ISSN 1610-241X (PRINT)
ISSN 1612-7501 (ONLINE)

CESifo Economic Studies

Volume 63 Number 3 September 2017

http://academic.oup.com/cesifo

A joint initiative of the University of Munich's Center for Economic Studies and the Ifo Institute

OXFORD UNIVERSITY PRESS

CESifo Economic Studies

With the increasing specialisation in the profession, there is an urgent need for building bridges between different areas in economics. The economic research papers published in academic journals are becoming increasingly less accessible to those not working in the same field. *CESifo Economic Studies* bridges the gap between specialised research in economics and the interest of non-specialists. The journal combines theory and empirical research with a strong emphasis on policy issues. *CESifo Economic Studies* publishes contributed papers by top-level academic economists which are written in a style accessible to all economists regardless of their specialisation. The primary focus of the journal is on policy problems, but it also publishes high-quality, non-technical articles which are of general interest to economists.

Abstracted and Indexed in Social Sciences Citation Index, Social Scisearch, Current Contents/Social and Behavioural Sciences, EconLit, Journal of Economic Literature, Scopus and Geobase.

SUBSCRIPTIONS

A subscription to *CESifo Economic Studies* comprises 4 issues. Prices include postage by surface mail, or for subscribers in the USA and Canada by airfreight, or in India, Japan, Australia and New Zealand, by Air Speeded Post. Airmail rates are available on request.

Annual Subscription Rate (Volume 63, 4 issues, 2017)
Institutional
Print edition and site-wide online access: £293/€436/US$562
Print edition only: £270/€402/US$518
Site-wide online access only: £230/€342/US$441
Personal
Print edition and individual online access: £66/€98/US$126
Please note: US$ rate applies to US & Canada, Euros applies to Europe, UK£ applies to UK and Rest of World

There are other subscription rates available, for a complete listing please visit http://academic.oup.com/cesifo/subscriptions.

Full prepayment, in the correct currency, is required for all orders. Orders are regarded as firm and payments are not refundable. Subscriptions are accepted and entered on a complete volume basis. Claims cannot be considered more than FOUR months after publication or date of order, whichever is later. All subscriptions in Canada are subject to GST. Subscriptions in the EU may be subject to European VAT. If registered, please supply details to avoid unnecessary charges. For subscriptions that include online versions, a proportion of the subscription price may be subject to UK VAT. Personal rate subscriptions are only available if payment is made by personal cheque or credit card and delivery is to a private address. For further details, see inside back cover.

The current year and two previous years' issues are available from Oxford Journals. Previous volumes can be obtained from the Periodicals Service Company at http://www.periodicals.com/oxford.html or Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. E-mail: psc@periodicals.com. Tel: (518) 537-4700. Fax: (518) 537-5899.

CESifo Economic Studies (ISSN 1610-241X) is published four times per year in March, June, September and December by Oxford University Press, Oxford, UK and distributed in the USA by Central Mailing Services c/o UKP Worldwide, 1637 Stelton Road B2, Piscataway, NJ 08854.

The US annual print subscription price is $562.00. Airfreight and mailing in the USA by agent named Central Mailing Services c/o UKP Worldwide, 1637 Stelton Road B1-2, Piscataway, NJ 08854. Periodicals postage paid at Piscataway, NJ and additional mailing offices.

US Postmaster: send address changes to Central Mailing Services c/o UKP Worldwide, 1637 Stelton Road B1-2, Piscataway, NJ 08854.

Oxford Journals Environmental and Ethical Policies
Oxford Journals is committed to working with the global community to bring the highest quality research to the widest possible audience. Oxford Journals will protect the environment by implementing environmentally friendly policies and practices wherever possible. Please see https://academic.oup.com/journals/pages/about_us/ethical_policies for further information on Oxford Journals' environmental and ethical policies.

CESifo Economic Studies

SUBSCRIPTIONS continued

For further information, please contact: Journals Customer Service Department, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. Email: jnls.cust.serv@oup.com. Tel (and answerphone outside normal working hours): +44 (0)1865 353907. Fax: + 44 (0)1865 353485. **In the US, please contact:** Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513, USA. Email: jnlorders@oup.com. Tel (and answerphone outside normal working hours): 800 852 7323 (toll-free in USA/Canada). Fax: 919 677 1714. **In Japan, please contact:** Journals Customer Service, Oxford University Press, 4-5-10-8F Shiba, Minato-ku, Tokyo 108-8386, Japan. Tel. +81 3 5444 5858. Fax. +81 3 3454 2929. E-mail: custserv.jp@oup.com.

Methods of payment. (i) Cheque (payable to Oxford University Press, to Oxford University Press, Cashiers Office, Great Clarendon Street, Oxford OX2 6DP, UK) in GB£ Sterling (drawn on a UK bank), US$ Dollars (drawn on a US bank), or EU€ Euros. (ii) Bank transfer to Barclays Bank Plc, Oxford Group Office, Oxford (bank sort code 20-65-18) (UK), overseas only Swift code BARC GB 22 (GB£ Sterling to account no. 70299332, IBAN GB89BARC20651870299332; US$ Dollars to account no. 66014600, IBAN GB27BARC20651866014600; EU€ Euros to account no. 78923655, IBAN GB16BARC20651878923655). (iii) Credit card (Mastercard, Visa, Switch or American Express).

PERMISSIONS

For information on how to request permissions to reproduce articles/information from this journal, please visit https://academic.oup.com/journals/pages/access_purchase/rights_and_permissions.

ADVERTISING

Inquiries about advertising should be sent to CESifo Economic Studies, Oxford Journals Advertising, Journals Division, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. Email: jnlsadvertising@oup.com. Tel: +44 (0)1865 353329, (0)1865 354637 or (0)1865 354767. Fax: +44 (0)1865 353774.

DISCLAIMER

INSTRUCTIONS TO AUTHORS

CESifo Economic Studies is an economics journal circulated to a wide audience including governments, businesses, and policy-makers as well as academics and students. Authors should make every possible effort to avoid technicalities in the form of either terminology which will be unfamiliar to a general readership or mathematical presentations. Where specialist terminology is used, it should be defined clearly and examples of its application given. The use of mathematics should be avoided. Authors who do employ mathematics in their articles should ensure that it is a necessary component of their presentation which clarifies an unfamiliar concept. If mathematics is used, authors should restrict it to an appendix as much as possible.

Format. On the first manuscript page provide the title, names of all authors, affiliations, any footnotes to the title, address to which proofs are to be sent, a short running title, and both fax number and e-mail address of the author to be contacted.

Abstracts. Authors are asked to provide a short abstract in English, which should not exceed 100 words. Please supply up to five keywords, and the Journal of Economic Literature index number.

Submitting your paper. Once you have prepared your manuscript according to the instructions above please visit the online submission web site: http://mc.manuscriptcentral.com/cesifo

Refereeing. All articles submitted to the journal will be refereed.

For detailed Instructions to Authors, please visit the journal's website: http://academic.oup.com/cesifo

MIX
Paper from responsible sources
FSC® C007785
www.fsc.org

Typeset by Cenveo publisher services, Bangalore, India
Printed by Bell and Bain Ltd, Glasgow, UK

OXFORD
UNIVERSITY PRESS

About CESifo Economic Studies

CESifo Economic Studies aims to bridge the gap between specialised research in economics and the interest of non-specialists. It publishes policy oriented articles written in a style accessible to all economists regardless of their specialisation

CESifo Economic Studies

Volume 63 Number 3 September 2017

http://academic.oup.com/cesifo

Contents

CESifo Economic Studies, 2017, 255–269
doi: 10.1093/cesifo/ifx012
Advance Access Publication Date: 8 September 2017
Original article

Vocational vs. General Education and Employment over the Life Cycle: New Evidence from PIAAC

Franziska Hampf* and Ludger Woessmann†

*ifo Institute, University of Munich, Munich, Bavaria, Germany. e-mail: hampf@ifo.de and †University of Munich, Munich, Bavaria, Germany, ifo Institute, CESifo, Munich, Germany, IZA, Bonn, Germany. e-mail: woessmann@ifo.de

Abstract

It has been argued that vocational education facilitates the school-to-work transition but reduces later adaptability to changing environments. Using the recent international data of the Programme for the International Assessment of Adult Competencies (PIAAC), we confirm such a trade-off over the life cycle in a difference-in-differences model that compares employment rates across education type and age. An initial employment advantage of individuals with vocational compared to general education turns into a disadvantage later in life. Results are strongest in apprenticeship countries that provide the highest intensity of industry-based vocational education. (JEL codes: J24, J64, I20)

Key words: vocational education, apprenticeship, employment, life cycle, PIAAC

1. Introduction

Around the world, there is an increasing interest in expanded vocational education as a way to get youth quickly and effectively into the labor market by endowing them with occupation-specific skills. Earlier analysis of labor markets in the 1990s, however, suggested possible adverse impacts of vocational education on employment opportunities later in life due to limited adaptability to technological and structural change (Hanushek et al. 2017). With the significant transformation of labor markets over the past two decades including such factors as globalization, technological change, altered training programs, and reforms of social security systems, it is important to revisit the potential efficacy of expanding vocational education in today's economic environment. This article provides new evidence whether the employment trade-off of vocational orientation over the life cycle is still relevant today.

The ramifications of the deep changes that have occurred on labor markets for the employment effects of vocational education over the life cycle are not obvious. On the one hand, the structural changes brought about by globalization and the rapid technological changes stemming from automation and digitalization (Autor, Dorn, and Hanson 2015) may make the obsolescence of occupation-specific skills over the life cycle even more pronounced (cf. Krueger and Kumar 2004). In these changing environments, long-run employment prospects may be enhanced by general skills such as basic cognitive skills, social interaction skills, and skills that facilitate continuous learning such as transversal skills, adaptability, creativity, problem-solving, and critical thinking skills. On the other hand, reduced options of generous early retirement schemes in the social security systems of many countries may dampen the incidence of reduced employment at older ages, thereby reducing the scope for differential employment patterns between vocational and general education late in the life cycle.[1]

This article uses the Programme for the International Assessment of Adult Competencies (PIAAC), conducted in 2011–2012, to estimate the employment effects of vocational vs. general education over the life cycle on modern labor markets in a sample of 16 countries. To address concerns of selection into types of education, we employ the difference-in-differences model introduced by Hanushek et al. (2017) that compares employment rates across age for people with general and vocational education. We make use of the individual skill measures available in PIAAC, among others, to account for potential differential changes in selectivity over time.

Our results confirm a strong trade-off between early advantages and late disadvantages in employment for individuals with vocational education. But there is strong heterogeneity depending on the specific institutional structure of schooling and work-based training in a country. While no significant pattern is detected in the six countries without sizeable vocational systems, the declining relative age–employment pattern of individuals with vocational education is found across the 10 countries with significant vocational systems, and it is strongest in countries with widely developed apprenticeship systems where industry is directly involved in education. In these apprenticeship countries, the cross-over age by which individuals with a general education have higher employment probabilities is as low as age 44 years, and somewhat higher around age 50 years for the group of vocational countries at large.

Our study contributes to a growing literature on the effects of vocational education on labor-market outcomes over the life cycle. An extensive literature looks at the effect of vocational education on the school-to-work transition, with varying results (see Shavit and Müller (1998), Ryan (2001), and Zimmermann et al. (2013) for studies with an international focus and Malamud and Pop-Eleches (2010) for a study identified from a Romanian reform). Using the International Adult Literacy Survey (IALS) of the mid-1990 s, Hanushek et al. (2017) extended this perspective beyond the entry phase of the labor market, showing that the relative labor-market advantage of vocational over general education decreases with age. Several recent country-specific studies that go beyond the entry phase similarly show consistent age patterns by education type, including Cörvers et al. (2011) for Germany, the Netherlands, and Great Britain; Weber (2014) for Switzerland; and Brunello and Rocco (2017) for Great Britain. While Stenberg and Westerlund (2015) and Golsteyn

1 For example, in Germany the entitlement age for early retirement after 12 months of unemployment has been gradually raised from 60 to 63 years since 2006, and the terms of early retirement have become less generous. As a consequence, the share of those retiring before age 65 years (61 years) among all retirees has declined from 75% (56%) in 1995 to 57% (25%) in 2012 (Deutsche Rentenversicherung Bund 2015).

and Stenberg (2017) also find such a pattern for Sweden, Hall (2016) is an exception that does not find a significant pattern based on the pilot of a Swedish reform in 1988–1993 that extended upper-secondary vocational programs by 1 year and increased their general content.

Our results extend the life-cycle analysis to a large sample of countries with recent data. While some have argued that pension reforms that limit early retirement may have dampened any relative employment effect at older ages, others have suggested that increasing globalization, automation, and digitization may have made adaptability to changing occupational structures ever more important. In fact, our results show a continuing trade-off for vocational education between ease of labor-market entry and limited adaptability at later ages that is very similar in size to the results in Hanushek et al. (2017) for the mid-1990 s. Apart from the updated period, the PIAAC data also provide a much richer testing of skills and a sample size that is almost twice as large as in IALS. Because we see our main contribution in showing that international results which refer to two decades ago also hold on today's labor markets, we keep the article intentionally short with a focus on the core results of employment over the life cycle.[2]

Our analysis also extends the emerging literature that uses the PIAAC data to study different aspects of education and the labor market. Thus, Levels, van der Velden, and Allen (2014) provide an analysis on mismatch; Hanushek et al. (2015) on returns to skills; Brunello and Rocco (2015) and Forster, Bol, and van de Werfhorst (2016) on aspects of vocational education; Broecke, Quintini, and Vandeweyer (2017) on inequality; Falck, Heimisch, and Wiederhold (2016) on returns to information and communication technology (ICT) skills; and Kahn (2016) on employment protection.

In what follows, Section 2 introduces the PIAAC database. Section 3 describes the difference-in-differences model. Section 4 presents our main results on the employment effects of education type over the life cycle and reports several robustness analyses indicating that results are not driven by varying selectivity into education types over time. Section 5 tests for heterogeneity across groups of countries with differing vocational systems. Section 6 concludes.

2. The PIAAC Data

Collected between August 2011 and March 2012, PIAAC was developed by the Organisation for Economic Co-operation and Development (OECD) to survey the skills of a representative sample of adults aged 16–65 years in each participating country. For our purposes, PIAAC provides internationally comparable data on individuals' type of education, labor-market status, and background variables in 16 countries.[3]

We classify the 16 countries into different categories according to the extent and intensity of vocationalization of their education systems using information from PIAAC and

2 Hanushek et al. (2017) provide additional analyses of income and adult education over the life cycle, lifetime earnings, within-occupational-group analysis using the German Microcensus, and analysis of exogenous variation from plant closures in Austrian administrative data.

3 Among the remaining eight PIAAC countries, the Russian data have issues of representativeness; Canada and Estonia do not provide data on educational attainment in the Public Use File; and Belgium, Cyprus, Italy, Poland, and the Slovak Republic do not provide consistent data on the type of education.

OECD's Education at a Glance (EAG) statistics.[4] We define 'vocational countries' as those countries whose vocational share is at least 40% in PIAAC and at least 50% in EAG. Based on these criteria, 6 countries (Ireland, Japan, Korea, Spain, the United Kingdom, and the USA) are classified as 'non-vocational countries' with limited vocational systems, whereas 10 countries are 'vocational countries' with significant vocational systems. Among the latter, three countries (Austria, Denmark, and Germany) are 'apprenticeship countries' with a share of combined school and work-based vocational programs that exceeds 40% in EAG. Together with these three countries, the Czech Republic is also classified among the 'non-school based vocational countries' that have a vocational sector with at least 25% in combined school and work-based programs. The remaining six vocational countries (Australia, Finland, France, the Netherlands, Norway, and Sweden) have mostly school-based vocational sectors.

Our sample includes all males aged 16–65 years who completed at least secondary education and are not currently in education.[5] The type of education is derived from responses to an internationally harmonized background questionnaire. For individuals with secondary education, the PIAAC data provide a variable indicating whether a respondent's highest level of education is vocationally oriented. For individuals with tertiary education, we follow Hanushek et al. (2017) and Brunello and Rocco (2015) in classifying the largely theory-based tertiary-type A programs (ISCED 5A) that are designed to provide sufficient qualifications for entry to advanced research programs and professions with high skill requirements as general. The more practical, technical, and occupational specific tertiary-type B programs (ISCED 5B) that lead to professional qualifications are classified as vocational.[6]

Apart from the education type, PIAAC provides detailed tests of individuals' cognitive skills in numeracy, literacy, and 'problem solving in technology-rich environments'. These skill measures have been shown to have substantial returns on the labor market (Hanushek et al. 2015) and allow us to account for differential selectivity into education type by age. Test scores are normalized to have mean zero and standard deviation one within each country. Apart from the richer testing of skills, PIAAC also provides substantially larger sample sizes per country than the IALS data set of the mid-1990 s, so that our full sample of 29,452 individuals is almost twice as large as in the IALS study by Hanushek et al. (2017).

Table 1 provides descriptive statistics of the main variables of our analysis for the sample of 10 countries with significant vocational systems. On average, 64% of individuals have completed a vocational education program in this country sample. Country-specific inspection suggests that the shares of individuals who completed a vocational program is rather stable over age cohorts in most of these countries, with the exceptions of Denmark and Finland (and, to a lesser extent, France) indicating a decline in vocational attendance

4 The categorization follows the one applied in Hanushek et al. (2017), updated with the more recent statistics of PIAAC and EAG 2008.

5 The restriction to males with their historically stable aggregate labor-force participation patterns during prime age circumvent concerns raised about our identification by cohort-specific selection into work by females.

6 While tertiary vocational programs are likely more heterogeneous in the mix of general skills obtained, our results are robust to restricting the analysis to the subsample of individuals completing just secondary education for whom PIAAC explicitly provides a classification of education type (not shown).

Table 1. Descriptive statistics

	(1) Full sample	(2)	(3)	(4) Individuals with	(5)
	Mean	Minimum	Maximum	Vocational education	General education
Employed	0.793	0	1	0.769	0.836
	(0.405)			(0.421)	(0.371)
General education	0.358	0	1	0	1
	(0.479)				
Age	44.36	17	65	44.64	43.86
	(12.62)			(12.73)	(12.40)
Years of schooling	13.97	9	22	12.95	15.81
	(2.309)			(1.51)	(2.36)
Literacy score	282.8	51.5	445.1	271.5	303.2
	(44.9)			(42.3)	(42.2)
Numeracy score	289.1	48.2	467.0	277.5	310.0
	(48.7)			(45.9)	(46.6)
Observations	18,938			12,164	6774
Countries	10			10	10

Notes: Means, standard deviations (in parentheses), minimum, and maximum. Sample includes males aged 16–65 years with at least secondary education in the 10 vocational countries. Data weighted by sampling weights, giving same weight to each country.
Source: PIAAC.

over time. Employment rates are 84% for individuals with a general degree and 77% for those with a vocational degree. Literacy and numeracy scores are also higher for individuals with a general education.

3. Empirical Model

We focus on the impact of vocational vs. general education types on employment over the life cycle, with our main hypothesis being that any relative labor-market advantage of vocational over general education decreases with age. As developed in Hanushek et al. (2017), our baseline model is a simple difference-in-differences approach that compares the age–employment patterns of workers of the two education types within each country:

$$E_i = \alpha_0 + \alpha_1 A_i + \alpha_2 A_i^2 + \beta_1 G_i + \beta_2 G_i \cdot A_i + X_i \gamma + \mu_c + \epsilon_i \tag{1}$$

where E_i is an indicator capturing whether individual i is employed (in paid work during the past week); age A and its square capture the normal age–employment pattern in the economy; G_i is an indicator for general (as opposed to vocational) education type; X is a vector of control variables including years of schooling and skills; and μ_c are country fixed effects.

Our main coefficient of interest is β_2, which captures the differential impact of general relative to vocational education on employment with each year of age. In addition, β_1 measures the initial employment probability of general relative to vocational education (normalized to age 16 years in the empirical application). While we doubt that β_1 adequately captures the impact of general education because it implicitly includes any selectivity into education types not captured by X, the identifying assumption for β_2 is the standard assumption of the difference-in-differences approach that the selectivity of people into general vs. vocational education (conditional on X) does not vary over time. Put differently, to interpret our cross-sectional analysis as a pattern over the life cycle, we assume that conditional on the available observables, today's older people in each education type are a good proxy for today's younger people when they grow older.[7]

In our analysis below, we provide several tests of this assumption. First, to account for possible time-varying selection of individuals with differing ability into education types, we condition on the literacy and numeracy scores observed in PIAAC and, importantly, their interactions with age. Second, we control for two additional characteristics that may depict selection into education type and their interactions with age, namely, parental education and the number of books at home when a person was 15 years old. Third, given the cross-country nature of our main analysis, we can also condition on the share of each 10-year age cohort in a country that completed an education type, thereby holding overall changes in the size of each education type constant. Fourth, we use propensity score matching to identify a sample of individuals with vocational education that is observationally comparable to that for general education, thereby disregarding any individuals who do not have common support in the other education type. Together, these analyses provide strong support for an interpretation of the cross-sectional analysis as a life-cycle result.[8]

In addition, we can perform a straightforward direct test of whether selectivity into education types changed over time in our setting: we can estimate whether the effect of observed predictors of choice of education type varies with individuals' age. As is evident from Table 2, both individual test scores and socioeconomic status at the time of making educational choices—proxied by the number of books at home when an individual was 15 years old—are strong predictors of education type. In particular, individuals with higher literacy scores and more books at home are more likely to select into general education programs. Numeracy score also enters significantly in the absence of literacy scores, whereas only literacy retains significance in a model that considers both of them jointly. Mothers' education is marginally significantly positive in a model without books at home, but loses significance with books at home. More importantly, the interaction terms of all these variables with individuals' age are statistically insignificant. That is, we can observe a number of significant predictors of choice of education type, but the effect of none of them varies with age in the very setting of our analysis. While this does not preclude the possibility that unobserved characteristics of individuals with different

7 Reassuringly, Brunello and Rocco (2017) and Golsteyn and Stenberg (2017) confirm a trade-off of labor-market outcomes by education type over the life cycle with longitudinal data in Britain and Sweden, indicating that age differences reflect actual age effects rather than cohort effects that are specific to education types.

8 These analyses also address potential effects of changes in the extent to which the curricula of vocational programs contain general material. For example, reforms of vocational programs such as the Dutch reform studied by Oosterbeek and Webbink (2007) and the Swedish reform studied by Hall (2016) may have contained such curricular implications. To ensure that our results are not driven by these reforms, we confirm that results are robust to excluding the Netherlands and Sweden from our analysis (not shown).

Table 2. Correlates of general education type

	(1)	(2)	(3)
Literacy score	0.047***		0.054***
	(0.009)		(0.017)
Literacy score × age	−0.001		−0.003
	(0.003)		(0.005)
Numeracy score		0.036***	−0.008
		(0.009)	(0.017)
Numeracy score × age		−0.001	0.001
		(0.003)	(0.005)
Books at home at age 15	0.038***	0.041***	0.039***
	(0.010)	(0.010)	(0.010)
Books at home at age 15 × age	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)
Mother has high school education	0.032	0.034	0.032
	(0.028)	(0.028)	(0.028)
Mother has high school education × age	−0.000	−0.000	−0.000
	(0.001)	(0.001)	(0.001)
Age	−0.008***	−0.008***	−0.008***
	(0.001)	(0.002)	(0.002)
Age^2	0.015***	0.014***	0.015***
	(0.002)	(0.002)	(0.002)
Years of schooling	0.120***	0.121***	0.121***
	(0.001)	(0.002)	(0.002)
Country fixed effects	Yes	Yes	Yes
Observations	18,340	18,340	18,340
Countries	10	10	10
R^2 (adjusted)	0.436	0.434	0.436

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Notes: Linear probability model. Dependent variable: 1 = education type of individual is general; 0 = vocational. Sample includes males aged 16–65 years with at least secondary education in the 10 vocational countries. Age variable subtracted by 16 and divided by 10. Regressions weighted by sampling weights, giving same weight to each country. Robust standard errors in parentheses.

Source: PIAAC.

education types may have changed differently over time, this result provides plausibility to our identifying assumption that conditional time-varying selectivity into education types does not drive our results.

4. Employment Effects of Education Type over the Life Cycle

Our results in Table 3 indicate that there is indeed a strong trade-off of employment patterns by education type over the life cycle. Initially, individuals completing vocational education programs have higher employment probabilities. But with increasing age, this advantage declines and ultimately turns around into an employment advantage of

Table 3. Vocational vs. general education and employment over the life cycle in PIAAC

	(1)	(2)	(3)	(4)	(5)	(6) Propensity score matching	(7) 20+ age sample	(8) 30+ age sample
General education	−0.100***	−0.090***	−0.085***	−0.082***	−0.084***	−0.090***	−0.093***	−0.135***
	(0.017)	(0.018)	(0.018)	(0.019)	(0.018)	(0.027)	(0.018)	(0.026)
General education × age	0.032***	0.024***	0.022***	0.021***	0.022***	0.027***	0.025***	0.034***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.006)	(0.008)
Age	0.270***	0.260***	0.257***	0.255***	0.260***	0.260***	0.252***	0.453***
	(0.013)	(0.013)	(0.013)	(0.015)	(0.013)	(0.015)	(0.013)	(0.027)
Age^2	−0.066***	−0.062***	−0.062***	−0.062***	−0.063***	−0.062***	−0.062***	−0.091***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)
Years of schooling	0.021***	0.016***	0.015***	0.015***	0.015***	0.020***	0.015***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Literacy score		0.001	−0.002	−0.000	−0.003	0.028	−0.008	−0.017
		(0.009)	(0.017)	(0.017)	(0.017)	(0.022)	(0.017)	(0.025)
Literacy score × age		0.014***	0.002	0.002	0.002	−0.007	0.004	0.007
		(0.003)	(0.006)	(0.006)	(0.006)	(0.008)	(0.006)	(0.008)
Numeracy score			0.006	0.003	0.007	−0.002	0.011	0.029
			(0.017)	(0.017)	(0.017)	(0.021)	(0.017)	(0.025)
Numeracy score × age			0.014**	0.014**	0.014**	0.017**	0.012**	0.007
			(0.006)	(0.006)	(0.006)	(0.008)	(0.006)	(0.008)

(continued)

Table 3. Continued

	(1)	(2)	(3)	(4)	(5)	(6) Propensity score matching	(7) 20+ age sample	(8) 30+ age sample
Share of country cohort with general education					−0.133**	−0.125	−0.144**	0.178*
					(0.066)	(0.080)	(0.066)	(0.093)
Mother's education (two indicators and their interaction with age)				Yes				
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,938	18,938	18,938	18,372	18,938	12,374	18,745	15,691
Countries	10	10	10	10	10	10	10	10
R^2 (adjusted)	0.138	0.146	0.149	0.148	0.149	0.122	0.150	0.175

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Notes: Linear probability model. Dependent variable: individual is employed. Sample includes males aged 16–65 years with at least secondary education in the 10 vocational countries. Age variable subtracted by 16 and divided by 10. Regressions weighted by sampling weights, giving same weight to each country. Robust standard errors in parentheses.

Source: PIAAC.

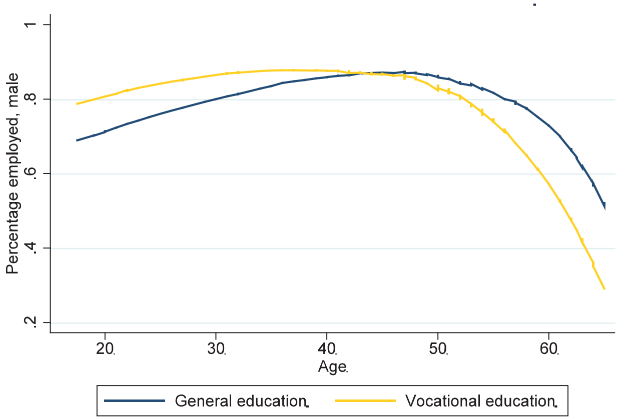


Figure 1. Employment by age and education type in apprenticeship countries.

Notes: Sample includes males who completed at least secondary education and are currently not students in the three 'apprenticeship countries' (Austria, Denmark, and Germany), based on a matched sample that uses propensity score matching to ensure common support between persons with a vocational and a general education in each country (see text for details of the matching algorithm). Smoothed scatterplot using locally weighted regressions (Stata lowess). Data source: PIAAC.

individuals completing general education programs (see also the descriptive pattern in Figure 1).[9] Using the sample of 10 countries with significant vocational systems, the first column of Table 3 shows the simplest model that conditions only on country fixed effects, a quadratic in age, and years of schooling. At age 16 years, the employment probability of persons with a vocational education is 10.0 percentage points higher. But with every 10 years of age, this declines significantly by 3.2 percentage points, which is even larger than the 2.1 percentage points found in the equivalent specification of Hanushek et al. (2017) for the mid-1990 s. The interacted specification implies that starting with age 48 years, persons with a general education have a higher employment probability.

As discussed above, the main concern with identification from the age gradient in relative employment in this difference-in-differences approach is that within countries, selectivity into the two education types may have changed over time. As a first check on this possibility, Column 2 adds the PIAAC literacy score and its interaction with age. On the one hand, this inclusion captures any change in selectivity of individuals with initially different basic skill levels into different education types that is reflected in differences in observed adult skills. On the other hand, these skills could in part be endogenous to specific education types and to work histories, thereby taking out more of the identifying variation than it should. Specifically, if the education programs and employment experiences of

9 It is apparent from the figure that the gap between the two curves moves to the advantage of general education in a rather linear fashion, favoring the linear-in-age interaction specification of the empirical model. However, specifications with interaction terms that are nonlinear in age indicate that the differential pattern of employment between vocational and general education is particularly pronounced starting in the mid-50 age range (not shown). While Figure 1 is based on a matched sample of vocationally educated and generally education individuals, the same qualitative pattern emerges in a purely descriptive figure of the full sample.

generally educated individuals lead them to gain and maintain more literacy skills relative to vocationally educated individuals, conditioning on adult literacy skills will lead to an underestimation of β_2, our coefficient of interest in Equation (1). In any event, while the association of literacy with employment indeed increases with age, the main pattern of results remains unchanged, with a slightly reduced coefficient on the type of education–age interaction. Given that the inclusion of controls for adult skills is likely to lead to conservative estimates in our setting, we keep including them throughout.[10]

While the inclusion of literacy scores follows the analysis with the IALS test in Hanushek et al. (2017), PIAAC in fact provides considerably richer testing of skills which allows us to estimate our main equation conditional on the different domains of cognitive skills tested in PIAAC. When we add the PIAAC numeracy score in addition to the literacy score (Column 3), literacy in fact loses significance, which is fully captured by numeracy. However, our qualitative results do not change.[11]

As another control for potential differential selectivity into education over time, Column 4 adds controls for the education level of respondents' mothers and its interaction with age. These turn out insignificant and hardly change our substantive results.[12] The same is true in a model without literacy scores (considering the potential endogeneity of adult skills) or when adding the number of books at home at age 15 years as another background control, which enter the model significantly with or without skill controls (not shown).

To account for potential effects of changes in the aggregate composition of the labor force by type of education over time, Column 5 adds the percentage of each 10-year age cohort completing general education in each country; results are hardly affected.[13] In this main specification, for each 10 years of age, the relative employment chances of those with a general education increase by 2.2 percentage points relative to those with a vocational education, which is effectively the same as found in the base specification of Hanushek et al. (2017) for the mid-1990s.

As another approach to address possible selection issues, Column 6 shows results of a model using propensity score matching to compare individuals with a vocational education only to observationally similar individuals with a general education. We use nearest-neighbor matching which, for each country, matches each individual with vocational education to one individual with general education based on age, years of schooling, literacy and numeracy scores, and mother's education, so that the estimate is only identified from

10 While the basic literacy and numeracy skills captured by the PIAAC tests may be part of the set of general skills of which general education programs provide more than vocational programs, they do not capture many other aspects of general skills such as other cognitive skills, social-interaction skills, and learning-to-learn skills.

11 Despite the high correlation between literacy and numeracy (0.85), our results are effectively unchanged when including only numeracy or when using the average of literacy and numeracy. Interestingly, the new PIAAC domain of 'problem solving in technology-rich environments' (not available in France and Spain) does not enter our employment equation significantly (individually or jointly with the other domains) and does not affect our results.

12 The same holds for father's education and parents' highest education, which are missing more observations.

13 The age pattern of employment by education type is also robust to adding the average skill scores of individuals with the particular education type by country and 10-year age cohort, as in Hanushek et al. (2017).

Table 4. Heterogeneity across country groups with different vocational intensity

	(1) All countries	(2) Nonvocational countries	(3) Vocational countries	(4) Nonschool-based vocational countries	(5) Apprenticeship countries	(6)	(7)	(8)
					All	Austria	Denmark	Germany
General education	−0.063***	−0.001	−0.084***	−0.123***	−0.134***	−0.083	−0.110**	−0.201***
	(0.014)	(0.024)	(0.018)	(0.032)	(0.035)	(0.062)	(0.046)	(0.067)
General education × Age	0.019***	−0.000	0.022***	0.041***	0.049***	0.064***	0.036**	0.043*
	(0.005)	(0.009)	(0.006)	(0.011)	(0.012)	(0.022)	(0.015)	*(0.022)*
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	29,452	10,514	18,938	8040	6004	1719	2365	1920
Countries	16	6	10	4	3	1	1	1

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Notes: Linear probability model. All models include the same controls as Column 5 of Table 3. Dependent variable: individual is employed. Sample includes males aged 16–65 years with at least secondary education. See Section 2 for country groups. Age variable subtracted by 16 and divided by 10. Regressions weighted by sampling weights, giving same weight to each country. Robust standard errors in parentheses.

Source: PIAAC.

common support between the two groups within each country. While this reduces the number of observations by 35%, our main result in fact becomes stronger, indicating that it is not driven by observations off the common support.

A final concern is selectivity at young ages because some young people are still in the education system, particularly in general programs. Thus, Columns 7 and 8 restrict the sample to persons at least 20 and 30 years of age, respectively. In fact, the age pattern of employment by education type gets stronger in these reduced samples (in contrast to Hanushek et al. 2017).

5. Heterogeneity across Countries

As indicated, countries differ widely in the treatment intensity of their aggregate institutional vocationalization. While the previous results were restricted to the 10 countries with significant vocational systems, the first column of Table 4 shows that the main results also hold in the full sample of 16 countries, albeit at reduced coefficient size. In fact, Column 2 shows that the pattern is not at all visible in the nonvocational countries, with effectively no employment differences across education types. This result may reflect the vagueness of the definition of general vs. vocational types of education programs in countries with limited vocational systems.

In contrast, results are substantially stronger in countries with nonschool-based vocational systems (Column 4) and, in particular, in countries with extensive apprenticeship systems (Column 5). The heterogeneous results across country groups may reflect an increasing treatment intensity of vocational specificity: The apprenticeship programs with their substantial industry-based education tend to provide the highest intensity of vocational experience (cf. Wolter and Ryan 2011). The cross-over age from which on employment is higher for general than for vocational education is as low as 44 years on average across the apprenticeship countries. In fact, despite the smaller sample sizes, the main pattern is significantly visible in all three apprenticeship countries (Columns 6–8), with the Austrian results providing confirmation in a country that had not participated in IALS. The overall pattern across country groups is consistent with the employment effects of education types increasing with the treatment intensity of occupation-specific education in the vocational system.[14]

6. Conclusions

Using recent data on labor markets in a large sample of countries, we aim to provide a deeper understanding of the merits and limitations of different education types for employment in an increasingly globalized era. We find strong evidence that a life-cycle perspective is important: while individuals who completed vocational education programs initially have better employment opportunities than individuals who completed general education programs, this pattern turns around at older ages. While estimates vary across specifications, the estimated cross-over age by which individuals with a general education have a

14 These results suggest that the opposing interpretation in Forster, Bol, and van de Werfhorst (2016) may stem from peculiarities in their standardized index of vocational systems, as well as their inclusion of countries with unclear identification of education types in PIAAC.

higher employment probability than individuals with a vocational education is around 50 years, and somewhat earlier around 45 years in the apprenticeship countries. These estimates are broadly in line with the range of estimates found for the mid-1990s in Hanushek et al. (2017), although they tend to indicate a slightly earlier cross-over age in the early 2010s. The findings are also consistent with the general pattern suggested by a number of recent country studies that show a similar age pattern of labor-market outcomes by education type over the life cycle.

The estimated impact of education type on the age–employment profile is consistent with vocational education improving the transition from school to work but reducing adaptability of older workers to economic change. This pattern is particularly pronounced in countries with apprenticeship systems, whose emphasis on industry-based education may provide the strongest treatment intensity of vocationalization.

From an individual perspective, the results imply that people should be aware that there is a trade-off between early advantages and later disadvantages of vocational vs. general education programs over the employment life cycle. The topics of facilitated entry vs. later adaptability indicate that there are both pros and cons of vocational education and of general education. The relative merits will depend on many factors, including the imminence of disruptions from technological or structural change in a specific sector or occupation in the country, the individual's inclination for adaptability and change in general, and the rate at which the individual discounts the future.

From a policy perspective, our results suggest caution about policies that concentrate just on the current employment situation and ignore the dynamics of growing economies. Current policy discussions often focus narrowly on issues of labor-market entry and youth unemployment. For a full assessment of how different education types affect the labor-market chances of workers, however, policy has to set the potential advantages of vocational programs in facilitating the transition from school to work against potential disadvantages when people have to adjust to changing conditions later in life. For countries with extensive vocational systems, the results may suggest that reducing the early specialization of students on specific occupational skills may be conducive to their long-run prospects on the labor market. In addition, the results indicate that it may be worth considering the establishment of a system for lifelong learning that does not only update workers' skills within their occupation but also conveys skills that facilitate their flexibility if changing labor-market conditions require occupational change.

Acknowledgements

The authors would like to thank Thijs Bol, Eric Hanushek, Guido Schwerdt, Simon Wiederhold, the editor Matz Dahlberg, two anonymous referees, and participants at the ROA workshop on 'Human Capital over the Life Cycle' in Maastricht and at the International Conference on PIAAC and PIAAC-Longitudinal in Mannheim for their comments. Financial support by the Leibniz Competition for the research project 'Acquisition and Utilization of Adult Skills' (SAW-2015-GESIS-2) is gratefully acknowledged.

References

Autor, D. H., D. Dorn, and G. H. Hanson (2015), "Untangling Trade and Technology: Evidence from Local Labour Markets", *Economic Journal* 125, 621–46.

Broecke, S., G. Quintini, and M. Vandeweyer (2017), "Wage Inequality and Cognitive Skills: Re-opening the Debate", in C. Hulten and V. Ramey, eds, *Education, Skills, and Technical*

Change: Implications for Future U.S. GDP Growth, University of Chicago Press, Chicago, IL. NBER Book Series Studies in Income and Wealth, forthcoming.

Brunello, G. and L. Rocco (2015), "The Effects of Vocational Education on Adult Skills and Wages: What Can We Learn from PIAAC?", OECD Social, Employment and Migration Working Papers 168, OECD, Paris.

Brunello, G. and L. Rocco (2017), "The Labor Market Effects of Academic and Vocational Education Over the Life Cycle: Evidence Based on a British Cohort", *Journal of Human Capital* 11, 106–66.

Cörvers, F., H. Heijke, B. Kriechel, and H. Pfeifer (2011), *"High and Steady or Low and Rising? Life-Cycle Earnings Patterns in Vocational and General Education"*, ROA Research Memorandum ROA-RM-2011/7, Research Centre for Education and the Labour Market, Maastricht.

Deutsche Rentenversicherung Bund (2015), *Rentenversicherung in Zeitreihen*, Deutsche Rentenversicherung Bund, Berlin.

Falck, O., A. Heimisch, and S. Wiederhold (2016), "Returns to ICT Skills", CESifo Working Paper 5720, CESifo, Munich.

Forster, A. G., T. Bol, and H. G. van de Werfhorst (2016), "Vocational Education and Employment Over the Life Cycle", *Sociological Science* 3, 473–94.

Golsteyn, B. and A. Stenberg (2017), "Earnings Over the Life Course: General versus Vocational Education", *Journal of Human Capital* 11, 167–212.

Hall, C. (2016), "Does More General Education Reduce the Risk of Future Unemployment? Evidence from an Expansion of Vocational Upper Secondary Education", *Economics of Education Review* 52, 251–71.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015), "Returns to Skills Around the World: Evidence from PIAAC", *European Economic Review* 73, 103–30.

Hanushek, E. A., G. Schwerdt, L. Woessmann, and L. Zhang (2017), "General Education, Vocational Education, and Labor-market Outcomes Over the Life-cycle", *Journal of Human Resources* 52, 48–87.

Kahn, L. M. (2016), "Permanent Jobs, Employment Protection and Job Content", IZA Discussion Paper 9961, Institute for the Study of Labor, Bonn.

Krueger, D., and K. B. Kumar (2004), "Skill-specific Rather Than General Education: A Reason for US-Europe Growth Differences?", *Journal of Economic Growth* 9, 167–207.

Levels, M., R. van der Velden, and J. Allen (2014), "Educational Mismatches and Skills: New Empirical Tests of Old Hypotheses", *Oxford Economic Papers* 66, 959–82.

Malamud, O. and C. Pop-Eleches (2010), "General Education versus Vocational Training: Evidence from an Economy in Transition", *Review of Economics and Statistics* 92, 43–60.

Oosterbeek, H. and D. Webbink (2007), "Wage Effects of an Extra Year of Basic Vocational Education", *Economics of Education Review* 26, 408–19.

Ryan, P. (2001), "The School-to-Work Transition: A Cross-national Perspective", *Journal of Economic Literature* 39, 34–92.

Shavit, Y. and W. Müller eds. (1998), *From School to Work: A Comparative Study of Educational Qualifications and Occupational Destinations*, Clarendon Press, Oxford.

Stenberg, A. and O. Westerlund (2015), "The Long-term Earnings Consequences of General vs. Specific Training of the Unemployed", *IZA Journal of European Labor Studies* 4, 1–26.

Weber, S. (2014), "Human Capital Depreciation and Education Level", *International Journal of Manpower* 35, 613–42.

Wolter, S. C. and P. Ryan (2011), "Apprenticeship", in: E. A. Hanushek, S. Machin, and L. Woessmann, eds, *Handbook of the Economics of Education*, Vol. 3, North Holland, Amsterdam, pp. 521–76.

Zimmermann, K. F., C. Biavaschi, W. Eichhorst, C. Giulietti, M. J. Kendzia, A. Muravyev, J. Pieters, N. Rodríguez-Planas, and R. Schmidl (2013), "Youth Unemployment and Vocational Training", *Foundation and Trends in Microeconomics* 9, 1–157.

CESifo Economic Studies, 2017, 270–294
doi: 10.1093/cesifo/ifx015
Advance Access Publication Date: 23 September 2017
Original article

Export Hysteresis, Capacity Constraints, and Uncertainty: A Smooth Transition Analysis for Euro Area Member Countries

Ansgar Belke* and Jan Wagemester‡

*University of Duisburg-Essen, Duisburg, Germany, CEPS Brussels, Brussels, Belgium and IZA Bonn, Bonn, Germany. e-mail: ansgar.belke@uni-due.de and †University of Duisburg-Essen, Duisburg, Germany. e-mail: jan@wagemester.com

Abstract

We argue that, under certain conditions described by a sunk cost hysteresis model, firms consider exports as a substitute for domestic demand. This is valid also on the macroeconomic level where the switch from the domestic market to the export market and vice versa takes place in a smooth manner. Areas of weak reaction of exports to changes in domestic demand are widened by uncertainty. Our econometric model for six euro area countries suggests domestic demand and capacity constraints as additional variables for export equations. We apply the exponential and logistic variant of a smooth transition regression model and find that domestic demand developments and uncertainty are relevant for short-run export dynamics particularly during more extreme stages of the business cycle. A substitutive relationship between domestic and foreign sales can most clearly be found for France, Greece, and Ireland (exponential smooth transition regression (ESTR) model) and France, Portugal, and Italy (logistic smooth transition regression (LSTR) model), providing evidence of the importance of sunk costs and hysteresis in international trade between the member states of the Economic and Monetary Union (EMU). What is more, our empirical results are robust to the inclusion of a variable measuring European policy uncertainty. In some cases (Italy, Greece, and Portugal) the results underscore the empirical validity of the export hysteresis under uncertainty model. (JEL codes: F14, C22, C50, C51, F10).

Key words: domestic demand pressure, exports, error-correction models, hysteresis, modelling techniques, smooth transition models, sunk costs, uncertainty

1. Introduction

A number of euro area countries which recorded large current account deficits in the period prior to the European debt and banking crisis starting in 2010 have seen a significant correction of their external imbalances, in particular the trade balance, over recent years. Falling imports have been an important part of this correction due to low domestic demand. However, at the same time, exports and export market shares have been continuously increasing in most of these countries since 2009. Shrinking unit labour costs and falling real effective exchange rates are able to explain only part of the gains in export market shares. Christodoulopoulou and Tkacevs (2014) find that only 60–70% of variation in exports can be explained by standard export equations. It thus seems likely that non-price-related factors have been important in explaining export performance. The residuals from a standard approach to model exports are potentially consistent with the parallel dramatic fall of domestic demand. A possible relationship between domestic demand and exports could be particularly important in the current economic situation of substantial macroeconomic adjustment needs and very low domestic demand.

The relation between domestic demand and exports is not straightforward and could be either negative (substitutive) or positive (complementary). A recent survey of literature on this topic is presented in Esteves and Rua (2013). Theoretical reasons for a positive link between domestic demand and exports may be due to increased efficiency from learning by doing effects (Belke et al. 2013) or due to liquidity generated by cash flow from exports which can help overcome liquidity constraints for domestic operations (Berman et al. 2011). Theory has identified a negative relationship between domestic demand and exports mostly at the firm level. Several studies have been concerned with the effects of domestic demand pressure on the inclination and capacity to export. These studies are not numerous, but go back several decades.[1]

The main argument is that—in the short run—exporting firms face capacity constraints or increasing marginal costs and thus have to substitute sales between their domestic and foreign markets. An increase in demand for exports cannot be satisfied in the short run as long as capacity is highly utilized and most of production is sold on the domestic market. Conversely, with low domestic demand, for instance during a domestic recession, firms will be able to shift more resources to export activities; to compensate for the decline in domestic sales, firms will increase their efforts to export. Besides pull factors (e.g. foreign demand), export performance can thus also be determined by push factors (such as low capacity utilization). Besides the studies mentioned above, more recent empirical literature (Ilmakunnas and Nurmi 2007; Máñez et al. 2008; Berman et al. 2011; Blum et al. 2011; Vannoorenberghe 2012 or Ahn and McQuoid 2013) generally identifies a significant negative effect of domestic demand pressure on exports for several countries, among them the UK, the USA, Germany, Spain, Israel, Turkey, Morocco, and India.

The main lesson from the literature is that any exercise of modelling export performance should take into account not only the factors driving external demand (and thus impact export activity from the demand side) but also those influencing domestic demand (which affect export activity mostly through the supply side). Moreover, the studies underline the necessity of clearly differentiating between the short run and the long run. One potential limitation of the previous literature is that the complementarity versus substitutability property of domestic demand and exports has often been analysed in a linear framework. The relationship between domestic demand and export performance may however vary with economic conditions and thus be of a non-linear nature.

Assuming a substitutive relationship between domestic demand and exports, following a domestic demand shock, firms will try to shift sales between the two markets. However,

1 Examples are Ball et al. 1966; Smyth 1968; Artus 1970, 1973; Dunlevy 1980; Zilberfarb 1980; and Faini 1994 and Sharma 2003.

entering the export market or shifting more sales towards it usually implies sunk costs. These are costs firm need to pay that are irreversible ex post (Baldwin and Krugman 1989), and the significance of this knowledge diminishes rapidly after leaving the export market (Belke et al. 2013).

In that respect, we can distinguish two cases. First, with a negative domestic demand shock and sunk costs for entering or shifting to the export market, firms will therefore be reluctant to pay these costs as long as capacity is still relatively highly utilized. Once capacity utilization falls below a certain threshold, firms might be more willing to pay sunk investment costs, as these costs and the effort of selling in the foreign market might be lower than the cost of running excess capacity.[2] Exports in this case can be considered as 'survival-driven'. Secondly, following a positive domestic demand shock, firms might not be able to serve both domestic and foreign markets due to highly utilized capacities. If they prefer producing for the domestic market, firms would consider shifting sales to that market once a certain high capacity utilization threshold has been crossed. With sunk costs, leaving the export market or shifting sales away from it implies that these costs would have to be paid again upon trying to re-enter the export market or reshifting sales towards it in the future.

Overall, these arguments suggest that only if certain low or high capacity utilization thresholds have been crossed, firms will change their export behaviour. Only if a domestic demand shock is accompanied by extreme changes in capacity utilization will firms shift their sales to another market. As long as capacity is utilized to a more normal degree and operates within these lower and upper thresholds, firms are working in a 'band of inaction' where sunk costs hinder firms from changing their export behaviour, even though capacities might exist for those firms that are not yet very active in foreign markets.[3] This also implies that, once capacity utilization thresholds have been crossed on either end and firms have shifted sales among markets, they will be reluctant to shift again once capacity returns back to more normal levels. There is thus strong persistence in export behaviour which can be traced back to the theory of hysteresis (Baldwin and Krugman 1989). Export hysteresis is the tendency of a temporary change in export behaviour to become permanent. It is particularly important in the current weak economic situation of several euro area member states; firms increase efforts to shift sales to the export market, given weak domestic demand, and this might not be a cyclical change but rather a persistent improvement as firms will often decide to stay in the foreign market even once domestic demand picks up again as they are trying to avoid repaying sunk costs.

We thus essentially present a story of dynamic investment in the presence of high fixed cost and capacity constraint. This story is consistent with firms switching from selling in the domestic market to the foreign market as soon as the level of domestic demand falls short of a given trigger threshold. But what about the heterogeneity element that induces switching only by some firms, but not in all firms? Here we refer to Belke and Goecke

2 Alternatively, some firms might be constrained by technical limitations that allow production at a certain capacity utilization rate only; facing a certain low capacity utilization threshold, they might face the decision to either not produce at all or shift their production to serving foreign markets.

3 In the European case and the countries under consideration, potential for shifting production to foreign markets seems to exist. As an example, Esteves and Rua (2013) specify that in 2010, only one-third of Portuguese manufacturing firms were exporting and for them the exports to sales ratio was on average around 30%.

(2005) who, starting from the idea of a 'band of inaction', focus on the issue of aggregation. They are able to derive an aggregation process, considering heterogeneity of sunk exit/entry costs and/or the extent of uncertainty of the future market situation and/or the elasticity of demand. This is resulting in different triggers for different firms. This (realistic) consideration of heterogeneity alters the hysteresis characteristics when aggregating from the microeconomic to the macroeconomic level. Due to heterogeneity in firm characteristics such as the magnitude of sunk costs or the productivity level, firms exit (and entry) sequentially and not all in a time from (into the) the market, and the resulting aggregated hysteresis loop thus shows no discontinuities. This is rather important in our context because, absent this feature, all firms would switch if there is a large negative domestic demand shock. This would contradict the abundant micro-evidence in the trade literature that actually the most engaged exporters are also faring best on the domestic market.[4]

Notably, in this model of export hysteresis, the band of inaction is widened by uncertainty (Belke and Goecke, 2005). This is because a forward-looking firm considers future effects of a present sunk cost 'investment'. If the exogenous variable demand is stochastic, a real option approach applies (Dixit 1989, Pindyck 1998, 1991; Belke and Goecke 2001). An inactive firm deciding on a present entry or to stay passive will include the option to enter later as a potential alternative. Demand which is presently contributing to cover costs may in a stochastic situation decrease in the future. By staying passive the firm can avoid future losses if this situation will realize. Moreover, an instantaneous entry kills the option to enter later and to 'wait-and-see' if the future demand movement will turn out to be (un)favourable. Thus, in a stochastic situation, the sunk costs and, additionally, an option value of waiting have to be covered to trigger an entry. Therefore, uncertainty implies an upward shift of the entry trigger demand. The same is valid for the exit trigger demand which will shrink in a situation with uncertainty. Belke and Goecke (2005) show that this line of reasoning is valid also at the macroeconomic aggregated level. Thus, uncertainty leads to a widening of the band of inaction also at the macroeconomic level, aggravating the hysteresis property of the firm's export behaviour.[5]

Our article builds on this sunk cost hysteresis model and explicitly tests for a short-run non-linear relationship between domestic demand and exports from a macroeconomic perspective. A particular asymmetric effect was already considered in Esteves and Rua (2013) for the case of Portugal. Belke, Oeking, and Setzer (2015) consider the relation of domestic demand and export of goods in several euro area countries. Our analysis goes beyond these studies by investigating six euro area countries with significant current account deficits in the pre-crisis period (Spain, Portugal, Italy, France, Ireland, and Greece) employing the export of both goods and services.

The focus upon these former current account deficit countries reflects our intent to analyse countries in which firms were experiencing a fall in domestic demand. After the start of EMU, a real appreciation sets in for peripheral Euro area member countries, especially in those experiencing a massive housing boom, namely, Ireland and Spain. As a consequence, their competitiveness went down further. At the same time nominal long-term interest rates converged among EMU member countries, inducing low real interest rates in the peripheral

4 The aggregation procedure of firm heterogeneity under consideration is explained in detail in Belke and Goecke (2005), pp. 196–201.

5 The aggregation procedure under consideration of firm heterogeneity and uncertainty is explained in detail in Belke and Goecke (2005), pp. 189–192.

countries with higher inflation. This in turn stifled spending and inflation even further, leading to growing current account deficits (see, for instance, Krugman et al. 2015, pp. 687ff.). Since currency devaluation was no option for these countries, it became clear that the necessary real exchange rate adjustment implied a period of low inflation or even deflation in combination with significant unemployment and protracted recession including weakness of domestic demand.[6] The 'doom loop' among banks and governments contributed significantly to this development.

Moreover, we go beyond the papers mentioned above by thoroughly conducting tests for structural breaks common to the countries under investigation and integrating an uncertainty variable in our estimations.

Following Belke et al. (2015), we implement a smooth transition regression (STR) model such that we can specify aggregated non-linearities with a high degree of flexibility. We argue that the strength of the relation between domestic demand and exports depends on capacity constraints and more generally the business cycle. Besides the possibility that substitutability will increase after reaching either the upper or lower threshold (i.e. giving rise to symmetry around the band of inaction), we also allow for the possibility that exports react sharper in a recession than during an economic expansion (giving rise to asymmetry around the band of inaction). This is achieved by relying on either an exponential or logistic variant of smooth transition specification. The aggregation at the macro level allows us to draw results on net effects of capacity utilization on the economies as a whole. This is of special importance in the discussion of macroeconomic adjustment and the reduction of current account imbalances in the euro area.

The article proceeds as follows. Taking the simple sunk cost-based hysteresis model as a starting point, we carry out some pretesting in terms of unit roots and cointegration in Section 2. Based on the cointegration results, we set up an error-correction export equation and incorporate non-linearities as suggested by our theoretical considerations. These STR models, including several robustness tests among them the incorporation of an uncertainty variable, are estimated in Section 3. Section 4 finally concludes and conveys an outlook on further research avenues.

2. Empirical Strategy

2.1 Data

Our data stem from different sources (cf. Table A1): data on real exports (x_t) and real domestic demand (dd_t) come from the national statistical offices (either obtained from Eurostat or Oxford Economics). Value-added exports (x_t^{va}) have been constructed by data from the World Input-Output Database (wiod.org); the annual data were converted to quarterly data by applying cubic spline interpolation. The real effective exchange rate has been obtained from Eurostat and is an index deflated by consumer price indices with a

6 See, for instance, Belke and Gros (2017). But, for example, in Ghironi and Melitz (2005), a negative current account emerges also in times of a positive productivity shock or a reduction of entry barriers. In this case, the home economy experiences the most attractive conditions and becomes nevertheless a net borrower on international markets to finance the creation of new firms that a positive productivity shock or the reduction of competitive barrier warrants. However, this was not the scenario the six EMU member countries in our sample were faced with.

country's 15 main trading partners (r_t). Alternatively, the same source provides an index deflated by unit labour costs with a country's 24 main trading partners (r_t^{ULC}). Data on foreign demand (y_t^*) from the ECB are based on trade-weighted imports for a country's 15 main trading partners. Capacity utilization data in the manufacturing industry (z_t) come from the Business and Consumer Surveys by the European Commission, available from Eurostat or Insee in the case of France. For Ireland, data on capacity utilization are not available. Instead, we use the output gap (interpolated data from AMECO). As an uncertainty variable for our robustness checks, we employ the economic policy uncertainty index relevant for the European Union as a whole because the respective index was not available for the individual Euro area member countries for such a long sample period like ours (http://www.policyuncertainty.com/europe_monthly.html). The final data set is quarterly and mostly available from 1980:Q1 to 2012:Q4.

2.2 Non-stationarity and cointegration tests

By focusing on the volume of exports for a specific country as our main purpose of this article, it is necessary to specify a function, which depends on foreign demand and the difference in price levels concerning trading partners in the long run. For this purpose, an equation

$$x_t = b_1 + b_2 y_t^* + b_3 r_t + b_4 d + b_5 d \cdot y_t^* + b_6 d \cdot r_t + e_t \qquad (1)$$

is specified, where x_t is the logarithm of exports, y_t^* the log of foreign demand, r_t the log of the real effective exchange rate, and d is a dummy. By taking logarithms on each side of the equation, we apply a non-linear framework to short-run effects. Before applying the Engle–Granger approach (1987) to test for cointegration, we need to check whether the variables introduced above are non-stationary. In an augmented Dickey–Fuller test (ADF test), we include an intercept for the real effective exchange rate series and an intercept plus a time trend for all other series and specify the auxiliary regression accordingly. In addition, using the two-break minimum Lagrange multiplier (LM) unit root test (Lee and Strazicich 2003) trend stationarity is established if the null hypothesis is rejected. The ADF test is complemented by LM unit root tests and corroborates our ADF results. The results are displayed in Table A2.

We adopt a methodology developed by Bai and Perron (1998, 2003) to account for possible instabilities in the long-run coefficients of Equation (1). Their basic idea is to choose break points such that the sum of squared residuals for all observations is minimized. The estimated break points by definition represent the linear combination of these segments which achieve a minimum of the sum of squared residuals (Bai and Perron, 2003). Table 1 shows the two most important break points for the six countries analysed, accompanied by the 95% upper and lower confidence intervals. In case of Spain the first break point occurs in 1993Q4 with 1993Q3 and 1994Q1 providing the 95% confidence intervals.

The results of Table 1 are also useful in the context of unifying our testing and estimation approach. One may ask, for instance, to what extent the differences in the cointegrating test and cointegrating equation estimation results across countries we usually gain for the export equations of the six EMU member countries analysed here (available on request)

Table 1. Break points with lower and upper 95%

Country	Break point	Lower 95%	Upper 95%
Spain	1993 Q4	1993 Q3	1994 Q1
	2004 Q1	2004 Q1	2004 Q4
Portugal	1986 Q3	1986 Q1	1987 Q1
	1993 Q4	1993 Q2	1994 Q3
Italy	1993 Q4	1993 Q3	1994 Q4
	1997 Q4	1997 Q3	1998 Q1
France	1983 Q2	1983 Q1	1984 Q1
	1993 Q4	1993 Q3	1994 Q1
Ireland	1993 Q4	1993 Q3	1994 Q1
	2001 Q2	2001 Q1	2001 Q3
Greece	1985 Q3	1985 Q2	1986 Q1
	1996 Q4	1996 Q3	1997 Q1

Notes: The table provides the two most important break points according to the Bai and Perron (1998) methodology for all countries under investigation.

are driven by the fact that the specific break points are different across countries?[7] To check this, we follow the option to see what happens when imposing a common break point for all countries. Do the data strongly reject such an assumption? To account for this issue, we have implemented one common break point for all countries in all estimations contained in this article with an eye on the results displayed in Table 1, that is by a (permanent) dummy denoting the most common break in 1993:04 which may proxy the fallout of the 1992/1993 crisis of the European Monetary System (EMS).[8]

The respective findings (long-term relation and Engle -Granger test for cointegration) based on a common break in the fourth quarter of 1993 are provided in Table 2, alternative specifications with country-specific break points were contained in the previous version of this article and are available upon request. To be more concrete, we focus on estimating the long-run equilibrium of Equation (1) by FMOLS (fully modified ordinary least squares). To test for cointegration, we apply the Engle–Granger test for cointegration. The results are displayed in the last column with the respective critical values from MacKinnon (1991). Because $\hat{e}_t \sim I(0)$, we can conclude that for each country the error-terms are stationary and a cointegration relationship between the variables is thus present. It should be mentioned that the findings are not greatly affected by the specification of the break points. The error terms from estimations based on common and individual break points turn out to be highly correlated and the short-term findings provided in the following are not affected by these findings.

The sign of the estimated coefficients (negative for our exchange rate variable and positive for the foreign demand variable) overall corresponds with our priors from theory. The effects are in line with theory. As an example, the effect of y (exp) is always positive. Our estimation results for the long-run relations largely match those of other studies, both in terms of sign and size of the coefficients (see, for instance, European Commission 2011).

7 We owe this point to an anonymous referee.

8 Estimation results for a common dummy denoting the introduction of the Euro are available on request.

Table 2. Long-run relationships and Engle–Granger test for cointegration based on a specification with common breaks

Country	Long-run relationship	Engle–Granger test
Spain	$x_t = 4,444^{***} - 0,189d - 1,023^{***}r_0 - 0,867^{***}r_1 + 1,161^{***}y_0 + 1,164^{***}y_1$	−4.5136***
	(147,62) (−0,56) (−9,80) (−4,17) (9,37) (20,84)	
Portugal	$x_t = 3,749^{***} - 0,765d - 1,075^{***}r_0 - 0,280r_1 + 1,246^{***}y_0 + 0,828^{***}y_1$	−2.6114*
	(88,38) (−0,92) (−20,52) (−0,59) (10,68) (21,71)	
Italy	$x_t = 4,751^{***} + 0,347d - 0,724^{***}r_0 - 0,609^{**}r_1 + 0,851^{***}y_0 + 0,555^{***}y_1$	−3.9215*
	(154,77) (0,88) (−17,10) (−2,55) (9,93) (17,38)	
France	$x_t = 4,772^{***} + 1,492^{***}d - 0,523^{***}r_0 - 1,230^{***}r_1 + 0,604^{***}y_0 + 0,629^{***}y_1$	−3.8773*
	(230,07) (5,13) (−25,93) (−8,74) (33,22) (26,50)	
Ireland	$x_t = 3,951^{***} - 1,649^{***}d - 1,405^{***}r_0 - 0,596^{**}r_1 + 1,585^{***}y_0 + 1,733^{***}y_1$	−4.0386*
	(69,19) (−3,92) (−19,89) (−2,18) (19,78) (17,69)	
Greece	$x_t = 3,501^{***} + 2,683^{**}d - 0,331^{***}r_0 - 2,417^{***}r_1 + 0,370^{***}y_0 + 1,287^{***}y_1$	−4.5760***
	(71,39) (2,53) (−6,45) (−3,92) (10,55) (6,66)	

Notes: The final column tests the null hypothesis that there is no cointegration (i.e. that the residual series has a unit root). The (approximate) critical values for the t-test are taken from MacKinnon (1991) for the respective number of variables. */**/*** statistical significance at the 10/5/1% level. The common break point is located in 1993:04. d denotes a dummy which takes a value of 1 from 1994:01. For the regressors r and y, 0 denotes the first and 1 the second subperiod. The number of observations is 131 (Portugal, France, and Ireland), 128 (Italy), 111 (Spain) and 105 (Greece).

We do not come up with a more detailed analysis here, as our main focus is on the short-run relation, and slightly different long-run specifications did not change the following results in a noteworthy way.[9]

2.3 Empirical model

As explained above, we apply a non-linear framework to capture any short-run non-linear impact in the relation between domestic demand and exports regarding the state of the economy. We consider each country's economic condition by looking at deviations of its capacity utilization from its mean. Looking at short-run adjustments and in particular at the short-run relation between exports and domestic demand, we take into account the long-run equilibrium estimated above. For this purpose, we apply an error-correction model. As already mentioned in the introduction, we take into account the possibility of a non-linear adjustment process to a linear long-run equilibrium relationship depending on the state of the economy. Based on an economy's export performance where individual firm-level decisions are aggregated, it may not seem adequate to assume that this threshold is a sudden and abrupt change which is identical for all firms and which is commonly known; the smooth transition regression (STR) model thus allows for gradual regime change when the timing of the regime switch varies on an aggregated level.

According to Engle and Granger (1987), for every (long-run) cointegration model an error-correction model describes the short-run dynamics of the system. Our main interest is in parameter β, the short-run elasticity of exports to a change in domestic demand concerning the state of economy, looking at the capacity utilizations and especially its deviations from its mean (z_t). The long-run equilibrium Equation (1) takes the possibility into account that a non-linear adjustment process leads, depending on z_t, to the long-run equilibrium. The error-correction model (see Equation (2) below) derived from Equation (1) can best be modelled as a smooth transition regression (STR). We will therefore estimate the following error-correction model with non-linear short-run adjustment in STR form:

$$\begin{aligned}\Delta x_t = &\left[\alpha_1 + \sum_{i=0}^{n-1} \beta_{1i}\Delta dd_{t-i} + \sum_{i=0}^{n-1} \theta_{1i}\Delta y^*_{t-i} + \sum_{i=0}^{n-1} \mu_{1i}\Delta r_{t-i} + \sum_{i=1}^{n-1} \eta_{1i}\Delta x_{t-i} + \delta_1\hat{\epsilon}_{t-1}\right]\\ &+ \left[\alpha_2 + \sum_{i=0}^{n-1} \beta_{2i}\Delta dd_{t-i} + \sum_{i=0}^{n-1} \theta_{2i}\Delta y^*_{t-i} + \sum_{i=0}^{n-1} \mu_{2i}\Delta r_{t-i} + \sum_{i=1}^{n-1} \eta_{2i}\Delta x_{t-i} + \delta_2\hat{\epsilon}_{t-1}\right]\\ &F(z_{t-j}, \gamma, c) + u_t,\end{aligned} \tag{2}$$

$$\hat{\epsilon}_{t-1} = x_{t-1} - \hat{b}_1 - \hat{b}_2 y^*_{t-1} - \hat{b}_3 r_{t-1} - \hat{b}_4 d - \hat{b}_5 d \cdot y^*_{t-1} - \hat{b}_6 d \cdot r_{t-1} \tag{3}$$

as a non-linear short-run STR model which includes gradual regime changes when the timing of the regime switch varies on an aggregated level. Δx_t represents a function of lagged equilibrium errors (the error-correction term $\delta_1\hat{\epsilon}_{t-1}$, where $\hat{\epsilon}_t$ refers to the error term of the long-run cointegration relation between x_t, y^*_t , and r_t determined in the previous step),

9 As robustness checks, we also included additional variables in the long-run relation, e.g. trade openness, or restricted the coefficient for foreign demand to unity. Other non-price competitiveness variables could also have an influence on exports. As Esteves and Rua (2013) point out, the long-run results need to be interpreted with caution, as further structural breaks or these potential omitted variables could have an influence on the outcomes. Since our focus is on the short-run results, the short-run non-linear estimation appears to be relatively insensitive to slightly different long-run specifications.

changes in domestic demand dd_t, foreign demand y_t^*, the real effective exchange rate r_t, and past changes of its own value. The parameter δ is referred to as the adjustment effect which gives information about the speed of adjustment when there is disequilibrium and parameters $\alpha,\ \beta,\ \theta,\ \mu,\ \eta$ are the short-run effects. Our main parameter of interest is β, the short-run elasticity of exports to a change in domestic demand.

The main difference between our short- and long-run specifications is the inclusion of the domestic demand variable. Based on the theoretical arguments given in the introduction above, domestic demand should enter our estimations in the short run only.[10] In contrast to the long-run estimation, we do not include a structural break in the short-run estimations of Equation (2) because this specification already includes the smooth transition non-linearities. Furthermore, a break in the long-run relation does not imply that short-run dynamics change as well; by excluding these breaks we are also able to reduce our model's complexity.

The first set of brackets in Equation (2) is a standard linear error-correction model. Non-linearity is introduced via the second set of brackets which includes the same regressors, but is multiplied with the transition function $F(z_{t-j},\gamma,c)$. The transition function in a STR model is a smooth, continuous, and bounded function between 0 and 1. We consider two popular forms of smooth transition models based on the transition function. These are the logistic STR model (LSTR) and exponential STR model (ESTR). The LSTR model uses a logistic transition function of the following form:

$$F(z_{t-j},\gamma,c)=\left[1+\exp\left(\frac{-\gamma}{\sigma_z}(z_{t-j}-c)\right)\right]^{-1}$$

with

$$\gamma>0,$$

with the transition variable z distinguishing different regimes in our non-linear estimation. In our case z is operationalized by the degree of capacity utilization to capture business cycle effects (in particular in the manufacturing industry). We look at deviations of z from a threshold value c which we set as the average value of capacity utilization over our sample time period.[11] Smoothness parameter γ determines strength and speed of the transition and σ_z is the standard deviation of the transition variable. As the smoothness parameter γ depends on the scaling of the transition variable, we follow Teräsvirta (1998) and normalize it by σ_z to be scale-free.

The logistic transition function increases monotonically from 0 to 1, as the value of transition variable z increases. We can therefore distinguish two different regimes in the extreme and a gradual transition between these two: (i) negative deviations of the transition variable from its threshold: $\lim_{z_{t-j}\to-\infty} F(z_{t-j},\gamma,c)=0$, when the model collapses to just

10 As a robustness test, we also included domestic demand in the above long-run cointegration relation. Its coefficient did neither turn out to be statistically significant nor did it help to constitute a better long-run relation.

11 As a robustness check, we also apply the same estimations by looking at deviations of *z* from its mean value. Final results remain similar. Results are available from the authors upon request.

the first set of brackets in Equation (2), that is the linear part, and (ii) positive deviations of the transition variable from its threshold: $\lim_{z_{t-j}\to+\infty} \; F(z_{t-j}, \gamma, c) = 1$. The coefficients α, $\beta,\ \theta,\ \mu,\ \eta,\ \delta$ gradually change between these two extreme values with changing z_{t-j}.

In our setting, this implies testing the hypothesis that domestic sales are substituted by foreign sales once capacity utilization falls below a certain threshold. Further reductions in capacity reduction strengthen the substitution of domestic demand by exports. Note that there is no threshold for the opposite case of high capacity utilization. In other words, the band of inaction is only constrained on one side (for negative but not for positive deviations of capacity utilization from its mean).

The ESTR model relies on an exponential transition function of the following functional form:

$$F(z_{t-j}, \gamma, c) = 1 - \exp\left[-\frac{\gamma}{\sigma_z}(z_{t-j} - \mathrm{c})^2\right] \tag{5}$$

with

$$\gamma > 0. \tag{5}$$

This transition function is symmetric (U-shaped) around $z_{t-j} = \mathrm{c}$, so that the two different regimes to distinguish between are: (i) large deviations of the transition variable from its threshold: $\lim_{\mathrm{z_{t-j}}\to\pm\infty} \; F(z_{t-j\mathrm{t-j}}, \gamma, \mathrm{c}) = 1$ and (ii) small deviations of the transition variable from its threshold: $\lim_{\mathrm{z_t-j}\to\mathrm{c}} F(z_{t-j\mathrm{t-j}}, \gamma, \mathrm{c}) = 0$, that is the linear part.

In our case, the ESTR model represents the hypothesis of symmetric hysteresis in exports. Here, both positive and negative deviations of the threshold variable capacity utilization from its average value c matter. This implies that as long as the deviation of the transitional variable from c is small, there would be no or only small substitution effects from domestic demand to exports (band of inaction). However, if the capacity utilization variable is either significantly above or below its average value, we would expect substitution effects.

The main difference between these two forms of non-linear error-correction model refers to different deviations of the transition variable from its threshold value (its mean): the LSTR case distinguishes positive vs. negative deviations and the ESTR model large vs. small deviations from equilibrium. The former will therefore yield asymmetric results around the threshold, and the latter symmetric deviations above or below the threshold.

3. Empirical Results

3.1 Specification tests

The modelling cycle for smooth transition models suggested by Teräsvirta (1994) starts with a test for non-linearity. The null hypothesis of linearity can be expressed as either H_0: $\gamma = 0$, or H_0: $\beta_1 = \beta_2$. However, both γ and β_2 are unidentified under the null hypothesis. Consequently, standard asymptotics cannot be applied because of the existence of nuisance parameters (van Dijk et al. 2002). To overcome this, Teräsvirta (1994) suggests an approximation of the transition function by a third-order Taylor approximation.

Thus, the corresponding LM test for linearity introduced by Luukkonen et al. (1988) can be expressed as:[12]

$$\Delta x_t = \phi_0 + \phi_1 W_t + \phi_2 W_t z_{t-j} + \phi_3 W_t z_{t-j}^{\ 2} + \phi_4 W_t z_{t-j}^{\ 3} + \epsilon_t. \tag{6}$$

Where $W_t = (\Delta dd_t, \Delta dd_{t-1}, \ldots, \Delta dd_{t-p}, \Delta y^*_t, \ldots, \Delta y^*_{t-p}, \Delta r_t, \ldots, \Delta r_{t-p}, \Delta x_{t-1}, \ldots, \Delta x_{t-p}, \hat{\epsilon}_{t-1})$ and $\phi_i = \left(\phi_{i1}, \ldots, \phi_{iq}\right)$ with q equal to the number of regressors (i.e. the number of elements in W_t).

The null hypothesis, which refers to the linear model being adequate, is tested as $H_0: \ \phi_i = 0$ with $i = 2, 3, 4$ against the alternative H_1 where at least one $\phi_i \neq 0$, implying that the higher-order terms are significant (Teräsvirta 1998). The test statistic has an χ^2 distribution with three degrees of freedom.[13] This procedure also enables the choice of an adequate transition variable. In the case of the linearity hypothesis being rejected, a method for choosing the latter lies in computing the test statistic for several transition functions, that is different values of the lag order j, and selecting the configuration for which its value is maximized (van Dijk et al. 2002). Teräsvirta (1994, 1998) has proven that this procedure is adequate.

According to Granger and Teräsvirta (1993), Teräsvirta (1994, 1998), as well as van Dijk, Teräsvirta and Franses (2002), the LM testing procedure described above can also be applied to distinguish between an exponential and a logistic transition function and thus to choose the appropriate specification. If the linearity null has been rejected, Equation (6) is used to test the following hypotheses successively:

$$\begin{aligned} H_{04}: &\ \ \phi_4 = 0, \\ H_{03}: &\ \ \phi_3 = 0 \mid \phi_4 = 0, \\ H_{02}: &\ \ \phi_2 = 0 \mid \phi_3 = \phi_4 = 0. \end{aligned} \tag{7}$$

The decision rule to select the most adequate transition function introduced by Teräsvirta (1994) is as follows. If the rejection of H_{03} is the strongest one in terms of lowest p-value or largest test statistic respectively, the ESTR model should be chosen or otherwise the LSTR model should be preferred.[14] Table 3 displays the empirical realizations of the non-linearity test statistics, while Table 4 provides the test statistic to distinguish between both configurations.

The common procedure behind selecting the lag length of the transition variable in the Teräsvirta testing and modelling cycle intuitively seems to pick the lags for which the chance to observe non-linearity is strongest (Belke et al. 2015). However, this would seem

12 In the case of small samples in combination with a large number of explanatory variables, *F*-versions of the LM test statistics are preferable, as they have better size properties (Granger and Teräsvirta, 1993, Teräsvirta, 1998, and van Dijk et al. 2002).

13 The number of degrees of freedom 3*ps* refers to the number of regressors *p*. Furthermore, the test assumes that all regressors, as well as the transition variable z_t, are stationary and uncorrelated with the error in Equation (4) u_{t+k} (Teräsvirta, 1998).

14 See Granger and Teräsvirta (1993) or Teräsvirta (1994) for details.

Table 3. Teräsvirta test for non-linearity

Country	Test statistic for $j=1$	Test statistic for $j=2$	Test statistic for $j=3$	Test statistic for $j=4$	Test statistic for $j=5$	Test statistic for $j=6$
Spain	372.18	178.31	85.41	920.17	118.78	111.00
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	[0.58]	[0.51]	[0.53]	[0.60]	[0.56]	[0.58]
Portugal	34.50	33.48	108.94	121.89	251.97	1270.97
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
	[0.34]	[0.38]	[0.37]	[0.33]	[0.41]	[0.45]
Italy	105.25	137.53	55.13	79.38	116.32	113.27
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	[0.46]	[0.46]	[0.42]	[0.50]	[0.51]	[0.59]
France	35.016	23.955	20.509	14.832	15.798	7.532
	(0.002)	(0.014)	(0.042)	(0.192)	(0.111)	(0.755)
	[0.39]	[0.41]	[0.38]	[0.39]	[0.39]	[0.39]
Ireland	188.90	249.53	182.05	204.51	100.73	89.36
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	[0.65]	[0.64]	[0.65]	[0.68]	[0.64]	[0.60]
Greece	1764.02	1619.83	146.17	97.69	137.47	180.74
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	[0.51]	[0.58]	[0.49]	[0.49]	[0.51]	[0.47]

Notes: The test statistic has asymptotic χ^2-distribution with $3\,m$ degrees of freedom under the null hypothesis ($m =$ number of regressors). The table shows the values of the test statistic and p-values in parentheses and $\bar{R}^2$ in brackets. The tests are conducted based on specifications with a common break point in 1993, fourth quarter. j denotes the lag length.

Table 4. Test for the appropriate specification

Country	Lags	(ii)	(iii)	(iv)
Spain	4	48.32	47.97	43.52
		(0.000)	(0.000)	(0.000)
Portugal	6	47.66	5.89	18.02
		(0.000)	(0.435)	(0.006)
Italy	6	47.11	28.36	8.29
		(0.000)	(0.001)	(0.405)
France	1	12.20	11.76	5.53
		(0.032)	(0.038)	(0.355)
Ireland	4	50.42	16.70	32.79
		(0.000)	(0.054)	(0.000)
Greece	2	72.42	54.98	70.47
		(0.000)	(0.000)	(0.000)

Notes: For the Teräsvirta test, χ^2 test statistic realizations are displayed with p-values in parentheses. The test is based on a long-run specification with a common break point. The choice is made for an individual lag length for each country. As mentioned in the text, however, the lag length is unified by us for the following estimations and both LSTR and ESTR are estimated. (ii), (iii), and (iv) refer to H_{02}, H_{03} and H_{04} in Equation (7), respectively.

to artificially favour our prior which is to find non-linearity in the data.[15] To react fully as possible to this important caveat, we provide findings from Table 3 on where a common lag order is just imposed for all countries to make the results more comparable.

The findings in Table 3 show that non-linearity is essentially never rejected for all lag orders. Table 4 shows that a distinction between both model configurations turns out to be difficult. Teräsvirta (1998) suggests estimating different models and choosing between the different specifications and different lag lengths only during evaluation of the estimation results. LSTR and ESTR models generally form very close substitutes. Tests as the ones above should thus be seen as a starting point for estimation instead of providing clear-cut outcome at this early stage of analysis. Taking the ambiguous findings into account, we therefore estimate both LSTR and ESTR models for all countries. To allow for a direct comparison of our findings, we always use a unified lag length of 2 for our transition function in our study.

3.2 Estimation

To evaluate our parameters, we estimate Equation (2) with non-linear least squares (NLS). Our main coefficient of interest β depends on the transition function $\mathrm{F}(\mathrm{z}_{\mathrm{t-j}}, \gamma, \mathrm{c})$ as depicted in either Equation (4) or (5). To choose the final specifications, we examine our estimation results by simple judgment regarding the plausibility of the parameter values and the regimes which the models imply, the models' convergence properties, goodness-of-fit measures, and a test of no residual autocorrelation. For this misspecification test, we apply a variant of the Breusch–Godfrey LM (BG) test suitable for non-linear estimation as suggested in Teräsvirta (1998). The test's null hypothesis is that there is no pth order serial correlation in our residuals u_t. The test regresses $\sim u_t$ (the estimated residuals) on $\sim u_{t-1}, \ldots, \sim u_{t-p}$ and the partial derivatives of the regression function with respect to γ.

Estimation results are found in Table 5 for countries with an ESTR specification and in Table 6 for countries with an LSTR specification. Our theoretical priors suggest a negative coefficient for the coefficient β, that is, a substitution effect from domestic demand to exports during times of low or high capacity utilization. When estimating the ESTR model, coefficient β_{1i} for $F(z_{t-j}, \gamma, \mathrm{c}) = 0$ (i.e. the linear model) shows us results for capacity utilization levels around the threshold level. The joint coefficient $\beta_{1i}+\beta_{2i}$ for the case when $F(z_{t-j\mathrm{t-j}}, \gamma, \mathrm{c}) = 1$ yields the results for positive and negative deviations from our threshold. In the LSTR case, β_{1i} represents low levels of capacity utilization, and $\beta_{1i}+\beta_{2i}$ high values of capacity utilization.

3.3 Estimation results

Let us first turn to the econometric specification based on an ESTR model (Table 5). For France, Greece, and Ireland, the effects for 1 or 2 lags display negative values for extreme levels of past capacity utilization, while negative contemporaneous effects are not identified except for the case of Italy. The contemporaneous coefficient is positive for Ireland; for

15 We owe this caveat to an anonymous referee. The results for potentially different country-specific lag lengths are contained in the previous version of this article and are available on request.

Table 5. Estimation results for domestic demand effects on exports based on ESTR specification

Contemporaneous coefficients		Spain	Portugal	Italy	France	Greece	Ireland
Domestic demand first regime (β_{10})		0.7377	−0.1910	1.0906***	−0.3952	−0.0258	0.4505***
		(1.6703)	(−1.3809)	(4.5453)	(−1.2791)	(−0.3100)	(3.9030)
Domestic demand second regime (β_{20})		−0.4091	0.3087**	−0.9441***	0.9451	−0.2483	3.6635
		(−0.3879)	(2.6784)	(−3.1921)	(1.5624)	(−1.0574)	(0.3619)
Sum	$\beta_{10}+\beta_{20}$	0.3286	0.1177	0.1465	0.5499	−0.2741	4.1140
Lagged coefficients with 1 lag							
Domestic demand first regime (β_{11})		−0.7260	−0.5166***	−1.8717**	2.4935**	0.5064**	−0.0039
		(−1.4029)	(−5.4342)	(−2.9751)	(2.0281)	(2.6547)	(−0.0767)
Domestic demand second regime (β_{21})		1.3413**	1.0025***	2.4579***	−2.2968	−0.7399**	−0.8538**
		(2.9055)	(8.1215)	(4.4976)	(1.5967)	(−2.6968)	(−3.0011)
Sum	$\beta_{11}+\beta_{21}$	0.6153	1.5191	0.586	0.197	−0.2034	−0.8577
Lagged coefficients with 2 lags							
Domestic demand first regime (β_{12})		0.0018	0.0819	0.4710	1.4331**	0.6935***	−0.0359
		(0.0051)	(1.2300)	(1.1663)	(2.1650)	(6.7361)	(−0.3502)
Domestic demand second regime (β_{22})		−0.0572	0.4357***	−0.7331**	−1.7396**	0.0641	0.6747**
		(−0.5623)	(5.1060)	(−2.1199)	(−3.0831)	(0.1176)	(2.5383)
Sum	$\beta_{12}+\beta_{22}$	−0.0554	0.5176	−0.2621	−0.3065	0.7576	0.6388
	γ cont	1.0319	2.8357**	5.0390**	2.3023**	2.5524***	0.0205
		(1.5611)	(2.0973)	(2.6401)	(2.0015)	(4.8027)	(0.3239)
	γ 1 lag	3.2174	2.8703**	13.9106***	36.0609	2.5483***	0.1179
		(0.9334)	(2.1083)	(3.9557)	(0.2189)	(5.2896)	(1.4186)
	γ 2 lags	−0.4824	2.4393**	4.8830	2.8426***	3.6794***	8.4277***
		(−1.5965)	(2.2700)	(1.2392)	(3.7361)	(4.9411)	(4.3715)

Notes: Coefficients estimated by NLS; Newey–West standard errors in parentheses. */**/*** statistical significance at the 10/5/1% level. For the joint significance of β_{1i} and β_{2i}, the linear restriction $\beta_{1i}+\beta_{2i}=0$ has been tested with χ^2 test statistics; p-value in brackets. The BG LM test is based on the null hypothesis of no serial correlation of the residuals of order $p=4$. β_{ji} $(j=1,2)$ is the coefficient for domestic demand in the non-linear error correction model. The two extreme regimes are $F(z_{t-j},\gamma,c)=0$ given by β_{1i} (i.e. for the ESTR model around the threshold value) and $F(z_{t-j},\gamma,c)=1$ given by $\beta_{1i}+\beta_{2i}$ (i.e. for the ESTR model for large deviations from threshold). Numbers of observations: Italy (125), Spain (100), Greece (95), Portugal (104), France (124), and Ireland (62).

Table 6. Estimation results for domestic demand effects on exports based on LSTR specification

		Spain	Portugal	Italy	France	Greece	Ireland
Contemporaneous coefficients							
Domestic demand first regime (β_{10})		0.5842	−0.0493	0.1413	0.5018***	−0.2646	0.5059**
		(1.0028)	(−0.3718)	(0.5148)	(4.0670)	(−1.1375)	(2.6860)
Domestic demand second regime (β_{20})		−0.2537	−0.4210***	1.1692**	−0.8151**	0.3259	0.0707
		(−0.3407)	(−3.2759)	(2.4850)	(−1.9890)	(0.4689)	(0.1721)
Sum	$\beta_{10}+\beta_{20}$	0.3305	−0.4703	1.3105	−0.3133	0.0613	0.5766
Lagged coefficients with 1 lag							
Domestic demand first regime (β_{11})		0.6569***	0.1893**	−1.7207***	−0.2950**	−0.1988	
		(4.4806)	(2.0073)	(−4.3676)	(−2.0709)	(−1.7014)	
Domestic demand second regime (β_{21})		−0.3124**	−0.2388	3.9681***	1.2857***	0.9990**	
		(−1.9651)	(−1.3240)	(4.4281)	(3.7642)	(2.4368)	
Sum	$\beta_{11}+\beta_{21}$	0.3445	−0.0495	2.2474	0.9907	0.8002	
Lagged coefficients with 2 lags							
Domestic demand first regime (β_{12})		−0.4857	0.0143	0.0291	0.8431	0.8653*	
		(−1.6104)	(0.0786)	(0.1011)	(1.5604)	(1.8661)	
Domestic demand second regime (β_{22})		1.2434***	0.5524**	0.7532	−1.1847*	−0.1405	
		(4.5419)	(2.3334)	(0.8090)	(−1.9363)	(−0.2531)	
Sum	$\beta_{12}+\beta_{22}$	0.7577	0.5667	0.7823	−0.3416	0.7248	
	γ cont	2.2329	4.1112**	4.0047**	36.0609	2.0200***	0.5575
		(1.7211)	(2.5128)	(2.9547)	(0.2189)	(3.3607)	(1.1220)
	γ 1 lag	9.6231	4.1054**	0.8425***	2.2374	2.1150***	
		(1.6827)	(2.6452)	(3.2729)	(0.9153)	(8.2651)	
	γ 2 lags	5.4792**	2.4577	1.7537***	7.2044	11.1546***	
		(2.2027)	(1.3123)	(6.0526)	(0.6369)	(6.4563)	

Notes: Coefficients estimated by NLS; Newey–West standard errors in parentheses. */**/*** statistical significance at the 10%/5%/1% level. For the joint significance of β_{1i} and β_{2i}, the linear restriction $\beta_{1i}+\beta_{2i}=0$ has been tested with χ^2 test statistics; p-value in brackets. The BG LM test is based on the null hypothesis of no serial correlation of the residuals of order $p=4$.

β_{ji} $(j=1,2)$ is the coefficient for domestic demand in the non-linear error-correction model. The two extreme regimes are $F(z_{t-j},\gamma,c)=0$ given by β_{1i} (i.e. for the ESTR model around the threshold value) and $F(z_{t-j},\gamma,c)=1$ given by $\beta_{1i}+\beta_{2i}$ (i.e. for the ESTR model for large deviations from threshold). Numbers of observations: Italy (125), Spain (100), Greece (95), Portugal (104), France (124), and Ireland (64).

Italy the results are ambiguous. This suggests a substitutive relationship between domestic and foreign sales when the economy is close to peak or trough. When capacity utilization is very low, firms react to a fall in domestic demand by increasing their efforts to export. Conversely, if the economy operates at high capacity utilization, capacity constraints imply that an increase in domestic demand triggers a reallocation of resources from external to domestic clients. A positive coefficient may imply that the short-run liquidity channel dominates, whereby the cash flow generated by exports is used to finance domestic operations, and the existence of increasing returns dominates the capacity constraints channel (Belke et al. 2015, and Berman et al. 2011). As argued above, also this general pattern is in line with the prevalence of hysteresis and the band of inaction due to switching costs for suppliers between serving the domestic and foreign market.

We now turn to our findings based on the LSTR specification (Table 6). The contemporaneous substitution coefficient is positive for France and Ireland but insignificant for the other countries in the first regime (beta 0, business cycle trough). For Portugal and France, the substitution coefficient becomes negative in a boom (which is reflected by the sum of both coefficients). While there is hardly any significance of the coefficients for a lag of two quarters, we find a positive coefficient for Spain and Portugal in case of negative capacity utilization (trough) and a negative one for Italy. However, the sum of both coefficients becomes positive for Italy, France, and Greece in case of positive capacity utilization (boom).

Overall, our empirical results presented in Tables 5 and 6 suggest that the relationship between domestic sales and exports depends on capacity utilization and the business cycle. Evidence of a substitutive relationship between domestic and foreign sales varies among countries and with different lag lengths. The findings are broadly in line with the gain in export market shares in several euro area (debt and banking) crisis countries during the subsequent recession. There is more diversity across countries during other stages of the business cycle suggesting that capacity constraints and the liquidity channel play a different role across countries and/or partly cancel each other out.

Seen on the whole, thus, more analytical rigour by imposing a common break point in all country-specific estimations and tests and common lags for the transition variable comes at a 'cost' in terms of less adaptation of country specifics and thus economically plausible results.[16]

3.4 Robustness check: taking stock of political uncertainty

In the following, we are reporting the results of an important robustness check of our estimations.[17] As a final check, we included economic policy uncertainty in our analysis, a variable playing a prominent role in explaining band of inactions in the reaction of exports

16 More evidence in favour of 'substitutability' for at least four countries in our sample (Spain, Portugal, Ireland, and Greece) is gained if one admits different break points for different countries. The results are available on request.

17 Belke et al. (2015) present results in a framework comparable to ours but using export goods (a measure which is more closely related to capacity utilization but is unfortunately not available for the sample period used by us) only and yield similar results. We also performed additional robustness tests by using different types of real effective exchange rates (deflated by unit labour costs and deflated by consumer price indices) and using the median instead of the mean value as threshold. The results are available upon request.

Table 7. Estimation results for domestic demand effects on exports with uncertainty—LSTR specification

Country	Spain	Portugal	Italy	France	Ireland	Greece
Lagged coefficients with 1 lag						
Domestic demand	0.1170	0.4589*	0.9710***	1.0858		−0.0183
first regime (β_0)	(0.5968)	(2.4077)	(5.8344)	(1.9293)		(−0.3000)
Domestic demand	0.4703*	−0.2768	−1.5012***	−0.6302		−0.4811***
second regime (β_{01})	(1.8622)	(−0.5282)	(−3.9177)	(−1.0603)		(−5.2344)
Lagged coefficients with 2 lags						
Domestic demand	0.8381***	0.7809	0.0475	0.9144**	0.0063	0.8299**
first regime (β_0)	(8.0184)	(7.2030)	(0.1425)	(2.616)	(0.0324)	(4.6914)
Domestic demand	0.0809	−0.61220**	0.6130	0.5287	0.0345	0.0091
second regime (β_{01})	(0.5704)	(−4.0477)	(1.5312)	(0.9899)	(0.1491)	(0.0556)

Notes: Coefficients estimated by NLS; Newey–West standard errors in parentheses. */**/*** statistical significance at the 10/5/1% level. For the joint significance of β_{1i} and β_{2i}, the linear restriction $\beta_{1i} + \beta_{2i} = 0$ has been tested with χ^2 test statistics; p-value in brackets. The BG LM test is based on the null hypothesis of no serial correlation of the residuals of order $p = 4$.

β_{ji} $(j = 1, 2)$ is the coefficient for domestic demand in the non-linear error-correction model. The two extreme regimes are $F(z_{t-j}, \gamma, c) = 0$ given by β_{1i} (i.e. for the ESTR model around the threshold value) and $F(z_{t-j}, \gamma, c) = 1$ given by $\beta_{1i}+\beta_{2i}$ (i.e. for the ESTR model for large deviations from threshold). Numbers of observations: Italy (61), Spain (61), Greece (61), Portugal (61), France (62), and Ireland (64).

to its main drivers (see, for instance, Belke and Kronen 2016). More precisely, we rely on the change of European policy uncertainty in period $t-1$ (i.e. lagged one quarter) as a transition variable in Equation (3). As derived in Section 1, the empirical results of the hysteretic export equation may turn out to be more pronounced because the band of inaction gets larger with increasing uncertainty. The findings for the LSTR model are given in Table 7.

For the following interpretation of the estimation results, we have to keep in mind that the first coefficient holds if economic policy uncertainty decreases, while the sum of both coefficients is relevant for an increase in economic policy uncertainty. This interpretation is based on Equation (3) for both values of the transition function. The evolution of the transition function reflects the band of inaction. The transition function increases from 0 to 1 if economic policy uncertainty increases. The second coefficient in Table 7 always provides the additional effect once the transition function increases from 0 to 1. Hence, the overall effect for the highest increase in uncertainty is given by the sum of both coefficients which reflects the case where the transition function is 1. We focus on two potential effects in the following: the effect on the substitution coefficient and the one on the global demand coefficient. In each case, we analyse the impact with a delay of the regressor of 1 or 2 lags. Due to the rich number of models estimated, several coefficients are as usual insignificant. Nevertheless, a few important results are worth mentioning.

The original effect of domestic sales on exports is always either positive or insignificant in the regime with a decrease in uncertainty. However, the effect for an increase in uncertainty (measured by the sum of both coefficients) becomes negative for Italy (1 lag) and Greece (1 lag) and is strongly reduced for Portugal (2 lags). This points to a substitution effect as a result of higher uncertainty. Interestingly, the sum of both coefficients becomes positive for Spain. These results broadly confirm robustness of the results gained before with respect to the inclusion of a formerly omitted variable 'policy uncertainty'. In other words, there is no omitted variable bias in our case. On the contrary, our model of export hysteresis presented in Section 1 is corroborated for some EMU member countries such as Italy, Greece and, a bit less so, also for Portugal. The 'non-case' of Ireland may be explained by the higher flexibility of the Irish economy compared to its Southern European counterparts. Flexible prices and immigration may have made capacity constraints less binding (see Belke et al. 2015).

For information only, differences between the two regimes are also observed with regard to the effect of world demand on exports. The effect of world demand on exports of France and Greece is negative in case of an increase in policy uncertainty. The opposite is observed for Portugal and Spain, where the effects on exports increase in case of higher uncertainty. While no effect is observed for Italy, the specific effect on Greece depends on the lag order, but the sum of coefficients for higher uncertainty is always significant. A useful extension for further research will be the consideration of country-specific economic policy uncertainty indices. However, these are not available over the full sample (see www.policyuncertainty.com).

4. Conclusions

In this article, we have analysed the relation between domestic demand and exports for six euro area countries using non-linear smooth transition estimations faced with a strong a priori restriction of common break points and common lags across individual country specifications. To illustrate the results gained in this article, it seems worthwhile to contrast them with those identified by us without these a priori restrictions.

Our empirical results based on individual and potentially different break point specifications which have been gained in a previous version of this article (available on request) clearly indicated that domestic demand behaviour is relevant for the short-run dynamics of several euro area member countries' exports. The estimation results suggested that on an aggregated level, contemporary and lagged domestic demand developments can affect a country's export performance significantly.

We found that in the cases of Spain, Portugal, and Italy, the symmetric non-linearity of the relation manifests itself in a contemporary substitutive relationship between domestic demand and export activity if deviations from average capacity utilization are large. This is somewhat independent of their sign, but we found stronger evidence for notably low levels of capacity utilization. In other words, the substitution effect from domestic demand to exports turns out to be stronger and more significant during more extreme stages of the business cycle. During periods of more average levels of capacity utilization, our empirical evidence pointed to a band of inaction in which the relation between domestic and foreign sales is complementary. On a micro level, theoretical reasons for these findings can be found in the sunk costs hysteresis approach. For France, the evidence for non-linearity was weaker. We found evidence of mostly complementary relationships. In the cases of Ireland and Greece, we detected an asymmetric non-linear relationship between domestic demand and exports. Domestic demand and exports are slightly substitutive during a business cycle trough and complements during normal times and in a boom.

Overall, our results mostly confirmed the short-run non-linear relationship between domestic and foreign sales depending on capacity constraints. A substitutive relationship with low capacity utilization shows up most clearly for Spain, Portugal, and Italy.

We also provided first ideas for why we believe there are valid reasons for the different findings in the other countries (such as the high number of multinational corporations in Ireland, the lower openness of the French economy, or the small Greek tradable sector). However, deriving more detailed explanations for these heterogeneous results for some countries in our sample provides an interesting area for future research. A further interesting avenue could lead to a more disaggregated, sectoral analysis to understand the underlying firm behaviour in more detail (Esteves and Prades 2017).

A final interesting avenue was taken in this article: we conducted all necessary estimations and tests based on a common break point implementation not to bias the results into the direction of the empirical model with the highest degree of non-linearity and on common lags for all countries to avoid the impression that the country-specific regression models were over-fitted till significance. The pattern of the results changed as follows. A substitutive relationship between domestic and foreign sales can now most clearly be found for France, Greece, and Ireland (ESTR model) and France, Portugal, and Italy (LSTR model), providing evidence of the importance of sunk costs and hysteresis in international trade in these EMU member countries.

What is more, our empirical results are robust to the inclusion of a variable measuring European policy uncertainty. In some cases (Italy, Greece, and Portugal) the results underscore the empirical validity of the export hysteresis under uncertainty model. While we do not feel legitimized to go more deeply into economic interpretations of the country-specific results due to the pronounced ex ante restrictions such as the imposition of common break points and of common lags across all country-specific empirical models, we would like to stress the finding of a general non-linear pattern of export activity of the Euro area member countries with a remarkable goodness of fit.

Seen on the whole, the macroeconomic perspective is able to offer insights on overall adjustment effects for euro area countries with previous imbalances. In recent years, the six countries under consideration which recorded large current account deficits before the European debt and banking crisis starting in 2010 have seen a significant correction of their external imbalances. This holds in particular for their trade balances, and exports have been a key adjustment factor. Our results provide one explanation for the rising exports besides standard competitiveness arguments; the observed increase in export market shares accompanying the reduction of the current account deficits could be due to non-price related factors, such a low domestic demand leading to survival-driven exports, instead of an increase in price competitiveness as expected by sustainable structural reforms. This argument appears to be especially relevant in the current period for the countries under consideration in which their capacities have been utilized only to a low degree and domestic demand has fallen strongly. Low domestic demand then did not only affect imports but at the same time exports and has thus strongly contributed to the external adjustment.

Regarding policy implications, our findings provide important insights for the discussion of macroeconomic adjustment and the reduction of imbalances in the euro area. Prima facie, our results for specific countries could suggest that domestic demand and exports are negatively related only in the short run, triggered by current economic conditions. To the extent that the closure of the output gap is driven by a pickup in domestic demand, a lot of the gains in export market shares of vulnerable euro area countries could be lost in the long run. In such a scenario, analyses of cyclically adjusted current account balances could possibly overestimate the structural adjustment to the degree that weak domestic economic conditions impact not only the import side of the net trade equation but also the export side.

On the other hand, at least three factors give rise to the hope that the gains in export market performance may be of a more long-run nature. First, domestic demand conditions in peripheral economies are likely to remain depressed as long as the debt burden of both private and public sector remains high. An extended period of deleveraging pressure increases the chances that the reallocation of resources to the export sector will also be more permanent, possibly also fostering increased export-oriented foreign direct investment into distribution networks and other hedging activities (Belke et al. 2013). This would make the hypothesized substitutive relationship between domestic demand and exports more long run. Secondly, our sunk cost hysteresis model suggests that once domestic producers have paid sunk costs for shifting production to exports, they remain in a band of inaction even as the business cycle improves. Reversing export market participation should not be expected as long as there are capacities for serving both domestic and foreign market. Thirdly, with increasing exports today and a pickup in domestic demand in the future, a complementary relation between domestic sales and exports might develop in the long run due to improvements in efficiency encouraged by learning-by-doing effects. In conclusion, the export increase could therefore be lasting and a substantial part of the gains in export market shares may not only a cyclical phenomenon, but indeed be of a more structural nature.

Acknowledgments

The authors would like to thank Gabriel Felbermayr and two anonymous referees for valuable comments and above all Joscha Beckmann but also Stéphane Dees, Philippe de Rougemont, Frauke Skudelny, Florian Verheyen, participants in the IMF research department’s brown bag seminar, the ECB’s DED seminar, and the CESifo-Delphi conference on ‘Current Account Adjustments’ for valuable comments and suggestions. The authors thank Christian Buelens for the provision of data on value-added exports.

References

Ahn, J. and A. F. McQuoid (2013), “Capacity Constrained Exporters: Micro Evidence and Macro Implications”, Department of Economics Working Papers 1301, Florida International University, Miami, FL.

Artus, J. R. (1970), “The Short-term Effects of Domestic Demand Pressure on British Exports Performance”, *International Monetary Fund Staff Papers* 17, 247–274.

Artus, J. R. (1973), “The Short-Run Effects on Domestic Demand Pressure on Export Delivery Delays for Machinery”, *Journal of International Economics* 3, 21–36.

Bai, J. and P. Perron (1998), “Estimating and Testing Linear Models with Multiple Structural Changes”, *Econometrica* 66, 47–78.

Bai, J. and P. Perron (2003), “Computation and Analysis of Multiple Structural Change Models”, *Journal of Applied Econometrics* 18, 1–22.

Baldwin, R. and P. Krugman (1989), “Persistent Trade Effects of Large Exchange Rate Shocks”, *Quarterly Journal of Economics* 104, 635–54.

Ball, R. J., J. R. Eaton, and M. D. Steuer (1966), “The Relationship between United Kingdom Export Performance in Manufactures and the Internal Pressure of Demand”, *Economic Journal* 76, 501–518.

Belke, A. and M. Goecke (2001), “Exchange Rate Uncertainty and Employment: An Algorithm Describing ‘Play”, *Applied Stochastic Models in Business and Industry* 17, 181–204.

Belke, A. and M. Goecke (2005), “Real Options Effects on Employment: Does Exchange Rate Uncertainty Matter for Aggregation?”, *German Economic Review* 6, 185–203.

Belke, A. and D. Gros (2017), “Optimal Adjustment Paths in a Monetary Union”, *Economic Modelling*, http://www.sciencedirect.com/science/article/pii/S0264999317302559.

Belke, A. and D. Kronen (2016), “Exchange Rate Bands of Inaction and Play-hysteresis in Greek Exports to the Euro Area, the US and Turkey: Sectoral Evidence”, *Empirica* 43, 349–90.

Belke, A., M. Goecke, and M. Guenther (2013), “Exchange Rate Bands of Inaction and Play-Hysteresis in German Exports – Sectoral Evidence for Some OECD Destinations”, *Metroeconomica* 64, 152–79.

Belke, A., A. Oeking, and R. Setzer (2015), “Domestic Demand, Capacity Constraints and Exporting Dynamics: Empirical Evidence for Vulnerable Euro Area Countries”, *Economic Modelling* 48, 315–25.

Berman, N., A. Berthou, and J. Héricourt (2011), “Export Dynamics and Sales at Home”, CEPR Discussion Paper No. 8684, Centre for Economic Policy Research, London.

Blum, B., S. Claro, and I. Horstmann (2011), *Occasional vs Perennial Exporters: The Impact of Capacity on Export Mode*, *Journal of International Economics* 90, 65–74.

Christodoulopoulou, S. and O. Tkacevs (2014), “Measuring the Effectiveness of Cost and Price Competitiveness in External Rebalancing of Euro Area Countries: What Do Alternative HCIs Tell Us?”, ECB Working Paper 1736, European Central Bank, Frankfurt.

Dixit, A. (1989), “Entry and Exit Decisions under Uncertainty”, *Journal of Political Economy* 97, 620–38.

Dunlevy, J. A. (1980), “A Test of the Capacity Pressure Hypothesis within a Simultaneous Equations Model of Export Performance”, *Review of Economics and Statistics* 62, 131–135.

Engle, R. F. and C. W. J. Granger. (1987), “Co-integration and Error Correction: Representation, Estimation and Testing”, *Econometrica* 55, 251–276.

Esteves, P. S. and E. Prades (2017), “On Domestic Demand and Export Performance in the Euro Area Countries: Does Export Concentration Matter?”, ECB Working Paper 1901, European Central Bank, Frankfurt.

Esteves, P. S. and A. Rua (2013), “Is There a Role for Domestic Demand Pressure on Export Performance?”, ECB Working Paper 1594, European Central Bank, Frankfurt.

European Commission (2011), “Export Demand Equations for Euro Area Countries”, *Quarterly Report on the Euro Area*, Box IV.1, European Central Bank, Frankfurt.

Faini, R. (1994), “Export Supply, Capacity and Relative Prices”, *Journal of Development Economics* 45, 81–100.

Ghironi, F. and M. Melitz (2005), “International Trade and Macroeconomic Dynamics with Heterogeneous Firms”, *Quarterly Journal of Economics* 120, 865–915.

Granger, C. W. J. and T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships* Advanced Texts in Econometrics, Oxford University Press, New York.

Ilmakunnas, P. and S. Nurmi (2007), “Dynamics of Export Market Entry and Exit”, Helsinki Institute of Economic Research Discussion Paper 197, Helsinki, November.

Krugman, P. R., M. Obstfeld, and M. J. Melitz (2015), *International Economics – Theory and Policy*, 10th ed., Pearson.

Lee, J. and M. C. Strazicich (2003), “Minimum LM Unit Root Test with Two Structural Breaks”, *Review of Economics and Statistics* 85, 1082–9.

Lee, J. and M. C. Strazicich (2004), “Minimum LM Unit Root Test with One Structural Break”, Appalachian State University Working Paper, University of Alabama, Alabama.

Luukkonen, R., P. Saikkonen, and T. Terasvirta (1988), “Testing Linearity against Smooth Transition Autoregression”, *Biometrika* 75, 491–499.

MacKinnon, J. (1991), "Critical Values for Cointegration Tests", in: R. F. Engle and C. W. J. Granger, eds: *Long-Run Economic Relationships*, London, Oxford University Press, pp. 267–276.

Máñez, J. A., M. E. Rochina, and J. A. Sanchis Llopis (2008), "Sunk Cost Hysteresis in Spanish Manufacturing Exports", *Review of World Economics* 144, 272–94.

Pindyck, R. S. (1988), "Irreversible Investment, Capacity Choice and Uncertainty", *American Economic Review* 78, 969–85.

Pindyck, R. S. (1991), "Irreversibility, Uncertainty, and Investment", *Journal of Economic Literature* 29, 1110–48.

Sharma, K. (2003), "Factors Determining India's Export Performance", *Journal of Asian Economics* 14, 435–46.

Smyth, D. J. (1968), "Stop-Go and United Kingdom Exports of Manufactures", *Oxford Bulletin of Economics and Statistics* 30, 25–36.

Teräsvirta, T. (1994), "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models", *Journal of the American Statistical Association* 89, 208–18.

Teräsvirta, T. (1998), "Modelling Economic Relationships with Smooth Transition Regressions", D. E. A. Giles and A. Ullah, eds, *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, pp. 507–552.

van Dijk, D., T. Teräsvirta and P. H. Franses (2002), "Smooth Transition Autoregressive Models—A Survey of Recent Developments", *Econometric Reviews* 21, 1–47.

Vannoorenberghe, G. (2012), "Firm-level Volatility and Exports", *Journal of International Economics* 86, 57–67.

Zilberfarb, B.-Z. (1980), "Domestic Demand Pressure, Relative Prices and the Exports Supply Equation - More Empirical Evidence", *Economica* 47, 443–50.

Appendix

Table A1. Data sources

Series	Source	Definition	Time periods available
Exports	National Statistical Offices	Real exports of goods and services (in prices of reference year)	• 1980Q1–2012Q4 • IT: 1981Q1–2012Q4
Exports (value added)	World Input-Output Database (interpolated)	Value-added exports (converted to prices of reference year)	1995Q1–2011Q1
Domestic demand	National Statistical Offices	Real domestic demand (in prices of reference year)	• 1980Q1–2012Q4 • IT: 1981Q1–2012Q4
Real effective exchange rate (CPI)	Eurostat	Index deflated by consumer price indices with a country's 15 main trading partners	1980Q1–2012Q4
Real effective exchange rate (ULC)	Eurostat	Index deflated by unit labour costs with a country's 24 main trading partners	1980Q1–2012Q4
Foreign demand	ECB	Trade-weighted imports for 15 main trading partners	1980Q1–2012Q4
Capacity utilization	Eurostat	Current level of capacity utilization in manufacturing industry based on business surveys	• PT: 1987Q1–2012Q4 • IT, GR: 1985Q1–2012Q4 • ES: 1987Q2–2012Q4
Capacity utilization	Insee	Capacity utilization rate based on quarterly business survey	FR: 1980Q1–2012Q4
Output gap	Federal Reserve Board	Gap between actual gross domestic product (GDP) and potential GDP as percentage of potential GDP	• IE: 1980Q1–2012Q4 • FR: 1980Q1–2012Q4
Policy uncertainty	www.policyuncertainty.com	Newspaper-based uncertainty index	1987Q1–2012Q4

Table A2. Unit root tests

Country	Series	ADF test		Lee–Strazicich test	
		Level *t*-statistics [lags]	First difference *t*-statistics [lags]	One break *t*-statistics	Two breaks *t*-statistics
Spain	dd_t	−1.054 [3]	−2.111** [2]	−0.6281	−0.6370
	x_t	−1.275 [0]	−10.565*** [0]	−1.7927	−2.0560
	x_t^{goods}	−1.875 [0]	−12.457*** [0]	−2.4443	−2.9754
	x_t^{va}	−2.407 [8]	−2.093** [10]	−0.7349	−0.7597
	y_t^*	−3.418* [1]	−4.569*** [0]	−1.9472	−2.0878
	r_t	−1.250 [1]	−8.763*** [0]	−1.8106	−1.9323
	r_t^{ULC}	−1.373 [1]	−7.905*** [0]	−1.0327	−1.0664
Portugal	dd_t	−0.199 [3]	−3.017*** [2]	−0.5972	−0.6117
	x_t	−0.731 [0]	−7.321*** [0]	−1.4594	−1.5466
	x_t^{goods}	−1.967 [4]	−3.257*** [3]	−2.6350	−2.9542
	x_t^{va}	−0.750 [8]	−1.843* [3]	−1.1552	−1.1895
	y_t^*	−2.742 [1]	−4.400*** [0]	−1.6444	−1.7162
	r_t	−1.353 [1]	−8.784*** [0]	−2.4693	−2.5850
	r_t^{ULC}	−0.917 [1]	−6.849*** [0]	−1.0068	−1.0402
Italy	dd_t	−0.153 [2]	−3.637*** [1]	−0.7875	−0.8090
	x_t	−1.318 [0]	−5.907*** [1]	−2.0700	−2.3491
	x_t^{goods}	−3.906** [2]	−8.076*** [0]	−2.5597	−2.9079
	x_t^{va}	−3.251* [7]	−2.585** [7]	−1.4249	−1.4481
	y_t^*	−2.944 [2]	−4.750*** [1]	−2.0089	−2.1816
	r_t	−2.501 [1]	−8.336*** [0]	−1.8317	−1.9321
	r_t^{ULC}	−2.279 [1]	−7.685*** [0]	−1.6470	−1.7732
France	dd_t	−1.692 [2]	−2.659***[1]	−0.9772	−1.0018
	x_t	−1.160 [1]	−4.640*** [1]	−1.0702	−1.1443
	x_t^{goods}	−2.297 [1]	−7.339*** [0]	−1.2483	−1.3156
	x_t^{va}	−1.509 [8]	−1.842* [7]	−0.7760	−0.8076
	y_t^*	−3.268* [1]	−4.703*** [0]	−2.0007	−2.0854
	r_t	−1.921 [0]	−10.654*** [0]	−2.6688	−2.7981
	r_t^{ULC}	−3.129* [1]	−8.750*** [0]	−1.5954	−1.6572
Ireland	dd_t	−1.650 [3]	−2.805*** [2]	−0.6024	−0.6188
	x_t	−0.764 [4]	−1.401 [6]	−1.1048	−1.1648
	x_t^{goods}	−1.273 [4]	−4.099*** [3]	−1.3362	−1.4306
	x_t^{va}	−2.308 [8]	−2.059** [7]	−0.5018	−0.5126
	y_t^*	−2.580 [2]	−5.141*** [1]	−1.8182	−1.9890
	r_t	−1.837 [0]	−9.162*** [0]	−1.8346	−1.9568
	r_t^{ULC}	−1.896 [1]	−7.549*** [0]	−1.2778	−1.3429
Greece	dd_t	−0.109 [5]	−2.906*** [4]	−1.1719	−1.2182
	x_t	−1.734 [4]	−5.125*** [3]	−2.4917	−2.8454
	x_t^{goods}	−3.015 [4]	−5.130*** [3]	−4.1321**	−4.8821***
	x_t^{va}	−1.232 [8]	−1.271 [6]	−0.8985	−0.9393
	y_t^*	−3.646** [1]	−4.249*** [0]	−1.8027	−1.9790
	r_t	−0.810 [0]	−12.329*** [0]	−3.5230*	−3.8786**
	r_t^{ULC}	−2.029 [1]	−9.804*** [0]	−1.9257	−2.0192

Notes: ADF test: the lag length is chosen by minimizing the Schwarz Information Criterion with a prior defined maximum lag length of 12. Critical values for an intercept: 1%: −3.43, 5%: −2.86, and 10%: −2.57. Critical values for both an intercept and a time trend: 1%: −3.96, 5%: −3.41, and 10%: −3.13. Critical values without deterministic trends (for first differences): 1%: −2.56, 5%: −1.94, and 10%: −1.62. Lee–Strazicich test: critical values with one break: 1%: −4.239, 5%: −3.566, and 10%: −3.211. Critical values with two breaks: 1%: −4.545, 5%: −3.842, and 10%: −3.504. Cf. Lee and Strazicich (2004) and Lee and Strazicich (2003).

CESifo WORKING PAPERS

6503
2017

May 2017

Do Parties Punish MPs for Voting Against the Party Line?

Björn Kauder, Niklas Potrafke, Marina Riem

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6503*
Category 2: Public Choice

Do Parties Punish MPs for Voting Against the Party Line?

Abstract

We examine whether parties punish politicians who vote against the party line in roll-call votes. Using data of German members of parliament over the legislative period 2009-2013, we take into account that the effect of punishment differs along the list of candidates because a candidate is punished more when he loses positions at the threshold of promising list positions. The dataset includes the voting behavior of 257 MPs in 218 roll-call votes. Our results do not show that parties account for the voting behavior by punishing politicians who have voted against the party line. Political parties may attract different groups of voters by tolerating politicians who vote according to their own credo. Qualities other than the voting behavior seem to matter to political parties when nominating candidates.

JEL-Codes: D720, D780, P160.

Keywords: voting against the party line, adherence to the party line, roll-call votes, proportional representation, party lists, selectorate.

Björn Kauder
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
kauder@ifo.de

*Niklas Potrafke**
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
portrafke@ifo.de

Marina Riem
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
riem@ifo.de

*corresponding author

May 2017
This paper has been accepted for publication in CESifo Economic Studies.
We would like to thank Matz Dahlberg, Panu Poutvaara and two anonymous referees for helpful comments and Matthias Hänsel, Sebastian Kropp, Marian Scheffer, and Alexander H. Schwemmer for excellent research assistance.

1. Introduction

Voting decisions depend on how notable a candidate is for the electorate or for the delegates that select a party's candidate for public office (the "selectorate", henceforth "the party"). A candidate can distinguish himself from co-partisans by past performance and effort in office, political experience and even physical attractiveness, but also by defecting from the party line on roll-call votes. The bailout of Greece in the aftermath of the financial crisis which began in 2007 is an excellent case in point. German politicians' views on the issue differed within and across parties. Most party leaderships advocated the rescue packages. Some members of parliament (MPs) did however not toe the party line in roll-call votes on the rescue packages. Because German journalists lean towards a critical view on the European crisis management, and because it was a controversial issue in the public discourse, the media celebrated the MPs that made a martyr of themselves by using roll-call votes as low-cost signaling devices. Newspaper articles quoted how many MPs voted in favor or against the Greek bailout packages or if they abstained from voting, and hyped individual MPs who voted – against the majority of their political party – against the Greek bailout packages.[1]

A first question is what determines defection from the party line on roll-call votes. MPs behave strategically when announcing a position on a roll-call vote because they have the electoral implication of their vote in mind (Mayhew 1974, Bütikofer and Hug 2015).[2] MPs that are more dependent on the party's reputation are less likely to vote against the party line (Thames 2005, Kunicova and Remington 2008, Sieberer 2010).[3] In Germany, MPs with a high expertise in European policy were more likely to vote against the European bailout packages (Wimmel 2013). Directly elected MPs in the 2005-2009 period were more likely to

[1] See, for example, "Griechenland-Abstimmung im Bundestag: So hat der Bundestag bisher in Sachen Griechenland abgestimmt" (*Focus Online*, 27 February 2015) and "Abstimmungen im Bundestag: Rekordmehrheit für Griechenland-Hilfe" (*Wirtschaftswoche*, 27 February 2015).
[2] A candidate may also score with his attractiveness. Studies have shown that voters favor physically attractive candidates (Klein and Rosar 2005, Lawson et al. 2010, Berggren et al. 2010, 2017).
[3] On roll-call votes in the European Parliament, see Roland (2009).

defect than list MPs; the likelihood of defection decreased with higher vote margins of direct MPs (Neuhäuser et al. 2013, Sieberer 2010). Becher and Sieberer (2008), however, do not find that direct MPs are more prone than list MPs to defect during the period 1983-1994; the likelihood to defect however increased if electoral competition increased, and executive offices and party affiliation explain patterns of defection in roll-call votes.[4]

Another pertinent question is how voters react to MPs voting against the party line. While most roll-call votes occur outside of the electoral campaign, the electorate might not be aware of the representatives' voting behavior. The electorate indeed often lacks interest to be informed about the incumbents' voting records and relies mainly on party identity, therefore voters cannot hold their representatives accountable (Stokes and Miller 1962). In any event, Ansolabehere and Jones (2010) show for the United States that voters have preferences over important bills and use their beliefs about legislators' roll-call votes and parties' policy orientation to vote for their representatives. Citizens do not pay much attention to their representatives' parliamentary activities. Beliefs are rather formed from facts learned from the media and campaigns and are drawn from party labels. Incumbents, however, worry about their votes and suspect that some roll calls may become visible to the electorate, i.e. when due to media coverage some roll calls are being politicized (Arnold 1990). Roll-call votes hence can be transformed into electorally important political issues and can have an impact at the polls (Fiorina 1974). Constituents punish politicians for being too partisan (Canes-Wrone et al. 2002), but not for being ideologically too extreme (Carson et al. 2010). In the United Kingdom, policy accountability of MPs is relatively weak and general rather than issue-specific (Vivyan and Wagner 2012).

[4] Politicians who ran in highly contested electoral districts were also more likely to attend parliamentary sessions (Bernecker 2014; on attendance rates and parliamentary activity see also Gehring et al. 2015 and Geys and Mause 2016). The vote margin may also influence tax policy and political rent extraction (Solé-Ollé 2003, Kauder and Potrafke 2016). Being directly elected also influenced committee membership in parliament and re-election prospects in the next election (Stratmann and Baur 2002, Stratmann 2006, Peichl et al. 2016).

Parties decide on direct candidates solely in the respective electoral district and hence only among a few fellow party members. List candidates, however, have to face elections in state party convents to be nominated and obtain one of the few promising party list positions (Schüttemeyer 2002, Oak 2006, Hennl 2014). List candidates therefore depend even more on the loyalty of their political party.[5] An intriguing issue is how parties punish MPs who voted against the party line. Empirical evidence is scarce. In Slovakia, defecting MPs received better pre-election list positions in the future (Crisp et al. 2013). Evidence from Italy suggests that parties allocate politicians who vote in line with the party to safe positions (Galasso and Nannicini 2015). In a descriptive study on European rescue packages, Wimmel (2014) portrays that some German MPs were punished for defecting from the party line.

Using German data for the legislative period 2009-2013, we empirically investigate whether German parties punished or rewarded list candidates that voted against the party line. The dataset includes the voting behavior of 257 MPs in 218 roll-call votes. As compared to previous studies we also take into account that the effect of punishment differs along the list of candidates because a candidate is punished more when he loses positions at the threshold of promising list positions. We acknowledge that parties would not react to list candidates not adhering to the party line when these list candidates have already deviated from the party line in the legislative period 2005-2009. The financial crisis, however, increased the public attention paid to roll-call votes and politicians who voted against the majority of their parties' MPs. The results do not show that parties account for the voting behavior in parliament by punishing politicians who have voted against the party line. We thus extend the literature that has mainly focused on how *voters* react to MPs not adhering to the party line.

[5] In the German mixed electoral system most direct candidates further "collateralize" their candidacy by also being on a party list. It is hardly possible to differentiate between direct and list candidates as also direct candidates depend on their parties' loyalty in order to be placed on a promising list position, especially when direct candidates compete for unsafe districts (Manow 2012).

2. Institutional backdrop

Two major political parties characterize the political spectrum in Germany: the Social Democratic Party (SPD) and the Christian Democratic Union (CDU; in Bavaria: CSU; the CDU and CSU form one faction in the German federal parliament. In the following, we label CSU MPs as CDU). The much smaller Free Democratic Party (FDP) and the Greens (Bündnis 90/Die Grünen) have played an important role as coalition partners. The Left Party has never been part of a federal government. In our period under investigation (2009-2013), a coalition of CDU and FDP was in office.

In federal elections, voters cast two votes in a personalized proportional representation system. The first vote determines which candidate is to obtain the direct mandate in one of the 299 electoral districts with a simple majority. The second vote determines how many seats the individual parties receive in parliament. Each party that received at least 5% of the second votes obtains a number of the 598 seats in the parliament that corresponds to the party's second vote share.[6] Candidates voted into the parliament with the first vote (direct mandate) obtain their seats first. Candidates from state-specific party lists obtain the remaining seats. Note that many candidates on party lists also run as direct candidates. The list position matters only for unsuccessful direct candidates and candidates that did not run for a direct mandate (we focus on these two groups in our analysis). When the number of direct mandates exceeds the party's vote share, the party obtains excess mandates. Because the other parties did not obtain equalizing mandates in the elections before 2013, excess mandates made it possible for an individual party to receive a larger number of seats as compared to the number of seats this party would have received based on the second vote result.

Before federal elections take place and voters decide on the direct candidates, each political party nominates candidates for their state-specific party list. The list position of each

[6] Candidates obtain a direct mandate even if their party fails to reach the 5% clause. If a party obtains less than 5% of the second votes, but at least three direct mandates, the party obtains a number of seats in the parliament according to the party's second vote share.

candidate is determined during state party convents. The voting procedure differs between political parties and states. Some parties suggest only one candidate for a certain list position and the party members cast a vote approving the candidate for the specified list positions. In those nominations usually vote shares are very high for the candidates. For some parties several candidates run for a certain list position on the state-specific party list. The party members vote for the presented candidates until a clear winner is determined. In those nominations vote shares are usually notably lower for the candidates. The list position on the state-specific party lists and the number of seats a party obtained in federal elections determines who and how many of the list candidates become a member of parliament.[7]

3. Parties' reaction to MPs voting against the party line

3.1 Descriptive statistics

We use data from the website of the German federal parliament (Bundestag), from the federal election administrator, and the German newspaper "Die Zeit" for the 17th legislative period, 2009-2013. We use data for the legislative period 2009-2013 only because important control variables such as earnings from side jobs, MPs' speeches and oral contributions are fully available only since the legislative period 2009-2013. Out of 651 MPs of the German federal parliament, 298 MPs were direct candidates (we excluded one MP who left his party during the legislative period) and 353 MPs were list candidates. 257 of these list candidates were elected into parliament in the 2009 election and re-ran as list candidates in the 2013 election. To measure how individual MPs deviate from the party line, we rely on the only voting procedure that reveals the voting behavior of each MP: roll-call votes. Roll calls have to be explicitly demanded by a parliamentary party group or by 5% of MPs. Recorded votes are hence relatively rare in the federal parliament and the topics of the roll-call votes must be

[7] To accurately measure if political parties punish or reward candidates we would preferably use vote shares from within-party elections. But as the nomination procedures differ between parties and states, vote shares are unfortunately not comparable. We thus simply use list positions.

important enough so that at least a group in the parliament requested a recorded vote. 218 roll-call votes took place between the beginning of the legislative period in 2009 and the end of the legislative period in 2013.[8] For each vote we record if the MP voted yes or no or abstained from the vote (note that MPs can choose "abstain" on the ballot paper; abstention is thus different from being absent). A deviating vote is recorded when the MP voted differently than the majority of his party. In our sample of 257 list MPs, 62 MPs never deviated from the party line. The remaining MPs had between 1 and 40 deviations. We measure how often an MP deviated from the party line over the entire legislative period from 2009 to 2013 by taking the ratio of the number of deviations over votes participated. Figures 1 and 2 indicate that deviations did not matter for the list position. In a similar vein, Figure 3 does not suggest that deviations have mattered for list positions close to the threshold of promising list positions (see below): if anything, politicians may have been punished for *not* deviating from the party line. A t-test on means does, however, not indicate a significant difference in deviation ratios between politicians that have been punished and those that have not.

The governing parties in the period 2009-2013 were the CDU and the FDP. Out of the 257 MPs in our sample, 64 were in the SPD, 64 in the FDP, 47 in the Left Party, and 62 from the Greens. Because most CDU politicians were elected into parliament as direct candidates, our sample includes only 20 MPs from the CDU. MPs are on average 8.66 years in parliament. 11 MPs held an office in their party and 6 MPs held the position of a minister during the legislative period. Individual MPs gave up to 140 speeches and 139 oral contributions, and did not attend up to 43% of the roll-call votes. MPs had earnings from side jobs of up to 724.000 euros during the period; 196 MPs, however, did not record any earnings from side jobs (see, for example, Arnold et al. 2014, Becker et al. 2009, and Geys and Mause 2013). Around 60% of MPs in the sample are male and married. MPs have 1.36 children on average. MPs are on average 46.63 years old at the beginning of the legislative period in

[8] Over the legislative period 2009-2013, for example, there were 287 legislative initiatives and 208 promulgations.

2009. Individual MPs gained up to 12 or lost up to 37 positions on their party lists between 2009 and 2013. Table 1 shows descriptive statistics for the variables included in our analysis.

3.2 Empirical strategy

The baseline regression model takes the following form:

$$\textit{Party's reaction}_i = \alpha + \beta \textit{ Deviation ratio}_i$$

$$+ \Sigma_k \gamma_k \textit{ Party}_{ik} + \Sigma_l \delta_l \textit{ Political}_{il} + \Sigma_m \varepsilon_m \textit{ Personal}_{im} + u_i$$

with i=1,…,257; k=1,…,4; l=1,…,7; m=1,…,4

where *Party's reaction*$_i$ describes the change in party list positions of each candidate *i* between the elections in 2013 and 2009. We measure the change in party list positions in three different ways: in a first step, we take the difference of the party list positions of candidate *i* between the elections in 2013 and 2009 (*Number of list positions lost*$_i$). The pool of candidates differs, however, between both elections. We thus use as a second measure the change in party list positions when we omit those candidates from the party lists that did not participate in both elections. We then calculate new party list positions for only those candidates that ran in both elections and calculate the difference of those new list positions (*Number of modified list positions lost*$_i$). Our third measure takes into account that the effect of punishment differs along the list of candidates: a candidate is punished more when he loses positions at the threshold of promising list positions than when he drops from the first onto the second list position. We thus use a dummy variable which takes the value 1 if candidate *i* had a list position in 2013 (unmodified) that was worse than the last list position that got into the parliament in 2009 (*Punishment*$_i$; note that our data set only includes politicians that were successful list candidates in 2009). To be sure, this variable cannot measure whether politicians were rewarded for voting against the party line; the variable rather measures whether MPs were punished or not.[9] As main explanatory variable, we count how often MP *i* defected and voted in roll-call votes against the party line, i.e. against the majority of his party. *Deviation ratio*$_i$ describes the ratio of defected over total participated votes by MP *i*.[10] We include many control variables that are likely to predict our dependent variable and that might also be correlated with our main explanatory variable *Deviation ratio*$_i$ to deal with endogeneity concerns because of potentially omitted variables: *Party*$_{ik}$ describes dummy variables for the political parties CDU, SPD, FDP, and Left Party (reference category: Greens). The parties' reaction on MPs deviating from the party line may well differ across parties. For example, we expect a conservative party such as the CDU to punish deviating from the party line to a larger extent than the Greens, a party that experiences quite some discourse within the party and promotes grassroots democracy. In a similar vein, deviating from the party line is likely to be more common within the Green party than within the CDU. Seven control variables describe political characteristics (*Political*$_{il}$). We measure the political experience of MP *i* by the years he was in parliament or held an office in his party (party leader, faction leader or party's secretary general) or was a minister. A prominent MP such as a (local) party leader is both less likely to be punished by the party and to deviate from the party line than MPs who are less prominent. Political characteristics also include an MP's activity in parliament as measured by speeches, oral contributions, absence rate (in roll-call votes), and earnings from side jobs. MPs who are active in parliament by, for example, giving many speeches are less likely to be punished than MPs who do not give many speeches. Parties often reward MPs' efforts. We believe that the MPs giving many speeches do not annoy their parties by discussing issues and expressing views who are not in line with the

[9] Rewarding an MP would require him to jump from an unsuccessful list position to a successful list position. We can however obviously observe voting behavior only for politicians who have already been in parliament and thus have had a successful list position already in 2009.

[10] We also coded abstention as deviation when the majority of the party voted yes or no, and yes and no as deviation when the majority abstained. Inferences do not change when *Deviation ratio* is based on the value 1 for deviation, 0.5 for abstention, and 0 for no deviation in case the majority voted yes or no, and when *Deviation ratio* is based on the value 0.5 for yes and no and 0 for abstaining in case the majority abstained.

views the majority of the party holds. We therefore also conjecture that MPs who give many speeches are less likely to deviate from the party line than MPs who do not do so. Clearly, some MPs giving many speeches may also be inclined to deviate from the party line – sometimes party leaders may even advocate diverging positions to signal grassroots democracy to their voters. MPs' outside earnings are likely to be positively correlated with parties' punishment because many voters and also party members believe that "good" politicians should devote their entire time to political activities and not to outside activities. MPs having pronounced outside earnings seem to be more independent from political office than MPs with low outside earnings. The MPs' independence, in turn, should make deviations from the party line more likely (independent MPs have to care less about potential punishment by the party to afford a living). We include four control variables $Personal_{im}$ that indicate whether an MP i is male, married, how many children he has and his age in 2009. Age may predict the deviation ratio and parties' reaction in manifold ways. For example, on the one hand, young MPs (freshmen) are less likely to deviate from the party line than older MPs when they believe in strict party discipline and fear to get punished for voting against the party line. Young politicians might be punished more than older politicians, because they had less time to build up strong networks within the party and the faction in the federal parliament. On the other hand, we would expect young politicians to be punished less than older politicians, because parties acknowledge young politicians' efforts to run for office, and in times of lacking political talents, parties cannot afford punishing young promising MPs. Self-confident, young MPs might want to express independence and their own views by intentionally voting against the party line in individual roll-call votes. We also include age squared; inferences regarding the *Deviation ratio* do no change. u_i describes an error term. We estimate OLS and probit models with standard errors robust to heteroskedasticity (Huber/White/sandwich standard errors – see Huber 1967 and White 1980).

For robustness checks that do not turn out to change the inferences regarding the *Deviation ratio,* we also include fixed state effects and interaction terms between the party dummy variables and the fixed state effects. It is conceivable that habits regarding both punishment and deviation from the party line differ across states and parties within states. For example, the CDU in Hesse has been described to be stalwart. F-tests indicate however that the fixed state effects and the interaction terms between the party dummy variables and the fixed state effects do not turn out to be jointly statistically significant.

We acknowledge that there might be unobserved characteristics that we might still be worried about after including our control variables. An example is within party clashes across different regions within a state. Regional representation is one of the most important predictors of designing the party lists for the national parliament. In many states and parties, regional representation is balanced and the party lists reflect the balance of power within parties. Large and powerful regions are served first. For instance, clashes between MPs from different regions within the faction of the state parliament or on party conventions contesting influential offices within the party (positions such as local chairmen or general secretary of a party) translate into designing the party lists for the federal parliament. We are hesitant to predict the extent to which these clashes or other unobserved characteristics bias our estimate of the *Deviation ratio*. To be more explicit about whether our estimate of the *Deviation ratio* would be upward or downward biased, we would need to know the correlation between the unobserved characteristic and the deviation ratio and parties' punishment (do within party clashes make MPs more or less likely to deviate from the party line? This may well depend on the balances of power within the party and individual political career concerns. Also, parties' punishment of individual MPs depends on the balances of power within the party).

Another endogeneity concern is reverse causality. To deal with potential reverse causality we also focus on the roll-call votes which took place before the parties nominated their candidates. Inferences do no change.

3.3 Regression results

Column (1) of Table 2 shows the results of OLS regressions with our first measure of change in party list positions. The coefficient of *Deviation ratio* is negative, but does not turn out to be statistically significant (the p-value is 0.451). We believe that we have estimated a quite precise zero. The estimated coefficient of the coefficient of *Deviation ratio* in column (1) is -5.056; the mean of the *Deviation ratio* is 0.02, the standard deviation is 0.03. The mean of the number of lost list positions is -0.35 and the standard deviation is 4.65. Increasing the *Deviation ratio* by one standard deviation and taking the insignificant estimate literally would have been associated with decreasing list positions by around 0.15 – some 0.03 standard deviations. The coefficients of the political party dummies for the CDU, SPD, FDP and Left Party are all negative and statistically significant. How many years an MP was in parliament or whether an MP had a function in his party or was a minister, an MP's activities in parliament as measured by speeches and oral contributions, and the absence rate lack statistical significance. When an MP, however, had high earnings from jobs other than his parliamentary duties, he benefitted in terms of list positions. Older MPs lost in terms of list positions. The coefficient of *Age* is positive and statistically significant at the 1% level. Other personal characteristics of an MP lack statistical significance.

In column (2) of Table 2 we run the same OLS regressions, but use the measure of change in party list positions when we omit those candidates from the party lists that did not participate in both elections. The *Deviation ratio* coefficients do again not turn out to be statistically significant, indicating that how we calculate list positions does not matter.

Our first two measures of changes in list positions still include rather irrelevant shifts in list positions throughout the entire party lists (Table 2). Table 3 shows the regression results of a probit model where the dependent variable is a dummy variable indicating whether an MP – who was elected in 2009 via the party list – had a list position in 2013 which was worse than the last list position that got into the parliament in 2009. We thus focus on changes in list positions where parties do not nominate MPs on promising list positions, describing actual punishment. The coefficient of *Deviation ratio* in column (1) is negative and statistically significant at the 5% level, indicating that politicians are punished for *not* deviating from the party line when it comes to whether politicians are placed on a promising list position or not. The numerical meaning of the marginal effect of *Deviation ratio* (not shown) is that the probability of punishment decreases by 1.02 percent when the deviation ratio increases by 1 percentage point. The result is, however, not robust to excluding the five MPs with the highest deviation ratio. The coefficient of *Deviation ratio* is no longer statistically significant when we exclude the outliers in column (2).[11] We do thus not arrive at the conclusion that MPs were punished for *not* deviating from the party line.

3.4 Robustness tests

We submitted all of our results to rigorous robustness tests using different specifications of our regressions and different samples. None of these robustness tests indicates any severe fragility of our results. Table 4 describes the individual robustness tests and indicates if and how inferences of our baseline models change. In the following, we describe only individual robustness tests in more detail.

In one of the robustness tests, we investigated whether parties punish male MPs differently than female MPs. We therefore estimated our regressions separately for male and female MPs. The coefficients of *Deviation ratio* are negative and significant for male MPs in all specifications. In the subsample of female MPs the coefficients of *Deviation ratio* lack statistical significance. It is conceivable that political parties react less to the voting behavior of female MPs because females are less active in politics, and parties often have quotas of how many females should be on their party lists.

[11] Inferences regarding Table 2 do not change when we exclude the five outliers with the highest deviation ratio (results not shown).

In our baseline estimations we included MPs that entered the German parliament via a party list. In Germany, however, almost all of the candidates who run for a direct mandate are also on a party list. For politicians who run in a safe district where they are very likely to win the direct mandate the position on the party list is not relevant. We expect that there is still no effect of deviation when we include all MPs that were on a party list irrespective of whether they entered the parliament via a direct or a list mandate (or when we include only MPs who won a direct mandate). The coefficients of the *Deviation ratio* in the OLS models with our first and second measure of change in list positions are however negative and often (marginally) statistically significant for all MPs and only the MPs who won a direct mandate. The results thus indicate that MPs with a direct mandate and a good list position, i.e. MPs for whom the list position is not relevant, are rewarded for deviating from the party line. These MPs even gain positions. In a similar vein, the coefficients of *Deviation ratio* in the probit model for only MPs who won a direct mandate indicate that MPs are rewarded for deviating from the party line. The coefficients of *Deviation ratio* in the probit model for all MPs do not suggest an association between the Deviation ratio and parties' reactions.

4. Conclusion

Ample literature exists on the voters' reaction to political candidates' characteristics and behavior. Studies have shown that voters reward MPs voting against the party line in the next election. But little empirical evidence exists how parties themselves react to MPs voting against the party line. We examine whether German parties punished candidates for the parliament that voted against the party line. Using different measures for parties' reaction, our results do not show that politicians are punished for deviating from the party line when it comes to whether politicians are placed on a promising list position or not.

Our findings show that parties tolerate when politicians vote according to their own credo. Parties do not punish defecting MPs by giving them a worse list position in the future.

Our findings are in contrast to an empirical study for Slovakia, where defecting MPs received *better* pre-election list positions in the next election (Crisp et al. 2013). In Germany, the CDU – contrary to the public conjecture – did also not sanction defecting MPs; CDU MPs did thus not face any consequences when they deviated from the party line and voted against the European rescue packages. Many MPs from the FDP, by contrast, did not obtain any list position when they voted against the European rescue packages (Wimmel 2014).

Why is it that parties do not have a negative view on MPs that defect from the party line? It is conceivable that parliamentary indiscipline benefits the party because parliamentary indiscipline may increase electoral support (more voters find their individual views being reflected in the party) and poor policy outcomes are less clearly attributed to unitary actors (Powell and Whitten 1993).[12]

[12] In majoritarian election systems, party leaders anticipate voters' punishment and ask legislators in safe districts to take risks and support the partisan cause because safe seats can afford to lose a modest amount of votes (Carson et al. 2010). An increase in party unity on voting at the aggregate level has adverse electoral costs for both parties over time (Lebo et al. 2007). Parties may however also incur costs from nominating notable individually strong candidates which are less dependent on the political party leaders and are hence more likely to break party unity (Cantor and Herrnson 1997, Heidar 2006, Kam 2009, Tavits 2009, 2010).

References

Ansolabehere, Steven and Philip E. Jones (2010), Constituents' responses to congressional roll-call voting, *American Journal of Political Science* 54, 583-597.

Arnold, R. Douglas (1990), *The Logic of Congressional Action*. New Haven, CT: Yale University Press.

Arnold, Felix, Björn Kauder, and Niklas Potrafke (2014), Outside earnings, absence, and activity: Evidence from German parliamentarians, *European Journal of Political Economy* 36, 147-157.

Becher, Michael and Ulrich Sieberer (2008), Discipline, electoral rules and defection in the Bundestag, 1983-94, *German Politics* 17, 293-304.

Becker, Johannes, Andreas Peichl, and Johannes Rincke (2009), Politicians' outside earnings and electoral competition, *Public Choice* 140, 379-394.

Berggren, Niclas, Henrik Jordahl, and Panu Poutvaara (2010), The looks of a winner: Beauty and electoral success, *Journal of Public Economics* 94, 8-15.

Berggren, Niclas, Henrik Jordahl, and Panu Poutvaara (2017), The right look: Conservative politicians look better and voters reward it, *Journal of Public Economics* 146, 79-86.

Bernecker, Andreas (2014), Do politicians shirk when reelection is certain? Evidence from the German parliament, *European Journal of Political Economy* 36, 55-70.

Bütikofer, Sarah and Simon Hug (2015), Strategic behaviour in parliament, *Journal of Legislative Studies* 21, 295-322.

Canes-Wrone, Brandice, David W. Brady, and John F. Cogan (2002), Out of step, out of office: Electoral accountability and House members' voting, *American Political Science Review* 96, 127-140.

Cantor, David M. and Paul S. Herrnson (1997), Party campaign activity and party unity in the U.S. House of Representatives, *Legislative Studies Quarterly* 22, 393-415.

Carson, Jamie L., Gregory Koger, Matthew J. Lebo, and Everett Young (2010), The electoral costs of party loyalty in Congress, *American Journal of Political Science* 54, 598-616.

Crisp, Brian F., Santiago Olivella, Michael Malecki, and Mindy Sher (2013), Vote-earning strategies in flexible list systems: Seats at the price of unity, *Electoral Studies* 32, 658-669.

Fiorina, Morris P. (1974), *Representatives, Roll Calls, and Constituencies*. Lexington, KY: Lexington Books.

Galasso, Vincenzo and Tommaso Nannicini (2015), So closed: Political selection in proportional systems, *European Journal of Political Economy* 40, 260-273.

Gehring, Kai, T. Florian Kauffeldt, and Krishna C. Vadlamannati (2015), Crime, incentives and political effort: A model and empirical application for India. Unpublished paper.

Geys, Benny and Karsten Mause (2013), Moonlighting politicians: A survey and research agenda, *Journal of Legislative Studies* 19, 76-97.

Geys, Benny and Karsten Mause (2016), The limits of electoral control: Evidence from last-term politicians, *Legislative Studies Quarterly* 41, 873-898.

Heidar, Knut (2006), Parliamentary party group unity: Does the electoral system matter? *Acta Politica* 41, 249-266.

Hennl, Annika (2014), Intra-party dynamics in mixed-member electoral systems: How strategies of candidate selection impact parliamentary behaviour, *Journal of Theoretical Politics* 26, 93-116.

Huber, Peter J. (1967), The behavior of maximum likelihood estimates under nonstandard conditions, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, 221-233.

Kam, Christopher J. (2009), *Party Discipline and Parliamentary Politics*. Cambridge, UK: Cambridge University Press.

Kauder, Björn and Niklas Potrafke (2016), Supermajorities and political rent extraction, *Kyklos* 68, 65-81.

Klein, Markus and Ulrich Rosar (2005), Physical attractiveness and electoral success. An empirical investigation on candidates in constituencies at the German federal election 2002, *Politische Vierteljahresschrift* 46, 263-287.

Kunicova, Jana and Thomas F. Remington (2008), Mandates, parties and dissent. Effect of electoral rules on parliamentary party cohesion in the Russian State Duma, 1994-2003, *Party Politics* 14, 555-574.

Lawson, Chappell, Gabriel S. Lenz, Andy Baker, and Michael Myers (2010), Looking like a winner: Candidate appearance and electoral success in new democracies, *World Politics* 62, 561-593.

Lebo, Metthew, Adam J. McGlynn, and Greg Koger (2007), Strategic party government: Party influence in Congress, 1789-2000, *American Journal of Political Science* 51, 464-481.

Manow, Philip (2012), Wahlkreis- oder Listenabgeordneter, Typus oder Episode? Eine Sequenzanalyse der Wege in den Bundestag, *Politische Vierteljahresschrift* 53, 53-78.

Mayhew, David (1974), *Congress: The Electoral Connection*. New Haven, CT: Yale University Press.

Neuhäuser, Alice, Maike Mischler, Graeme D. Ruxton, Markus Neuhäuser (2013), Gründe für von der Fraktionsdisziplin abweichendes Abstimmungsverhalten bei Bundestagsabgeordneten, *AStA Wirtschafts- und Sozialstatistisches Archiv* 7, 91-99.

Oak, Mandar P. (2006), On the role of the primary system in candidate selection, *Economics & Politics* 18, 169-190.

Peichl, Andreas, Nico Pestel, and Sebastian Siegloch (2016), Incumbency advantage in a mixed electoral system: The case of Germany. Unpublished paper.

Powell Jr., G. Bingham and Guy D. Whitten (1993), A cross-national analysis of economic voting: Taking account of the political context, *American Journal of Political Science* 37, 391-414.

Roland, Gérard (2009), Legislative behavior in the European Parliament, *CESifo Economic Studies* 55, 15-29.

Schüttemeyer, Suzanne S. (2002), Wer wählt wen wie aus? Pfade in das unerschlossene Terrain der Kandidatenaufstellung, *Gesellschaft - Wirtschaft - Politik* 51, 145-161.

Sieberer, Ulrich (2010), Behavioral consequences of mixed electoral systems: Deviating voting behavior of district and list MPs in the German Bundestag, *Electoral Studies* 29, 484-496.

Solé-Ollé, Albert (2003), Electoral accountability and tax mimicking: The effects of electoral margins, coalition government, and ideology, *European Journal of Political Economy* 19, 685-713.

Stokes, Donald E. and Warren E. Miller (1962), Party government and the saliency of Congress, *Public Opinion Quarterly* 26, 531-546.

Stratmann, Thomas (2006), Party-line voting and committee assignments in the German mixed-member system. In: Roger D. Congleton and Birgitta Swedenborg (eds.): *Democratic Constitutional Design and Public Policy. Analysis and Evidence*. Cambridge, MA: MIT Press, 111-130.

Stratmann, Thomas and Martin Baur (2002), Plurality rule, proportional representation, and the German Bundestag: How incentives to pork-barrel differ across electoral systems, *American Journal of Political Science* 46, 506-514.

Tavits, Margit (2009), The making of mavericks: Local loyalties and party defection, *Comparative Political Studies* 42, 793-815.

Tavits, Margit (2010), Effect of local ties on electoral success and parliamentary behaviour, *Party Politics* 16, 215-235.

Thames, Frank C. (2005), A House divided, *Comparative Political Studies* 38, 282-302.

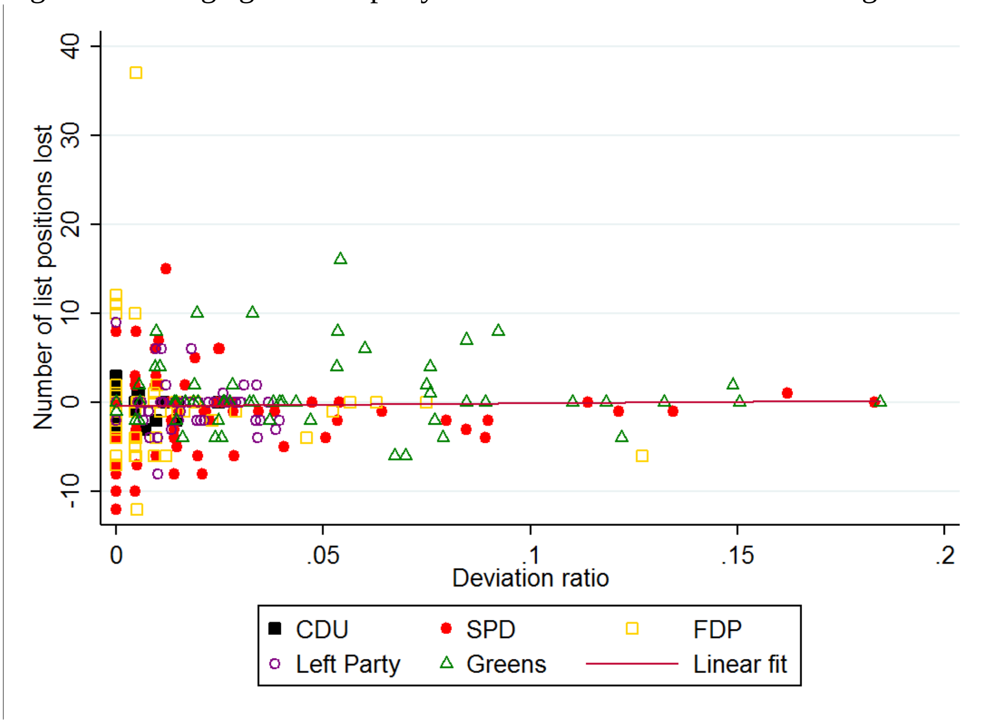
Vivyan, Nick and Markus Wagner (2012), Do voters reward rebellion? The electoral accountability of MPs in Britain, *European Journal of Political Research* 51, 235-264.

White, Halbert (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817-838.

Wimmel, Andreas (2013), Fachliche Expertise und abweichendes Verhalten bei Abstimmungen zur Euro-Krise im Deutschen Bundestag, *Zeitschrift für Politikberatung* 6, 125-136.

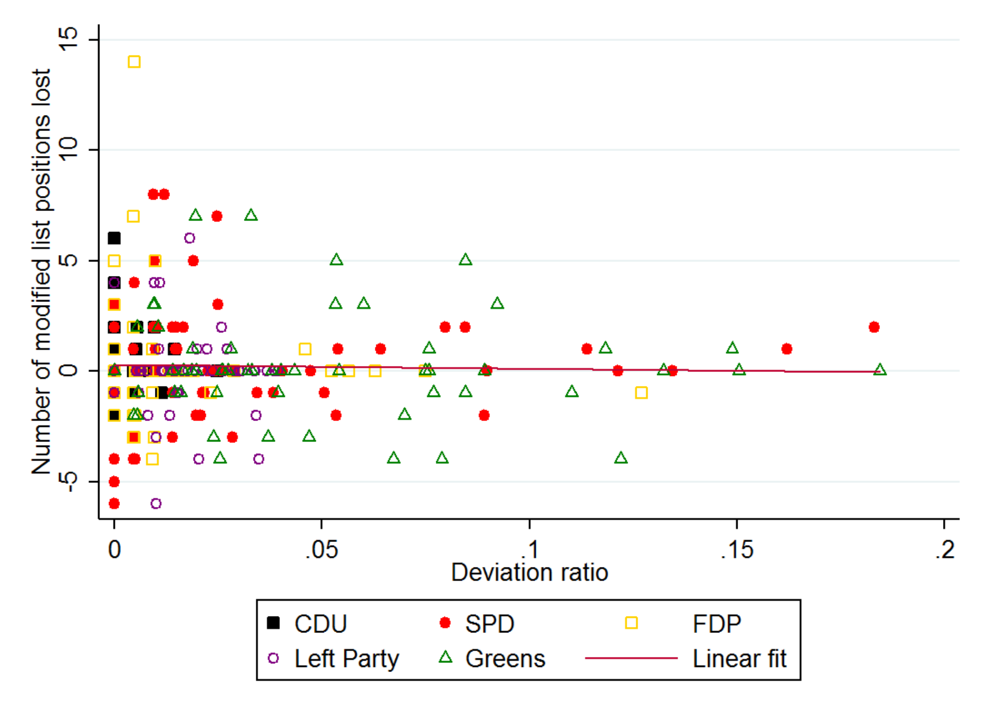
Wimmel, Andreas (2014), Nach der ESM-Abstimmung im Bundestag: Abstrafung der Euro-Kritiker? *Gesellschaft - Wirtschaft - Politik* 63, 13-22.

Figure 1: Voting against the party line is not correlated with a change in the list position



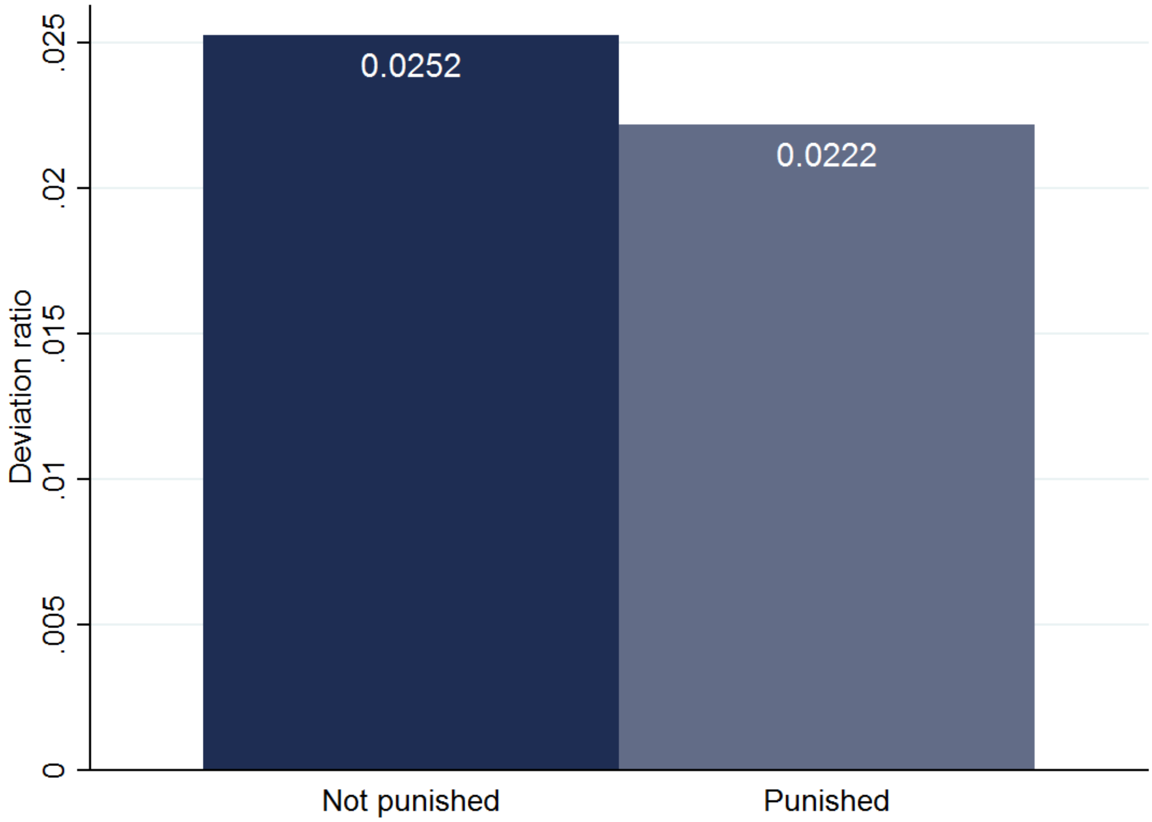
Note that negative values on the vertical axis indicate gained list positions.
Source: Own illustration.

Figure 2: Voting against the party line is not correlated with a change in the modified list position



Note that negative values on the vertical axis indicate gained modified list positions.
Source: Own illustration.

Figure 3: MPs who deviated more often from the party line were not more often punished



Source: Own illustration. A t-test on means does not indicate a significant difference between "not punished" and "punished" (t-value 0.41).

Table 1: Descriptive statistics

	Obs.	Mean	Std. Dev.	Min.	Max.
Number of list positions lost	257	-0.35	4.65	-12.00	37.00
Number of modified list positions lost	257	0.25	2.39	-6.00	14.00
Punishment	257	0.09	0.29	0	1
Deviation ratio	257	0.02	0.03	0.00	0.18
CDU	257	0.08	0.27	0	1
SPD	257	0.25	0.43	0	1
FDP	257	0.25	0.43	0	1
Left Party	257	0.18	0.39	0	1
Greens	257	0.24	0.43	0	1
Years in parliament	257	8.66	5.79	0.69	32.98
Function in party	257	0.10	0.55	0.00	3.98
Minister	257	0.08	0.52	0.00	3.98
Speeches	257	36.21	22.67	1	140
Oral contributions	257	13.33	13.72	0	139
Absence rate	257	0.08	0.08	0.00	0.43
Earnings from side jobs	257	16.67	62.01	0.00	724.00
Male	257	0.58	0.50	0	1
Married	257	0.62	0.49	0	1
Number of children	257	1.36	1.35	0	7
Age	257	46.63	9.91	23	69

Years in parliament, *Function in party*, and *Minister* measured in years; *Earnings from side jobs* measured in 1000 euros; *Age* measured in 2009.

Table 2: Regression results (OLS model).
Dependent variable: Number of list positions lost (column 1) and Number of modified list positions lost (column 2).

	(1)	(2)
Deviation ratio	-5.056	-2.462
	(0.451)	(0.487)
CDU	-3.782**	-0.213
	(0.018)	(0.779)
SPD	-3.577***	-0.403
	(0.002)	(0.430)
FDP	-1.687**	0.272
	(0.047)	(0.484)
Left Party	-1.526**	-0.377
	(0.039)	(0.389)
Years in parliament	0.173	0.077
	(0.175)	(0.115)
Function in party	-0.405	-0.258
	(0.288)	(0.108)
Minister	-0.557	-0.435**
	(0.229)	(0.034)
Speeches	-0.013	-0.001
	(0.418)	(0.880)
Oral contributions	-0.030	-0.008
	(0.164)	(0.513)
Absence rate	-1.040	-1.088
	(0.807)	(0.624)
Earnings from side jobs	-0.009*	-0.004*
	(0.068)	(0.086)
Male	0.381	0.195
	(0.498)	(0.508)
Married	-0.704	-0.467
	(0.235)	(0.145)
Number of children	0.097	-0.090
	(0.670)	(0.463)
Age	0.120***	0.074***
	(0.000)	(0.000)
Observations	257	257
R^2	0.173	0.166

Number of list positions lost describes how many list positions an MP lost on the 2013 list compared to the 2009 list of his party. *Number of modified list positions lost* describes how many list positions an MP lost on the 2013 list compared to the 2009 list of his party after omitting those candidates from the party lists that did not participate in both elections. *Deviation ratio* describes the ratio of defected over total participated votes by an MP.
Standard errors robust to heteroskedasticity (Huber/White/sandwich standard errors).
p-values in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Regression results (Probit model).
Dependent variable: Punishment.

	Full sample (1)	Outlier excluded sample (2)
Deviation ratio	-7.699**	-6.493
	(0.018)	(0.105)
CDU	-1.464**	-1.411**
	(0.019)	(0.023)
SPD	-0.766**	-0.736**
	(0.027)	(0.032)
FDP	-0.361	-0.320
	(0.332)	(0.391)
Left Party	-0.383	-0.375
	(0.313)	(0.321)
Years in parliament	-0.000	-0.000
	(0.988)	(0.997)
Speeches	-0.012	-0.011
	(0.117)	(0.125)
Oral contributions	0.004	0.004
	(0.648)	(0.625)
Absence rate	-3.501	-3.375
	(0.154)	(0.167)
Earnings from side jobs	-0.011**	-0.011**
	(0.039)	(0.040)
Male	0.200	0.182
	(0.431)	(0.479)
Married	-0.479*	-0.481*
	(0.084)	(0.082)
Number of children	0.105	0.105
	(0.286)	(0.287)
Age	0.059***	0.059***
	(0.000)	(0.000)
Observations	257	252
Pseudo R^2	0.214	0.212
Chi-squared	38.56	37.14
Prob > Chi-squared	0.000426	0.000701
Log likelihood	-62.66	-62.48

Punishment describes a dummy variable which takes the value 1 if a candidate had a list position in 2013 that was worse than the last list position that got into the parliament in 2009. *Deviation ratio* describes the ratio of defected over total participated votes by an MP. We exclude *Function in party* and *Minister*, because having a function in a party and being a minister predict failure perfectly.
Standard errors robust to heteroskedasticity (Huber/White/sandwich standard errors).
p-values in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Robustness tests.

Robustness test	Do inferences change?
Separating roll-call votes into categories to investigate whether punishment depends on the different topics of the votes: general foreign policy, military actions, domestic policy in general, domestic policy during the financial and economic crisis, energy topics, European politics in general, European rescue packages, and in particular Greek rescue packages.	Weak evidence for general foreign policy and for domestic policy during the financial and economic crisis.
Identifying the ten most important roll-call votes using Google Trends and using deviating from the party line in these votes as dependent variable.	No.
Using the total number of deviations from the party line of an MP instead of *Deviation ratio*.	No.
Counting being absent as a deviation. *Deviation ratio* then describes the ratio of defected over *total* votes.	No.
Estimating the regressions separately for each political party.	In some regressions significant effects for the SPD, Greens and Left Party.
Running regressions separately for male and female MPs	Significant effect for male MPs in all models. No effect for female MPs.
Measuring a promising list position to enter the federal parliament by using (a) the average position in the 1998, 2002, and 2005 national elections that sufficed to enter parliament and (b) the position in the next (2013) election that sufficed to enter parliament.	No.
Including all MPs that were on a party list irrespective of whether they entered the parliament via a direct or a list mandate, or including only MPs who won a direct mandate.	*Deviation ratio* in the OLS models is negative and often (marginally) significant for all MPs and only the MPs who won a direct mandate, and in the probit model for only MPs who won a direct mandate. No effect in the probit model for all MPs.
Testing whether parties are more attentive to the voting behavior of MPs during the time party list positions are voted on inside the parties (usually two years before the election): including the ratio of deviating over participated votes separately for each year of the legislative period.	Significant effect of *Deviation ratio* only in the probit model, which is strongest in the years 2011 and 2012 (the two years before the election year).
Testing for a selection effect: MPs who feared that they would be punished with non-viable list positions may have retired.	Retiring MPs on average deviated less than MPs who ran in the next election (moreover, retiring cannot be explained by deviating from the party line).