can be estimated with a high degree of confidence.

This is at odds with the mainstream view of the first wave of globalization. The literature of the 1980s brought about the view that technological advances revolutionized international trade in the second half of the 19th century (Harley 1980; Harley 1988, O'Rourke 1997). These studies show price convergence between single markets across the Atlantic after 1850. They do not analyze, however, the evolution of market integration in the half-century before.

The evidence presented here casts doubt upon new transport technology as the main reason for the increase in 19th century international market integration. This is because the strongest push to market integration was made during the first half of the 1800s, when it is unlikely that the railroad and the steamship had reached their full commercial potential for staple goods (Findlay and O'Rourke 2005, pp. 35–36; Keller and Shiue 2008, pp. 14–16).

While these findings cast serious doubts upon allegedly established results, their validity may be questioned. However, earlier evidence by Federico and Persson (2007) and Jacks (2005) point in the same direction, although based on less comprehensive statistical methods. My study employs more dynamic information than Federico and Persson (2007), and uses more cross-sectional evidence than Jacks (2005).

4.1.2 *Discussion*

How is this result reflected in the wider literature on 19th century commodity market integration? In order to keep track of the different approaches to trade and commodity markets in the literature, I am going to group them in demand and supply arguments. I further deconstruct supply arguments in technical and non-technical ones.

I start with the non-technical supply-side arguments that include political trade liberalization, and organizational improvements. North (1958; 1968) develops various arguments about why transport costs declined

[8] The international component would be larger in the third subperiod if it was not for Belgium's surprisingly low value here. Still, the results are an improvement over Jacks (2005), where international trade costs increase implausibly fast in the 1860s, which is clearly due to the wrong exchange rates used. Belgium suffered more than others from the harvest failures in the early 1870s, and needed to import large amounts of wheat at high prices (Crawford 1895, Table VI). However, the problem regarding the conversion to gold seems not to be fully solved yet, since for example comparisons of silver prices in the literature do not show Belgian prices differing from the UK (Crawford 1895, p. 79).

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

early in the 19th century even in the absence of 'revolutionary' new technologies. Among others, better commercial and navigating skills led to less port time and therefore increased the capital utilization of ships. Also, better organization helped to find more suitable return cargoes, increasing the profitability in both directions.

The literature on trade liberalization in the 19th century is perhaps best summarized by three political initiatives: first, the repeal of the British Corn Laws, paving the way for wheat imports as Britain's population grew (Schonhardt-Bailey 2006; Sharp 2010); second, the German customs union that lead to the formation of a large commodity market on the European continent (Dumke 1991). Finally, the network of bilateral trade agreements between European nations of the 1860s, better known as the Cobden–Chevalier network.[9] Of course the increase of import tariffs in the late 1870s in a number of net grain importers represented a substantial protectionist backlash.

On the technical side, there are 'non-revolutionary' improvements as brought forward by Kaukiainen (2001), and Brautaset and Grafe (2005). Kaukiainen (2001) compared how fast speed of information changed before 1860 as compared to the gain of information speed through the introduction of the telegraph shortly after 1860. He finds that between London and the world, business letters gained more in speed between 1820 and 1860 than news transmitted by cable after 1860. He attributes this largely to the introduction of coastal steam ship connections, and better road networks (but not to the railway), and emphasizes that the process was demand driven, and not primarily caused by the invention of steam ships (Kaukiainen 2001, p. 21).

In the same spirit as Kaukiainen, but dealing directly with commodity transport, Brautaset and Grafe (2005) collects an extensive data set of freight rates. The authors show that transport costs per unit of the Norwegian shipping fleet declined already in the first half of the 19th century for a wide set of products and routes. (At that time, the Norwegians commanded the world's third largest shipping fleet.) The authors attribute their findings primarily to larger ships and higher frequencies of travel, thereby constructing economies of scale argument in the production of shipping services. It says that in the first half of the 19th century, by increasing the volume of traded commodities, transport costs per unit could be brought down while transport technology was not radically changed.

Note that while both papers develop supply-side arguments, these arguments are ultimately rooted in the demand of transport services.

[9] As can be learned from Lampe (2009, p. 1016), however, the Cobden–Chevalier network had virtually no effect on wheat trade as wheat customs were already very low.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Two aspects are important here: one is the impact of overall economic growth on trade, and the other is increasing international specialization. Jacks et al. (2008) shows that for the period after 1870, about half of international trade increase can be attributed to economic growth alone. This would substantially reduce the phenomenon left to be explained.

Whether this argument can be successfully applied to the first half of the 19th century depends on both the growth rate of the international economy and the increase in international specialization, especially into commodity exporter and importer nations. According to the Crafts–Harley view of the industrial revolution, British economic growth advanced most in the early 19th century, accompanied by rapid structural change (Crafts 1985). This triggered demand for primary products; in the case of wheat, Britain had become a net importer already in the late 18th century (Fay 1932, p. 25). Even if the other European nations such as Germany and France were lagging behind Britain, their industrialization and thus their economic restructuring set in well before mid-century (Spree 1977; Lévy-Leboyer and Bourguignon 1990). Despite the lack of quantitative studies, I presume that their economic growth triggered substantial demand for shipping services even in the absence of transport cost declines. As argued above, it brought about dynamic effects through scale economies in the production of transport services. These lowered unit transport costs and thus created even more incentives for trade.[10]

4.2 Protection and de-globalization

Comparing the development of international and national market integration, I observe considerable differences nations (Table 2). The integration of UK wheat trade into the Atlantic market improved throughout the whole 19th century. It started with an international component of 48% in the first period and ended with 92% in the last period. Domestically, the British economy was already well integrated early in the 19th century with an average local component of only 7% in Period 1 declining to 4% in Period 4 after 1881.

Belgium's experience is similar with overall well-integrated markets in all subperiods as can be seen from the small local components.[11] While the country was initially domestically well integrated to an extent of 40% (internationally 56%), later in the century its prices aligned more with the outside world: 90% of price changes were determined internationally in the last quarter of the 1800s.

[10] This is in my view a somewhat neglected aspect in O'Rourke and Williamson (2002, p. 777), where the authors argue that price convergence and not trade volume is the only unambiguous indicator for market integration analysis.

[11] On the exceptional low international market integration in period 3 see Footnote 9.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

In contrast to the experiences of Belgium and the UK, Sweden, Austria-Hungary and Germany exhibit increasing national components between the third and the last quarter of the 19th century. Sweden's national comovement increases from 11% to 29%, Austria-Hungary's from 12% to 38%, and Germany's from 1% to 11%, while their international components decrease accordingly. Similarly, France and Spain do not globalize in the second half of the 19th century but show constant international variance shares of about 20 and 70%, respectively, leaving more than 40% of price swings in the last subperiod determined domestically in Spain and almost 20% in France.

The fact that UK and Belgium are internationally the best integrated can easily be reconciled with the liberal trade policies of the UK and Belgium (O'Rourke and Williamson 1999). In contrast to these free trading nations, the other countries in my sample resorted to protectionism at the end of the 19th century. These countries are the same as those whose domestic price components increased after 1880 or at least did not decline, and whose international variance shares declined or did not increase: Sweden, Germany, France, Spain, and Austria-Hungary. This shows that protective tariffs did not only drive up price gaps (O'Rourke 1997), but effectively reduced the impact of international shocks on domestic economies. The fact that local price components did not increase significantly between the third and the last period (with the exception of Spain) shows also that de-globalization was driven by national factors such as tariff policies and not by local shocks.

Surprisingly, these results are not found in Jacks (2005, p. 397). There, international trade cost estimates do not increase more for protectionist countries than for free traders in the late 19th century, and their overall level increases only very slightly.

At the same time, Federico and Persson (2007) does find higher price dispersion between protectionists and free traders, and between protectionists. What the present study adds, however, is that the USA were less well integrated than their usual role as the largest wheat exporter in the 19th century suggests, and that Germany's northern cities were surprisingly well integrated even after 1879.

The importance of the USA is illustrated by their role as the world's largest wheat exporter in the second half of the 19th century, and by the fact that it is usually classified as a 'free trader' (Federico and Persson 2007). In fact, the USA raised considerable import tariffs on grains (Lampe 2009, p. 1020), but since they were a net-exporter of wheat, it is legitimate to expect no effect on their degree of market integration. Still, I find a larger national variance share for the USA (21%) than for Germany (11%) in the period from 1881–1907, and that the USA did globalize more than France only after 1890 (Table 3). As late as between 1881 and 1907,

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table 3 Subperiod robustness

	1816–1840			1841–1865			1866–1890			1891–1907		
	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC
All countries	0.39	0.43	0.18	0.64	0.28	0.08	0.69	0.21	0.09	0.72	0.19	0.09
Austr.-Hung.	0.31	0.30	0.39	0.74	0.07	0.19	0.66	0.24	0.11	0.67	0.31	0.03
Belgium	0.69	0.25	0.05	0.89	0.10	0.01	0.68	0.28	0.03	0.93	0.06	0.02
France	0.71	0.12	0.17	0.74	0.21	0.05	0.55	0.37	0.08	0.54	0.39	0.07
Germany	0.59	0.20	0.21	0.83	0.04	0.12	0.81	0.12	0.07	0.89	0.08	0.04
UK	0.36	0.56	0.08	0.82	0.16	0.02	0.93	0.04	0.03	0.94	0.05	0.02
USA	0.23	0.69	0.08	0.20	0.76	0.04	0.79	0.18	0.04	0.75	0.23	0.02
Sweden	0.02	0.71	0.26	0.27	0.57	0.15	0.53	0.27	0.20	0.55	0.19	0.26

Logged and Hodrick-Prescott (6.25) filtered prices.
Median values of Bayesian parameter distributions. Standard deviations omitted.
Based on 48 cities in all subperiods.

the international variance share is 72% for America, compared to 71% for France (not different considering the uncertainty of the estimates). These results for the USA may reflect the distance to the European markets and the size of the US internal market.

The higher degree to which German markets globalized in the last period is remarkable (82%, compared to 72% in the USA). This can partially be explained by a sample composition effect: three out of four German markets are major northern cities and important wheat trading places (Königsberg, Hamburg, and Berlin). Evidence for smaller towns before 1855 shows that not all German wheat prices mirrored international conditions that well (Uebele 2010). Having said that, the result remains remarkable especially after Germany raised protective tariffs against wheat imports in 1879 (O'Rourke 1997).

France experiences less overall change, especially compared with Germany. Already in the first period, France's international price component is considerable with 55% and does not increase dramatically afterwards. At the same time, the national component ranges between 18% and 31%. Germany, in contrast, begins with a lower level of only 21%, jumps to 81% after 1830 and then stays above 80%. Its national component reaches 10% only after 1880. These findings may reflect the different national building histories with France's national unification well before the 19th century, and Germany's late nation building process culminating only in 1871.

4.3 Single cities in a global context

With the dynamic factor model it is possible to track the degree of national and international market integration even for single cities. This section

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

shows how these processes evolved over time and offers tentative explanations. Table A3 (in three parts) documents the findings discussed here.

Two categories emerge from scrutinizing the results for single cities: one for countries where single city market integration happened similarly to each other, and one for countries with large and consistent differences between single cities.

The first group consists of the UK, France, Belgium, and Sweden. Here we find only small differences between single city results. Exceptions with high local components are Pau and Saint-Brieuc in France, and Carmarthen in the UK. Apart from that, less than 20% of price changes are due to local shocks in France, and usually well below 10% in Belgium and the UK.

In Britain, some markets are earlier integrated internationally than others: Exeter's prices vary with Atlantic prices to a degree of 76%, followed by Dover with 61% in the first period. After 1830, however, all markets except Liverpool have international components of about 60%. The differences between cities are even smaller than in tiny Belgium, and all cities are almost perfectly synchronized with the Atlantic economy after 1855 with international components of about 90%. In France, there is more variation between cities, but no apparent tendency of certain regions to be domestically separated or much earlier integrated internationally than others. In Sweden, local components are even larger and differ more than in France, but without a consistent pattern, either. Despite the heterogeneity between Sweden's regions, most of them globalized after 1855 and de-globalized after 1880.

The picture looks different in Germany, Austria-Hungary, Spain, and the USA. In Germany, Munich is separated from the domestic market with not more than 15% of its price variations determined domestically in the whole 19th century, and local components of 20–30% throughout. In contrast, Hamburg, Berlin, and Königsberg are well synchronized with each other as well as with the Atlantic economy. Their international components reach 80–90% from the second quarter on.[12]

In Austria-Hungary, the western part globalized already in the first half of the 19th century, while its eastern markets were separated from both the Atlantic economy and the national market. In Period 1, Vienna and Ljubljana have high domestic variance shares, which indicates domestic integration within Austria, while Krakow and Lwow are separated with local shares of more than 70%. After 1830, the two Austrian markets globalize with international variance shares of 60–70%, and domestic variance shares below 20%. Between 1856 and 1880, all markets of the

[12] Königsberg seems to be completely detached in the first subperiod. However, it is well integrated domestically with Berlin and Hamburg, while Munich's price changes, determined to 51% internationally, differ strongly from the other German markets.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Habsburg Empire share their price changes internationally to a high degree (between 64% for Vienna and 88% for Krakow). A considerable domestic market evolves only after 1880, when the prices in all cities (including Budapest) are determined nationally to a degree of 30–45%.

Within the USA, local components are quite low with the exception of San Francisco, which is completely separated between 1856 and 1880 with a local component of 94%. In the last period, San Francisco's degree of separation is cut down to 30%, but this benefits only the international component, which rises from 2% to 70%. The markets in the Midwest and on the East Coast have at the same time similar domestic and international components. Between the third and the fourth period, those markets' international components increased from 32–49% to 65–79%, and the domestic components shrunk from 38–61% to 17–33%.

Quite unique in this sample is Spain, because no trend toward more market integration is conceivable, neither domestically nor internationally. Still, regional similarities can be identified. In the south, Cordoba and Granada both reduced their separation from the domestic market by at least half until the end of the 19th century, starting from a level of 40% in Period 2. The four markets of the inner plateau, Toledo, Segovia, Leon and Burgos, are quite well integrated domestically after mid-century with domestic shares of more than 70%, but not internationally. This is in contrast to the cities on or close to the Atlantic coast, Oviedo, La Coruna and Santander, which feature domestic shares of not more than 55%, and frequently much below. Despite their access to the Atlantic, their international integration is almost non-existant, especially after 1880 with international variance shares below 10%. Finally, Gerona in Catalonia shares 57% of its price changes internationally, which in the last period is in line with Zaragoza and Lerida, two cities close by.

It should have become apparent that transport costs and trade policy alone provide an insufficient explanation for these findings. A third factor may be called the length of a country's history as a unified nation. The UK, France, and Sweden from the first group have been unified nations well before the beginning of the 19th century and therefore are more likely to share a common set of institutions supporting high internal commodity market integration despite differing levels of domestic transport costs. Britain profits from cheap water transport in contrast to Sweden and France, which explains Britain's higher degree of early domestic market integration. Belgium became unified only in 1830, but Brussels, Bruges, and Ghent are geographically close, and the region has been one of the most commercially advanced in Europe since early modern times (De Vries and van der Woude 1997).

As for the domestic market integration of the USA, my results indicate that the East Coast and the Midwest were domestically well integrated

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

while the West Coast was never part of the former network, despite completion of the 'Pacific Railroad' in 1869 (see Cooper 2010). Distinguishing between comovement shared only domestically and internationally, the dynamic factor model reveals that the similarity between San Francisco's and other American prices was driven by international markets and not by domestic ones.

Germany's late national unification is clearly reflected in Munich's separation from the German market until the last quarter in the 1800s and its only partial domestic integration. In a similar way, Krakow's and Lwow's isolation from the Habsburg Empire reflects their geographic distance from Austria as well as their loose political ties to the empire, especially Krakow's, which was not part of the empire between 1809 and 1846. The development of a domestic market may be too late to be pinned down to the customs union of 1850 and may rather be reconciled with external tariffs raised in the late 19th century (Komlos 1979; Lehmann and O'Rourke 2011).

Spain, another protectionist nation, is the worst integrated in the sample, both nationally and internationally. Apart from tariff policies, the low internal market integration may also be explained with missing waterways due to its geography, while the international separation of the Atlantic coast markets is an especially clear signal for the effectiveness of Spain's prohibitive trade policy until the 1860s and import tariffs in the late 19th century (Pena and Albornoz 1984; Jacks 2005; O'Rourke et al. 2007; Federico 2011).

When trying to square these results with the literature, only country studies can be consulted, for there is no study so far that discusses international and domestic market integration in a unified setup discussing the integration of single cities within nations. However, this study shows that it is relevant to assess each country's path of market integration in the context of globalization, and not as an isolated process. The qualitative discussion thus identified three key factors driving market integration in the 19th century.

5 Conclusion

This article presents an unconventional reading of 19th century commodity market integration. To obtain these results, the article uses a large annual wheat price data set and employs a method that allows for analyzing international and national price comovement in a unified setting. Loosely spoken, it allows for seeing the forest and the trees at the same time.

Three results emerge from this venture:

(1) While in the bulk of the literature transport cost reducing technology is emphasized, I argue that the timing of increases in price comovement

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

found here does not fit that logic. As an explanation, the article relates the findings to the recent literature on demand-side driven trade growth, non-revolutionary technology improvements and variations of the organizational as well as political environment in which trade occurs.

(2) For the period after 1880, despite the widespread introduction of steam-related transport infrastructure, the evidence presented here shows that protectionist nations effectively deglobalized by experiencing less international and stronger nation-specific price variations. In contrast, prices in free-trading nations moved along with the latent world price. The US markets behaved rather like those of protectionist nations in this era and less like those of free traders.

(3) National market integration paths differed with respect to their single cities. In one group of countries, most cities behaved similarly, while in another group, single cities integrated earlier than others, and a domestic market emerged only at the end of the 19th century. Thus, in addition to geography/transport costs and trade policy, a country's history as a unified nation may account for these differences.

Of these results, the last particularly adds value to the literature, while the first two corroborate or augment earlier results. The first confirms findings by Jacks (2005) and Federico and Persson (2007), however with a method that uses both time-series *and* cross-sectional information. The second result is basically in line with Federico and Persson (2007), but goes beyond Jacks (2005) that does not find any effects of protectionism.

More research is needed in order to get a firmer grip at the explanation for the European commodity market integration in the first half of the 19th century. Especially, the period of the Napoleonic Wars has been severely neglected so far, mainly due to data problems (but see Federico 2011). For the period immediately following, works such as Kaukiainen (2001) and Brautaset and Grafe (2005) demonstrate that quantitative analysis is both possible and relevant.

While this seems to be an entirely scholarly debate, the conclusion drawn for policy-makers is that markets can improve despite the lack of major technological changes, and that protectionism threatens to roll back market integration even in the presence of well-developed transport infrastructure.

References

Baxter, M. and R. King (1999), "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time-Series", *Review of Economics and Statistics* **81**, 575–593.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Brautaset, C. and R. Grafe (2005), "The Quiet Transport Revolution: Norway's Sailing Fleet and 19th Century Transport Costs", *Oxford University Economic and Social History Series No. 62*.

Burns, A. F. and W. C. Mitchell (1946), *Measuring Business Cycles*, National Bureau of Economic Research, New York.

Christiano, L. J. and T. J. Fitzgerald (2003), "The Band Pass Filter", *International Economic Review* **44**, 435–465.

Cooper, B. C. E. (2010), *The Classic Western American Railroad Routes*, Chartwell Books, New York, USA, Worth Press, Bassingbourn, UK.

Crafts, N. F. (1985), *British Economic Growth During the Industrial Revolution*, Clarendon, Oxford.

Crawford, R. (1895), "An Inquiry into Wheat Prices and Wheat Supply", *Journal of the Royal Statistical Society* **58**, 75–120.

De Vries, J. and A. van der Woude (1997), *The First Modern Economy: Success, Failure, and Perseverance of the Dutch Economy, 1500–1815*, Cambridge University Press, Cambridge.

Dumke, R. H. (1991), "Tariffs and Market Structure: The German Zollverein as a Model for Economic Integration", in W. R. Lee, ed., *German Industry and German Industrialisation*, Routledge, London, pp. 77–115.

Fay, C. R. (1932), *The Corn Laws and Social England*, Cambridge University Press, Cambridge.

Federico, G. (2011), "When did European Markets Integrate?", *European Review of Economic History* **15**, 93–126.

Federico, G. and K. G. Persson (2007), "Market Integration and Convergence in the World Wheat Market, 1800–2000", in T. J. Hatton, K. H. O'Rourke and A. M. Taylor, eds., *The New Comparative Economic History: Essays in Honor of Jeffrey G. Williamson*, MIT Press, Cambridge, MA, pp. 87–114.

Findlay, R. and K. H. O'Rourke (2005), "Commodity Market Integration, 1500–2000", in M. D. Bordo, A. M. Taylor and J. G. Williamson, eds., *Globalization in Historical Perspective*, University of Chicago Press, Chicago, pp. 13–62.

Harley, K. (1980), "Transportation, the World Wheat Trade, and the Kuznets Cycle, 1850–1913", *Explorations in Economic History* **17**, 217–250.

Harley, K. (1988), "Ocean Freight Rates and Productivity, 1740–1913: The Primacy of Mechanical Invention Reaffirmed", *Journal of Economic History* **48**, 851–876.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Jacks, D. (2005), "Intra- and International Commodity Market Integration in the Atlantic Economy, 1800–1913", *Explorations in Economic History* **42**, 381–413.

Jacks, D., D. Novy and C. M. Meissner (2008), "Trade Costs, 1870–2000", *American Economic Review* **98**, 529–534.

Jacobs, A. and H. Richter (1935), "Die Getreidepreise in Deutschland seit dem Ausgang des 18. Jahrhunderts", *Vierteljahrshefte zur Statistik des Deutschen Reichs* **44**, 273–321.

Jörberg, L. (1972), *A History of Prices in Sweden, 1732–1914*, Gleerup, Lund.

Kaukiainen, Y. (2001), "Shrinking the World: Improvements in the Speed of Information Transmission", *European Review of Economic History* **5**, 1–28.

Keller, W. and C. Shiue (2008), "Institutions, Technology, and Trade", NBER Working Paper.

Komlos, J. (1979), "Austro-Hungarian Agricultural Development 1827–1877", *Journal of European Economic History* **8**, 37–60.

Kose, A., C. Otrok and C. H. Whiteman (2003), "International Business Cycles: World, Region and Country-Specific Factors", *American Economic Review* **93**, 1216–1239.

Lampe, M. (2009), "Effects of Bilateralism and the MFN Clause on International Trade: Evidence for the Cobden–Chevalier Network, 1860–1875", *The Journal of Economic History* **69**, 1012–1040.

Lehmann, S. and K. O'Rourke (2011), "The Structure of Protection and Growth in the Late Nineteenth Century", *Review of Economics and Statistics* **93**, 606–616.

Lévy-Leboyer, M. and F. Bourguignon (1990), *The French Economy in the Nineteenth Century: an Essay in Econometric Analysis*, Cambridge University Press, Cambridge.

North, D. C. (1958), "Ocean Freight Rates and Economic Development 1750-1913", *Journal of Economic History* **18**, 537–555.

North, D. C. (1968), "Sources of Productivity Change in Ocean Shipping, 1600–1850", *Journal of Political Economy* **75**, 953–970.

Oberschelp, R. (1986), *Beiträge zur niedersächsischen Preisgeschichte des 16. bis 19. Jahrhunderts*, August Lax, Hannover.

O'Rourke, K. H. (1997), "The European Grain Invasion, 1870–1913", *Journal of Economic History* **57**, 775–801.

O'Rourke, K. H., J. R. Rosés and J. G. Williamson (2007), "Globalization, Growth and Distribution in Spain 1500–1913", CEPR Discussion Paper No. 6356.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

O'Rourke, K. H. and J. G. Williamson (1999), *Globalization and History: The Evolution of a 19th Century Atlantic Economy*, MIT Press, Cambridge, MA.

O'Rourke, K. H. and J. G. Williamson (2002), "After Columbus: Explaining Europe's Overseas Trade Boom, 1500–1800", *Journal of Economic History* **62**, 417–456.

Pena, D. and N. S. Albornoz (1984), "Wheat Prices in Spain, 1857-1890: An Application of the Box-Jenkins Methodology", *Journal of European Economic History* **13**, 353–373.

Sánchez-Albornoz, N. (1974), "Congruence Among Spanish Economic Regions in the Nineteenth Century", *Journal of European Economic History* **3**, 725–746.

Sargent, T. J. and C. A. Sims (1977), "Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory", Federal Reserve Bank of Minneapolis Working Paper No. 55, 45–108.

Schonhardt-Bailey, C. (2006), *From the Corn Laws to Free Trade: Interests, Ideas, and Institutions in Historical Perspective*, MIT Press, Cambridge, MA.

Sharp, P. (2010), "1846 and All That': the Rise and Fall of British Wheat Protection in the Nineteenth Century", *Agricultural History Review* **58**, 76–94.

Spree, R. (1977), *Die Wachstumszyklen der deutschen Wirtschaft von 1840 bis 1880*, Duncker und Humblot, Berlin.

Uebele, M. (2009), "Historical Business Cycles and Market Integration: Evidence from Comovement", *PhD Thesis*, Humboldt-Universität zu Berlin.

Uebele, M. (2010), "Demand Matters: German Wheat Market Integration 1806-1855 in a European Context", CQE Working Papers 11/2010.

Appendix A

Table A1 Robustness for trend filter

		1806–1830			1831–1855			1856–1880			1881–1907		
		INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC
Full	HP	0.34	0.47	0.18	0.58	0.30	0.12	0.67	0.25	0.08	0.72	0.20	0.08
sample	BK[a]	0.08	0.71	0.20	0.50	0.38	0.12	0.64	0.27	0.09	0.73	0.20	0.07
avg.	FD	0.18	0.60	0.22	0.49	0.37	0.13	0.60	0.30	0.10	0.62	0.29	0.08
	CF[a]	0.38	0.42	0.20	0.48	0.36	0.16	0.75	0.18	0.06	0.62	0.29	0.08

[a]Band-pass window 2–8 years. 48 cities, 1806–1907.
Results may vary and not add up to one due to sampling error and rounding.
HP: Hodrick-Prescott, BK: Baxter-King, FD: First Differences, CF: Christiano-Fitzgerald.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table A2 Varying the number of markets per nation

Avg. over	Sample size	1806–1830			1831–1855			1856–1880			1881–1907		
		INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC
Full sample	26	0.37	0.44	0.19	0.58	0.30	0.12	0.69	0.23	0.08	0.73	0.19	0.08
	31	–	–	–	0.51	0.34	0.15	0.62	0.26	0.12	0.67	0.23	0.09
	48	0.35	0.48	0.17	0.54	0.34	0.11	0.72	0.20	0.08	0.69	0.20	0.10
	60	–	–	–	0.58	0.29	0.13	0.59	0.31	0.10	0.72	0.20	0.08
Austria–Hungary	26	0.16	0.38	0.45	0.64	0.09	0.26	0.72	0.20	0.08	0.69	0.20	0.10
	31	–	–	–	0.64	0.06	0.29	0.76	0.13	0.12	0.62	0.33	0.05
	48	0.15	0.41	0.61	0.61	0.09	0.31	0.75	0.13	0.12	0.57	0.37	0.06
	60	–	–	–	0.57	0.09	0.34	0.73	0.14	0.14	0.55	0.40	0.05
Belgium	26	0.72	0.23	0.04	0.88	0.10	0.02	0.71	0.27	0.01	0.94	0.04	0.01
	31	–	–	–	0.88	0.10	0.03	0.71	0.28	0.01	0.96	0.03	0.01
	48	0.56	0.40	0.05	0.90	0.08	0.03	0.65	0.33	0.01	0.90	0.09	0.02
	60	–	–	–	0.93	0.04	0.03	0.61	0.37	0.02	0.90	0.07	0.02
France	26	0.80	0.08	0.11	0.65	0.27	0.08	0.69	0.25	0.06	0.68	0.26	0.05
	31	–	–	–	0.66	0.27	0.08	0.70	0.24	0.06	0.70	0.25	0.05
	48	0.55	0.31	0.14	0.69	0.22	0.09	0.67	0.28	0.05	0.63	0.27	0.10
	60	–	–	–	0.73	0.19	0.09	0.61	0.34	0.05	0.66	0.24	0.10
Germany	26	0.31	0.47	0.23	0.82	0.03	0.15	0.89	0.04	0.07	0.86	0.07	0.07
	31	–	–	–	0.81	0.04	0.15	0.89	0.03	0.09	0.86	0.08	0.07
	48	0.21	0.57	0.22	0.82	0.04	0.15	0.86	0.04	0.11	0.81	0.12	0.08
	60	–	–	–	0.82	0.06	0.13	0.74	0.15	0.12	0.82	0.11	0.07
UK	26	0.19	0.76	0.05	0.62	0.36	0.03	0.88	0.11	0.02	0.88	0.09	0.03
	31	–	–	–	0.59	0.38	0.04	0.88	0.11	0.02	0.88	0.08	0.03
	48	0.48	0.44	0.07	0.64	0.32	0.04	0.91	0.08	0.02	0.90	0.06	0.04
	60	–	–	–	0.65	0.32	0.03	0.73	0.24	0.02	0.91	0.04	0.05
US	26	0.37	0.53	0.10	0.25	0.69	0.06	0.16	0.78	0.06	0.73	0.25	0.02
	31	–	–	–	0.22	0.68	0.10	0.20	0.61	0.19	0.74	0.24	0.02
	48	0.39	0.50	0.11	0.24	0.70	0.06	0.20	0.74	0.06	0.73	0.25	0.02
	60	–	–	–	0.26	0.64	0.10	0.23	0.61	0.16	0.74	0.24	0.02
Sweden	26	0.11	0.60	0.29	0.23	0.58	0.19	0.64	0.18	0.18	0.47	0.25	0.28
	31	–	–	–	0.20	0.60	0.21	0.65	0.17	0.18	0.47	0.25	0.28
	48	0.06	0.68	0.26	0.15	0.67	0.18	0.67	0.14	0.19	0.47	0.27	0.26
	60	–	–	–	0.12	0.70	0.18	0.51	0.29	0.19	0.46	0.28	0.26
Spain	26	–	–	–	–	–	–	–	–	–	–	–	–
	31	–	–	–	0.16	0.52	0.32	0.23	0.54	0.23	0.23	0.56	0.22
	48	–	–	–	–	–	–	–	–	–	–	–	–
	60	–	–	–	0.17	0.46	0.38	0.07	0.71	0.22	0.24	0.41	0.36

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table A3 Results (single cities)

		1806–1830			1831–1855			1856–1880			1881–1907		
		INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC
A.-H.	Vienna	0.12	0.76	0.12	0.72	0.17	0.09	0.64	0.27	0.10	0.64	0.36	0.01
		(0.03)	(0.11)	(0.10)	(0.05)	(0.09)	(0.07)	(0.04)	(0.07)	(0.06)	(0.02)	(0.02)	(0.00)
	Lwow	0.24	0.01	0.75	0.36	0.03	0.60	0.79	0.00	0.21	0.49	0.34	0.17
		(0.05)	(0.03)	(0.05)	(0.04)	(0.07)	(0.07)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
	Ljubljana	0.21	0.62	0.16	0.59	0.17	0.21	0.75	0.17	0.08	0.56	0.40	0.04
		(0.05)	(0.10)	(0.09)	(0.05)	(0.09)	(0.08)	(0.04)	(0.05)	(0.04)	(0.02)	(0.02)	(0.01)
	Krakow	0.02	0.25	0.73	0.53	0.09	0.37	0.88	0.04	0.08	0.52	0.45	0.03
		(0.03)	(0.09)	(0.09)	(0.04)	(0.08)	(0.07)	(0.06)	(0.05)	(0.02)	(0.02)	(0.02)	(0.01)
	Budapest	–	–	–	–	–	–	–	–	–	0.65	0.34	0.01
		–	–	–	–	–	–	–	–	–	(0.02)	(0.02)	(0.01)
Belgium	Brugges	0.50	0.44	0.05	0.93	0.05	0.03	0.66	0.32	0.02	0.90	0.06	0.04
		(0.15)	(0.15)	(0.02)	(0.04)	(0.04)	(0.01)	(0.06)	(0.06)	(0.01)	(0.01)	(0.01)	(0.01)
	Ghent	0.59	0.39	0.02	0.93	0.04	0.03	0.66	0.33	0.01	0.89	0.10	0.01
		(0.15)	(0.15)	(0.01)	(0.04)	(0.04)	(0.01)	(0.06)	(0.06)	(0.01)	(0.01)	(0.02)	(0.01)
	Brussels	0.58	0.36	0.06	0.95	0.03	0.02	0.69	0.30	0.01	0.92	0.08	0.00
		(0.16)	(0.16)	(0.02)	(0.03)	(0.03)	(0.01)	(0.06)	(0.06)	(0.01)	(0.01)	(0.01)	(0.00)
France	Bayeux	0.66	0.18	0.16	0.78	0.12	0.09	0.75	0.22	0.04	0.71	0.21	0.08
		(0.19)	(0.19)	(0.03)	(0.05)	(0.05)	(0.01)	(0.05)	(0.05)	(0.00)	(0.01)	(0.02)	(0.01)
	Saint-Brieuc	0.58	0.22	0.20	0.84	0.10	0.06	0.72	0.24	0.03	0.43	0.25	0.31
		(0.18)	(0.18)	(0.03)	(0.04)	(0.04)	(0.01)	(0.05)	(0.05)	(0.00)	(0.02)	(0.03)	(0.02)
	Toulouse	0.38	0.48	0.13	0.48	0.46	0.07	0.63	0.36	0.02	0.77	0.17	0.05
		(0.16)	(0.17)	(0.05)	(0.05)	(0.05)	(0.02)	(0.05)	(0.05)	(0.00)	(0.01)	(0.02)	(0.01)
	Bordeaux	0.40	0.52	0.08	0.76	0.21	0.03	0.71	0.28	0.01	0.72	0.23	0.05
		(0.17)	(0.18)	(0.03)	(0.05)	(0.05)	(0.01)	(0.05)	(0.05)	(0.00)	(0.02)	(0.02)	(0.01)

(continued)

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table A3 Continued

		1806–1830			1831–1855			1856–1880			1881–1907		
		INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC
	Chateauroux	0.44	0.44	0.12	0.79	0.15	0.06	0.65	0.33	0.02	0.77	0.20	0.03
		(0.18)	(0.19)	(0.03)	(0.05)	(0.05)	(0.01)	(0.05)	(0.05)	(0.00)	(0.02)	(0.02)	(0.01)
	Mende	0.68	0.24	0.08	0.40	0.36	0.23	0.55	0.38	0.07	0.64	0.17	0.19
		(0.17)	(0.17)	(0.02)	(0.04)	(0.05)	(0.03)	(0.05)	(0.05)	(0.01)	(0.01)	(0.02)	(0.02)
	Barleduc	0.67	0.23	0.10	0.88	0.02	0.10	0.68	0.28	0.05	0.68	0.21	0.11
		(0.17)	(0.17)	(0.02)	(0.04)	(0.03)	(0.02)	(0.05)	(0.05)	(0.01)	(0.02)	(0.03)	(0.02)
	Arras	0.69	0.23	0.08	0.93	0.02	0.05	0.75	0.23	0.02	0.76	0.17	0.07
		(0.19)	(0.19)	(0.02)	(0.04)	(0.03)	(0.01)	(0.05)	(0.05)	(0.00)	(0.02)	(0.02)	(0.01)
	Pau	0.42	0.24	0.35	0.47	0.46	0.07	0.62	0.37	0.01	0.76	0.22	0.02
		(0.10)	(0.11)	(0.04)	(0.05)	(0.05)	(0.02)	(0.05)	(0.05)	(0.00)	(0.01)	(0.01)	(0.01)
	Lyon	0.61	0.33	0.07	0.83	0.06	0.11	0.70	0.27	0.03	0.77	0.20	0.03
		(0.17)	(0.17)	(0.02)	(0.04)	(0.04)	(0.01)	(0.05)	(0.05)	(0.00)	(0.01)	(0.01)	(0.01)
	Marseille	–	–	–	–	–	–	0.61	0.16	0.23	0.76	0.04	0.20
		–	–	–	–	–	–	(0.05)	(0.05)	(0.01)	(0.02)	(0.01)	(0.01)
	Paris	0.55	0.32	0.12	0.91	0.02	0.06	0.65	0.17	0.18	0.79	0.16	0.05
		(0.18)	(0.19)	(0.03)	(0.04)	(0.03)	(0.01)	(0.03)	(0.04)	(0.01)	(0.02)	(0.02)	(0.01)
Germany	Berlin	0.16	0.75	0.08	0.91	0.06	0.03	0.92	0.02	0.06	0.84	0.13	0.04
		(0.08)	(0.09)	(0.05)	(0.02)	(0.02)	(0.01)	(0.05)	(0.05)	(0.02)	(0.01)	(0.02)	(0.02)
	Koenigsberg	0.00	0.67	0.32	0.77	0.14	0.09	0.95	0.01	0.03	0.83	0.12	0.05
		(0.02)	(0.06)	(0.06)	(0.03)	(0.05)	(0.04)	(0.06)	(0.06)	(0.01)	(0.01)	(0.02)	(0.02)
	Munich	0.51	0.15	0.34	0.71	0.02	0.27	0.70	0.01	0.28	0.73	0.09	0.18
		(0.10)	(0.09)	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)	(0.07)	(0.06)	(0.02)	(0.03)	(0.03)
	Hamburg	0.18	0.72	0.09	0.85	0.01	0.13	0.98	0.00	0.01	0.88	0.08	0.04
		(0.07)	(0.09)	(0.05)	(0.04)	(0.04)	(0.02)	(0.05)	(0.05)	(0.01)	(0.01)	(0.02)	(0.01)

Logged and Hodrick-Prescott (6.25) filtered prices.
Median values of Bayesian parameter distributions. Standard deviations in brackets.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table A3 Results (single cities, continued)

		1806–1830			1831–1855			1856–1880			1881–1907		
		INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC
UK	London	0.53	0.46	0.01	0.67	0.32	0.01	0.86	0.13	0.01	0.92	0.07	0.01
		(0.18)	(0.18)	(0.01)	(0.03)	(0.03)	(0.00)	(0.04)	(0.04)	(0.00)	(0.02)	(0.02)	(0.00)
	Dover	0.61	0.36	0.03	0.69	0.29	0.02	0.88	0.11	0.01	0.92	0.05	0.03
		(0.19)	(0.19)	(0.01)	(0.03)	(0.03)	(0.00)	(0.04)	(0.04)	(0.00)	(0.02)	(0.02)	(0.00)
	Exeter	0.76	0.15	0.09	0.69	0.27	0.04	0.94	0.05	0.01	0.93	0.01	0.05
		(0.20)	(0.19)	(0.02)	(0.04)	(0.04)	(0.00)	(0.04)	(0.04)	(0.00)	(0.01)	(0.01)	(0.00)
	Gloucester	0.60	0.32	0.07	0.62	0.36	0.02	0.94	0.05	0.01	0.98	0.01	0.01
		(0.20)	(0.20)	(0.02)	(0.03)	(0.03)	(0.00)	(0.04)	(0.04)	(0.00)	(0.01)	(0.01)	(0.00)
	Worcester	0.53	0.41	0.06	0.61	0.36	0.03	0.92	0.07	0.01	0.96	0.03	0.01
		(0.19)	(0.19)	(0.01)	(0.04)	(0.04)	(0.00)	(0.04)	(0.04)	(0.00)	(0.01)	(0.01)	(0.00)
	Cambridge	0.48	0.47	0.05	0.66	0.33	0.01	0.90	0.09	0.01	0.95	0.04	0.01
		(0.17)	(0.17)	(0.01)	(0.03)	(0.03)	(0.00)	(0.04)	(0.04)	(0.00)	(0.02)	(0.01)	(0.00)
	Norwich	0.54	0.41	0.05	0.69	0.30	0.00	0.89	0.10	0.01	0.95	0.02	0.03
		(0.18)	(0.18)	(0.01)	(0.03)	(0.03)	(0.00)	(0.04)	(0.04)	(0.00)	(0.02)	(0.01)	(0.00)
	Leeds	0.31	0.65	0.03	0.63	0.36	0.00	0.91	0.07	0.01	0.92	0.06	0.02
		(0.13)	(0.13)	(0.02)	(0.03)	(0.03)	(0.00)	(0.04)	(0.04)	(0.00)	(0.02)	(0.02)	(0.00)
	Liverpool	0.34	0.58	0.08	0.54	0.39	0.07	0.90	0.08	0.02	0.96	0.04	0.00
		(0.15)	(0.15)	(0.01)	(0.04)	(0.04)	(0.00)	(0.05)	(0.05)	(0.00)	(0.01)	(0.01)	(0.00)
	Manchester	0.29	0.66	0.04	0.61	0.38	0.01	0.88	0.09	0.03	0.84	0.09	0.07
		(0.13)	(0.13)	(0.01)	(0.03)	(0.03)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.01)
	Newcastle	0.21	0.64	0.14	0.60	0.36	0.04	0.89	0.05	0.07	0.74	0.09	0.17
		(0.09)	(0.09)	(0.03)	(0.04)	(0.04)	(0.00)	(0.05)	(0.05)	(0.01)	(0.02)	(0.03)	(0.01)

(continued)

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table A3 Continued

		1806–1830			1831–1855			1856–1880			1881–1907		
		INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC
	Carmarthen	0.60	0.21	0.19	0.67	0.22	0.11	0.85	0.13	0.02	0.95	0.01	0.04
		(0.18)	(0.17)	(0.02)	(0.03)	(0.03)	(0.00)	(0.04)	(0.04)	(0.00)	(0.01)	(0.01)	(0.01)
USA	New York	0.28	0.54	0.18	0.31	0.58	0.11	0.47	0.50	0.03	0.78	0.20	0.02
		(0.07)	(0.11)	(0.09)	(0.03)	(0.03)	(0.02)	(0.06)	(0.06)	(0.01)	(0.01)	(0.01)	(0.00)
	Philadelphia	0.38	0.55	0.07	0.28	0.69	0.03	0.49	0.47	0.04	0.79	0.20	0.01
		(0.06)	(0.10)	(0.08)	(0.03)	(0.03)	(0.02)	(0.05)	(0.05)	(0.01)	(0.02)	(0.02)	(0.00)
	Alexandria	0.52	0.40	0.07	0.13	0.64	0.24	0.41	0.47	0.12	0.76	0.21	0.03
		(0.07)	(0.09)	(0.06)	(0.02)	(0.03)	(0.03)	(0.05)	(0.05)	(0.01)	(0.02)	(0.02)	(0.00)
	Cincinnati	–	–	–	0.25	0.72	0.03	0.32	0.61	0.06	0.65	0.33	0.02
		–	–	–	(0.03)	(0.03)	(0.02)	(0.06)	(0.06)	(0.01)	(0.02)	(0.02)	(0.01)
	Ithaca	–	–	–	–	–	–	0.45	0.52	0.03	0.65	0.29	0.06
		–	–	–	–	–	–	(0.06)	(0.06)	(0.01)	(0.02)	(0.02)	(0.01)
	Chicago	–	–	–	–	–	–	0.49	0.38	0.13	0.76	0.17	0.07
		–	–	–	–	–	–	(0.04)	(0.04)	(0.01)	(0.02)	(0.02)	(0.01)
	Indianapolis	–	–	–	–	–	–	0.43	0.54	0.03	0.68	0.29	0.02
		–	–	–	–	–	–	(0.06)	(0.06)	(0.01)	(0.02)	(0.02)	(0.01)
	San Francisco	–	–	–	–	–	–	0.02	0.04	0.94	0.70	0.01	0.30
		–	–	–	–	–	–	(0.01)	(0.01)	(0.02)	(0.01)	(0.00)	(0.01)

Logged and Hodrick-Prescott (6.25) filtered prices.
Median values of Bayesian parameter distributions. Standard deviations in brackets.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table A3 Results (single cities, continued)

		1806–1830			1831–1855			1856–1880			1881–1907		
		INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC
Sweden	Stockholm	0.06	0.65	0.29	0.11	0.85	0.04	0.77	0.12	0.11	0.57	0.35	0.08
		(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.01)	(0.04)	(0.04)	(0.01)	(0.02)	(0.03)	(0.02)
	Uppsala	0.18	0.47	0.34	0.18	0.78	0.04	0.32	0.21	0.47	0.35	0.37	0.28
		(0.05)	(0.06)	(0.04)	(0.03)	(0.03)	(0.01)	(0.02)	(0.05)	(0.03)	(0.01)	(0.03)	(0.03)
	Soedermanland	0.06	0.69	0.25	0.13	0.82	0.05	0.81	0.17	0.02	0.32	0.25	0.43
		(0.02)	(0.06)	(0.05)	(0.03)	(0.03)	(0.01)	(0.03)	(0.03)	(0.01)	(0.02)	(0.04)	(0.03)
	Oestergoetland	0.08	0.74	0.18	0.14	0.78	0.09	0.87	0.10	0.03	0.56	0.39	0.06
		(0.03)	(0.05)	(0.03)	(0.03)	(0.03)	(0.01)	(0.04)	(0.04)	(0.01)	(0.02)	(0.03)	(0.02)
	Kalmar	0.00	0.84	0.14	0.11	0.78	0.11	0.84	0.13	0.04	0.37	0.28	0.35
		(0.06)	(0.05)	(0.04)	(0.02)	(0.03)	(0.01)	(0.03)	(0.03)	(0.01)	(0.02)	(0.03)	(0.02)
	Halland	0.01	0.72	0.26	0.08	0.31	0.61	0.79	0.10	0.11	0.43	0.29	0.28
		(0.06)	(0.06)	(0.05)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.01)	(0.02)	(0.03)	(0.03)
	Skaraborg	0.01	0.81	0.17	0.12	0.66	0.21	0.78	0.13	0.10	0.58	0.23	0.19
		(0.03)	(0.04)	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)	(0.04)	(0.01)	(0.02)	(0.03)	(0.02)
	Oerebro	0.05	0.76	0.19	0.07	0.72	0.20	0.83	0.11	0.06	0.41	0.36	0.23
		(0.02)	(0.04)	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)	(0.04)	(0.01)	(0.02)	(0.04)	(0.03)
	Vaestmanland	0.14	0.69	0.17	0.06	0.86	0.08	0.89	0.09	0.02	0.49	0.42	0.08
		(0.04)	(0.06)	(0.03)	(0.02)	(0.03)	(0.01)	(0.04)	(0.04)	(0.01)	(0.02)	(0.03)	(0.02)
	Gaestrikland	0.05	0.62	0.33	0.12	0.78	0.10	0.66	0.00	0.34	0.46	0.12	0.42
		(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.01)	(0.05)	(0.04)	(0.02)	(0.02)	(0.03)	(0.02)
	Haelsingland	0.01	0.54	0.43	0.04	0.76	0.19	0.22	0.01	0.77	0.46	0.13	0.41
		(0.07)	(0.06)	(0.06)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.02)
Spain	Burgos	–	–	–	0.14	0.44	0.43	0.16	0.80	0.03	0.34	0.45	0.21
		–	–	–	(0.03)	(0.07)	(0.06)	(0.02)	(0.03)	(0.01)	(0.02)	(0.03)	(0.03)

(continued)

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013

Table A3 Continued

	1806–1830			1831–1855			1856–1880			1881–1907		
	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC	INT	NAT	LOC
Cordoba	–	–	–	0.05	0.57	0.39	0.01	0.53	0.46	0.09	0.82	0.09
	–	–	–	(0.01)	(0.07)	(0.07)	(0.00)	(0.03)	(0.03)	(0.01)	(0.03)	(0.03)
Gerona	–	–	–	0.41	0.36	0.23	0.41	0.24	0.35	0.57	0.25	0.19
	–	–	–	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)
Granada	–	–	–	0.10	0.50	0.41	0.03	0.60	0.37	0.03	0.78	0.19
	–	–	–	(0.01)	(0.06)	(0.06)	(0.01)	(0.03)	(0.03)	(0.01)	(0.04)	(0.04)
Lerida	–	–	–	0.21	0.69	0.10	0.08	0.58	0.34	0.37	0.46	0.17
	–	–	–	(0.02)	(0.05)	(0.05)	(0.01)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)
Oviedo	–	–	–	0.11	0.43	0.46	0.37	0.46	0.17	0.04	0.06	0.90
	–	–	–	(0.02)	(0.05)	(0.05)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)
Segovia	–	–	–	0.07	0.39	0.55	0.16	0.72	0.12	0.46	0.40	0.14
	–	–	–	(0.02)	(0.07)	(0.06)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)
Zaragoza	–	–	–	0.11	0.70	0.20	0.08	0.72	0.20	0.46	0.25	0.29
	–	–	–	(0.02)	(0.05)	(0.04)	(0.01)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)
Santander	–	–	–	0.31	0.22	0.48	0.22	0.55	0.24	0.08	0.30	0.62
	–	–	–	(0.04)	(0.05)	(0.05)	(0.03)	(0.03)	(0.02)	(0.01)	(0.04)	(0.04)
Leon	–	–	–	–	–	–	0.22	0.72	0.06	0.21	0.51	0.28
	–	–	–	–	–	–	(0.03)	(0.03)	(0.01)	(0.01)	(0.03)	(0.03)
Toledo	–	–	–	–	–	–	0.03	0.75	0.23	0.12	0.74	0.14
	–	–	–	–	–	–	(0.01)	(0.03)	(0.03)	(0.01)	(0.04)	(0.03)
Coruna	–	–	–	–	–	–	0.39	0.49	0.12	0.06	0.28	0.65
	–	–	–	–	–	–	(0.04)	(0.04)	(0.01)	(0.01)	(0.03)	(0.04)

Logged and Hodrick-Prescott (6.25) filtered prices.
Median values of Bayesian parameter distributions. Standard deviations in brackets.

Downloaded from http://cesifo.oxfordjournals.org/ by guest on September 12, 2013