Vol. 50 No. 2 (NOV, 2013), 163–177

LATIN AMERICAN JOURNAL OF ECONOMICS

THE REAL OPTIONS APPROACH TO VALUATION: CHALLENGES AND OPPORTUNITIES*

Eduardo Schwartz**

This paper provides an overview of the real options approach to valuation mainly from the point of view of the author who has worked in this area for over 30 years. After a general introduction to the subject, numerical procedures to value real options are discussed. Recent developments in the valuation of complex American options has allowed progress in the solution of many interesting real option problems. Two applications of the real options approach are discussed in more detail: the valuation of natural resource investments and the valuation of research and development investments.

JEL classification: G12

Keywords: Real options

1. Introduction

The real options approach is an extension of financial options theory to options on real/non-financial assets. Options are contingent decisions that provide the opportunity to make a decision after uncertainty unfolds. Uncertainty and the agent's ability to respond to it (flexibility) are the source of value of an option. Whenever possible, real options valuations are aligned with financial market valuations.

Most investments are subject to options valuation. There are four main types of options associated with investment projects–the option to expand, to postpone, to abandon, and to temporarily suspend an investment. For example, the option to expand a project is valuable when a firm may want to invest in a negative net present value (NPV) project if it provides the firm the possibility of developing a new project. Consider the valuation of a mine of which, at current commodity prices, only half is economically feasible for development. This investment will provide the option to develop the remainder of the mine when and if market prices change. In this case, the option to expand is valuable and must be considered when quantifying the value of the mine. On

* This lecture was delivered at the 2013 Finance UC Conference in Santiago, Chile. Hay Jin Kim provided valuable assistance.

** California Chair in Real Estate and Land Economics Professor of Finance, Anderson School of Management, University of California, Los Angeles. Email: eduardo.schwartz@anderson.ucla.edu.

doi 10.7764/LAJE.50.2.163

the other hand, even with a positive NPV project, the option to delay the investment is valuable as it gives the firm the opportunity to wait until more market information is available. Furthermore, the option to abandon a project is important and valuable in research and development (R&D) investments as it provides the flexibility to abandon a project in the presence of negative outcomes. In contrast to the traditional approach that uses expected cash-flows to value investment projects, the real option approach takes into account the entire distribution of cash-flows, allowing the firm to react/respond during the course of the investment. Finally, the option to temporarily suspend production is valuable whenever a firm has the opportunity to open and temporarily close a facility. For instance, when a commodity price is low, the firm can choose to close its facility and re-open it later when prices are higher.

Thus, flexibility can be an important component of value for many investment projects and the option-pricing framework provides a powerful tool for analyzing such flexibility. Furthermore, the real options approach to valuation is currently being applied in practice and extended in several directions. In particular, this method has been broadened to take into account competitive interactions and their impact on option exercise strategies.

The remainder of this paper is organized as follows. Section 2 compares the two main approaches to value investment projects. Section 3 briefly describes three procedures used to solve option valuation problems. Section 4 presents two particular applications of the real option approach in investment projects. Finally, Section 5 concludes.

2. Two valuation approaches

Even in the absence of real options there are two main approaches to value investment projects. The traditional valuation technique, known as discounted cash-flows (DCF) or net present value (NPV), requires forecasts. It uses a single expected value of future cashflows. A simplified version of the traditional approach is:

$$NPV = C_0 + \sum_{t=1}^{N} \frac{C_t}{(1+k)^t} \tag{1}$$

where C_t is the expected cash flow in period t and k is the risk-adjusted discount rate. By defining cash flows as the profits obtained by the investment project, Equation (1) can be rewritten as:

$$NPV = C_0 + \sum_{t=1}^{N} \frac{q_t S_t - Cost_t}{(1+k)^t} \quad (2)$$

where q_t is the quantity produced and S_t is the spot price, assumed to be the only source of uncertainty in this simplified version.

There are two main drawbacks to the traditional approach that makes it inappropriate for valuing projects in many practical situations. First, DCF assumes that future firm decisions are fixed at the outset and ignores the flexibility in decision making during the course of the investment project. Second, when there are options (e.g., exit option) involved in the investment project, it is difficult to find an appropriate discount rate k to calculate the NPV in Equation (1).

Alternatively, the risk-neutral (RN) valuation or certainty-equivalent (CE) approach can effectively capture the flexibility embedded in real options valuation. In the CE approach, the adjustment for risk is in the probability distribution of cash flows instead of the discount rate. The NPV of a project is then calculated by discounting the certainty equivalent cash flows CEQ_t by the risk-free rate:

$$r_f : NPV = C_0 + \sum_{t=1}^{N} \frac{CEQ_t}{(1+r_f)^t} = C_0 + \sum_{t=1}^{N} \frac{q_t F_t - Cost_t}{(1+r_f)^t}. \quad (3)$$

As can be observed in Equation (3), in order to calculate the certainty equivalent cash flows, futures prices F_t are used instead of the spot prices S_t. Futures prices are the expected future spot prices under the risk-neutral distribution.

Cox and Ross (1976), Harrison and Kreps (1979), and Harrison and Pliska (1981) show that the absence of arbitrage implies the existence of a probability distribution, such that securities are priced at their discounted (at the risk-free rate) expected cash flows under these risk-neutral or risk-adjusted probabilities. Moreover, these probabilities are unique if markets are complete–all risks can be hedged. If, on the other hand, markets are not complete, their probabilities are not unique, but any of them can be used for pricing.[1]

1. The risk-neutral valuation approach can be generalized to include stochastic discount rates: $V_0 = E^Q[e^{\int_0^T -r_f(t)dt} X_T]$

There are three cases to consider in real option risk-neutral valuation. The first case is when the risk-neutral distribution is known, as in the Black-Scholes framework; unfortunately, the only pure example of this case in the real world are gold mines. In such case, the futures prices are $F_{o,T} = S_o(1 + r_f)^T$.

The second case is when the risk-neutral distribution is unknown but can be obtained from futures prices or other traded assets (e.g., copper mines and oil deposits). In Section 4 this topic will be explored further.

The last case is when the risk-neutral distribution is unknown and futures prices do not exist. In this case the risk-neutral distribution can be obtained by using an equilibrium model, such as the CAPM. This is the most common case in R&D projects, Internet companies, and information technology where no futures prices exist.

Thus, using the risk-neutral framework to value investment projects allows for use of all information contained in futures prices when these prices exist, to take into account all flexibilities/options the projects may have and use the powerful analytical tools that have been developed in contingent claims analysis.

3. Solution procedures to option valuation problems

There are three main solution methods for solving option valuation problems: the dynamic programming approach, partial differential equations, and the simulation approach. The first approach uses dynamic programming techniques to lay out possible future outcomes and folds back the value of the optimal future strategy using risk-neutral distributions. The binomial method is a dynamic programming approach widely employed to value simple options. It can also be used to price American-type options. However, this solution method becomes inadequate when there are multiple factors affecting the value of the option or when there are path dependencies.

The second method directly solves the partial differential equations (PDE) that result from most option pricing problems. This approach leads to closed-form solutions in very few cases, such as the Black-Scholes equation for European call options. In most option valuation problems the PDE has to be solved numerically. This is a very flexible method, and it is appropriate for valuing American options. Finding a solution however becomes extremely complicated when there are more than three state variables; thus, PDE is an inadequate method for solving the more complex real option problems. Furthermore, this method is technically sophisticated as it needs to approximate boundary conditions.

In general option pricing problems can also be solved by simulation. The simulation approach is very powerful; however, it is forward-looking whereas the optimal exercise of an American option has dynamic programming features.

Longstaff and Schwartz (2001) developed a simulation approach to valuing American options. An American option gives its holder the right to exercise at multiple points in time (finite number) before its maturity date. At each exercise point, the holder optimally compares the immediate exercise value with the value of continuation. As standard theory implies that the value of continuation can be expressed as the conditional expected value of discounted future cash flows, the basic idea behind the simulation approach is that the conditional expected value of continuation can be estimated from the cross-sectional information from the simulation by least-squares. The conditional expectation function is estimated by regressing discounted ex-post realized cash flows from continuation on functions of the current (or past) values of the state variables. The fitted value from this cross-sectional regression is shown to be an efficient estimator of the conditional expectation function. Thus, by estimating the conditional expectation function for each exercise date in each of the possible simulated paths, an optimal stopping rule for the option and hence its current value can be accurately estimated.

Among all three solution methods, the most useful tool for solving real options valuation problems is the simulation approach. It is easily applied to multi-factor models and directly applicable to path-dependent problems. Furthermore, it allows the state variables to follow general stochastic processes. It is intuitive, transparent, flexible, easily implemented, and computationally efficient.

4. Applications of the real options approach

Since Longstaff and Schwartz (2001) developed the simulation approach, complex real option valuation problems have been analyzed in numerous areas. Two particular applications of the real option approach will be discussed in this section: natural resource investments and pharmaceutical R&D investments.

4.1. Natural resource investments

Commodity-linked bonds were increasingly issued in the late 1970s. For instance, the Mexican government issued bonds backed by oil in 1979. A year later, in 1980, Sunshine Mining Company issued bonds backed by silver in the United States. In addition, gold-backed bonds have been around for a long time. Schwartz (1982) attempts to value commodity assets. He proposes a model, based on the Black-Scholes option pricing framework extended by Merton (1973) and Cox and Ross (1976), to deal with the problem of accurately valuing commodity-linked bonds. These bonds' payouts (coupon and/or principal) are directly linked to the market price of the underlying commodity (such as oil, copper, or gold), the interest rate, and the value of the firm.

The earlier work on natural resource investment valuation assumed that commodity prices follow a simple stochastic process similar to that of stock prices. This simplistic assumption is found to be appropriate for gold but inadequate for other types of commodities: Supply and demand adjustments induce mean reversion in commodity prices. On the supply side, an increase in the commodity price will induce high-cost producers to enter the market which, in turn, will decrease the market price; once the price is low, those high-cost producers will exit the market to avoid negative profits, increasing the market price once again. On the demand side, when the market price is high, consumers will substitute the commodity and demand for it will fall, resulting in a decrease in the commodity price.

Brennan and Schwartz (1985) acknowledge that the evaluation of mining and other natural resource projects is difficult due to uncertainty in commodity prices; they propose that mine and oil deposits can be interpreted and valued as complex options on the underlying commodities. This is one of the first papers on real options valuation. The valuation model takes explicitly into account managerial control over the output rate and the possibility of abandoning the project if output prices decrease substantially. The approach relies on constructing a self-financing portfolio of riskless bonds and futures contracts whose cash flows replicate those of the investment project to be valued. The main assumptions underlying this replication are "that the convenience yield on the output commodity can be written as a function of the output price alone, and that the interest rate is nonstochastic."

Subsequent works on commodity asset valuation have focused on making more realistic assumptions about the stochastic processes followed by commodity prices. For instance, Cortazar and Schwartz (2003) develop a three-factor model of the term structure of oil futures prices that can be estimated from available futures price data. The procedure is flexible and can take into account the dynamics of futures prices. The true stochastic process of spot prices is modeled as:

$$\begin{aligned} dS &= (v - y)Sdt + \sigma_1 S dz_1 \\ dy &= -\kappa y dt + \sigma_2 dz_2 \\ dv &= a(\overline{v} - v)dt + \sigma_3 dz_3 \end{aligned} \tag{4}$$

where v can be interpreted as the long-term drift of the process and y can be interpreted as the convenience yield. Both v and y are mean reverting processes, which in turn induce some mean reversion in S . However, in order to value options, a risk-neutral process is needed. For instance, assuming that the market prices of risk (λ_1,λ_2, and λ_3) are constants, the risk-adjusted processes are

$$\begin{aligned} S &= (v - y - \lambda_1)Sdt + \sigma_1 S dz_1 \\ dy &= (-\kappa y - \lambda_2)dt + \sigma_2 dz_2 \\ dv &= (a(\overline{v} - v) - \lambda_3)dt + \sigma_3 dz_3 \end{aligned} \tag{5}$$

Futures prices can be analytically derived from these equations. The joint processes can then be estimated by a Kalman filter. Using daily prices of all futures contracts traded on the New York Mercantile Exchange (NYMEX) between 1991 and 2001, the estimation results indicate that the model fits the data extremely well.

Some challenges facing all commodity price models are that assumptions about the functional forms of market prices of risks are needed and that we can have confidence in the model's fit only for the period for which we have futures data (typically not more than six years into the future). One of the main challenges of the three-factor model is dealing with longer maturities where no futures prices exist: Should we accept the model predictions for maturities where there are no futures prices? Should we assume that futures prices are constant or that they increase at a fixed rate? What discount rate should we use for longer maturities?

More recently, Trolle and Schwartz (2009) note that once options data are included in the valuation of commodities (in addition to futures price data), it is critical to understand the dynamics of volatility in commodity markets for pricing, hedging, and risk management of commodity options and real options. While volatility in commodity markets is stochastic, it is not clear the extent to which volatility is spanned by the factors that affect futures prices. Schwartz and Trolle analyze this issue in the crude-oil market and develop a tractable model for pricing commodity derivatives in the presence of unspanned stochastic volatility. The model is then estimated on NYMEX crude-oil derivatives using "an extensive panel data set of 45,517 futures prices and 33,104 option prices, spanning 4082 business days." First, the covariance matrix of the futures returns are factor-analyzed retaining the first three principal components (PCs). Then, the straddle returns or changes in implied volatilities which are used as proxies for changes in actual volatility are regressed on PCs and PCs squared (to take into account possible non-linearities):

$$\begin{aligned} r_t^{strad,i} &= \beta_0^i + \beta_1^i PC_t^{fut,1} + \beta_2^i PC_t^{fut,2} + \beta_3^i PC_t^{fut,3} \\ &\quad + \beta_4^i (PC_t^{fut,1})^2 + \beta_5^i (PC_t^{fut,2})^2 + \beta_6^i (PC_t^{fut,3})^2 + \epsilon_t^i \end{aligned} \tag{6}$$

The resulting R^2 for the regresions are typically low, between 0% and 21%. Thus, the factors that explain futures prices cannot explain changes in volatility.

Next, in order to check the existence of systematic factors affecting volatility, the covariance matrix of the residuals from these regressions are factor-analyzed. The results indicate that the first two PCs explain over 80% of the variation in the residuals.

Based on these results, Trolle and Schwartz (2009) develop models with one and two volatility factors, in addition to the factors affecting commodity prices. These were the first models estimated using also options data, in addition to futures price data.

Schwartz and Trolle (2010) use this model to price expropriation risk in a natural resource project. Their approach focuses on some of the important economic trade-offs that arise from a government holding an "option" to expropriate an oil field, abstracting from the various operational options that are typically embedded in natural resource investments. While the main benefit from exercising the expropriation option is that the government receives all the profits rather than a fraction through taxes, the expropriation costs are that a private firm may produce oil more efficiently, that the government may have to pay compensation to the firm, and that the government may face "reputational" costs. Given these variables, exercise of the expropriation option by the government is optimally determined. Spot prices, futures prices, and volatilities are described by the dynamics proposed in Trolle and Schwartz (2009). Furthermore, the expropriation option is modeled as an American-style option. At every point in time, the government must compare the value of immediate exercise with the conditional expected value (under the risk-neutral measure) of continuation. The optimal exercise time for each simulated path can then be used to value the expropriation option.

The results indicate that for a given contractual arrangement, the value of the expropriation option increases with the spot price, the slope of the futures curve, and the volatility of the spot (futures) price. On the other hand, for a given set of state variables, the value of the expropriation option decreases with the tax rate and various expropriation costs. Furthermore, the increase in the field's value to the government due to the expropriation option is found to be always smaller than the decrease in the field's value to the firms due to the "deadweight losses" associated with the expropriation process, i.e., production inefficiency and reputational costs.

4.2. R&D investments

The main focus of real options valuation in R&D investment projects has been on the pharmaceutical industry. However, the pharmaceutical R&D framework can easily be applied to other research-intensive industries.

The pharmaceutical industry has become a research-oriented sector that makes a major contribution to health care. The success of the industry in generating a stream of new drugs with important therapeutic benefits has created an intense public policy debate over issues such as the financing of research, the prices charged for its products, and the socially optimal degree of patent protection. There is a trade-off between promoting innovative efforts and securing competitive market outcomes. While expected monopoly profits from drug sales during the life of the patent compensate the innovator for the risky investment, the onset of competition after patent expiration limits the deadweight losses to society that arise from monopoly pricing under the patent. Moreover, regulation has had important effects on the cost of innovation in the pharmaceutical industry.

The analysis of R&D projects in the pharmaceutical industry must take into account some of its unique characteristics. First, the development of a new drug takes a long time; the average is between 10 and 12 years. Second, there is uncertainty about the costs of development and the time to completion. Although the average time to completion is 12 years, the development of a new drug could take 20 years, by which time the patent may expire and the project will therefore be abandoned. Third, there is also a high probability of failure for either technical or economic reasons; for example, 80% of projects that start clinical trials are later abandoned. Technical reasons include catastrophic events, while economic reasons comprise the high cost of production and the inefficacy of the drugs. Fourth, new drugs require approval by the U.S. Food and Drug Administration (FDA). Finally, once the drug has been approved, there is uncertainty about the level and duration of future cash flows as the time to completion and the length of the patent are also uncertain. As a result of the particular procedure followed by the pharmaceutical researchers and the high costs[2] involved in the development stages, an abandonment option is clearly valuable in the pharmaceutical industry and must be considered in the valuation method.

Schwartz (2004) proposes a methodology to quantify the value of a single R&D project that is patent-protected. Equivalently, the approach aims to determine the price of the patent. Taking into consideration all the unique features of the pharmaceutical industry described above, Schwartz treats the patent-protected R&D project as a complex option on the variables underlying the value of the project. There are two variables that are taken into consideration: expected costs to completion and anticipated cash flows. By allowing these two variables to follow stochastic processes through time, uncertainty is introduced into the analysis. The expected cost to completion is assumed to follow:

$$dK = -Idt + \sigma(IK)^{\frac{1}{2}}dz. \tag{7}$$

2. In the United States alone, annual expenditures on prescription drugs were more than US$300 billion in 2010. For instance, Pfizer Inc., the largest pharmaceutical company in the world, invested over US$108 billion in R&D from 1997 to 2011, successfully developing 14 new drugs. In the same period, Amgen Inc. spent over US$33 billion on R&D, with an average investment of $3.6 billion per successful drug.

As can be seen in Equation (7), the cost to completion, K, decreases with investments I and is affected by a random shock, such that the volatility of the cost process can be expressed as:

$$Var(\tilde{K}) = \frac{\sigma^2 K^2}{2 - \sigma^2}. \tag{8}$$

Concurrently, the cash flows follow a geometric Brownian motion (which may be correlated with the cost-process):

$$dC = \alpha C dt + \phi C dw \tag{9}$$

and the associated risk-adjusted process used for valuation is:

$$dC = (\alpha - \eta) C dt + \phi C dw = \alpha^* C dt + \phi C dw. \tag{10}$$

Due to the absence of futures prices, the risk-adjusted process for the cash-flows (i.e., η) is obtained using the returns of traded pharmaceutical companies.

In this model, the value of the project once investment has been completed will depend on the cash flows and on time: $V(C,t)$. An analytical solution can be found by solving the PDE:

$$\frac{1}{2}\phi^2 C^2 V_{CC} + \alpha^* C V_C + V_t - rV + C = 0 \tag{11}$$

setting the following boundary condition:

$$V(C,T) = M \bullet C. \tag{12}$$

Condition (12) implies that the value of the project at expiration date T of the patent is a multiple or fraction M of the cash-flows. Thus,

$$\begin{aligned} V(C,t) = {} & \frac{C}{r - \alpha^*}[1 - \exp(-(r - \alpha^*)(T - t))] \\ & + MC \exp(-(r - \alpha^*)(T - t)). \end{aligned} \tag{13}$$

Based on Equation (13), the stochastic process for the (true) return on the project once investment is completed can be calculated:

$$\frac{dV}{V} = (r + \eta)dt + \phi dw \tag{14}$$

and finally, the market price of risk (η) can be obtained using the prices of the traded pharmaceutical companies.

The value of the investment project before the investment has been completed also depends on the expected cost of completion, $F(C,K,t)$. In this case no analytical solution exists for the PDE since the time at completion is uncertain:

$$\begin{aligned} \max_I \Bigg[& \frac{1}{2}\phi^2 C^2 F_{CC} + \frac{1}{2}\sigma^2 (IK) F_{KK} + \phi\sigma\rho C (IK)^{\frac{1}{2}} F_{CK} \\ & + \alpha^* C F_C - I F_K + F_t - (r + \lambda)F - I \Bigg] = 0 \end{aligned} \tag{15}$$

subject to the boundary condition at completion of investment: $F(C,0,\tau) = V(C,\tau)$ where τ is a random variable representing the time of project completion. Nevertheless, this problem can be solved through simulation. By approximating the value of the investment project as an American option, the optimal investment and abandonment strategies are determined, i.e., the project will be abandoned when cash flows are too low and/or expected costs to completion are too high.

Miltersen and Schwartz (2004) analyze pharmaceutical R&D investments projects with competitive interactions among research firms. This work concentrates on the competitive interaction and its effect on the valuation and optimal investment strategies. Thus, the real options valuation technique is extended to incorporate game-theoretical concepts. Under this framework, two firms invest in R&D for different drugs, both targeted to cure the same disease. If both firms are successful, there are duopoly profits in the marketing phase. Clearly, future expected profits can affect managerial decisions in the development phase, which in turn affect the outcome in the marketing phase. The implementation of this extended real options framework show that competition in R&D

brings about higher production at lower prices, higher probability of success, and shorter average development time. These benefits are, however, offset by higher total development costs and lower values for R&D investment projects.

Hsu and Schwartz (2008) adapt the pharmaceutical model of R&D valuation to incorporate the design of optimal research incentives to examine the problem of pharmaceutical underinvestment in vaccines to treat diseases such as malaria, tuberculosis, and subtypes of HIV common in Africa. These diseases kill more than 5 million people each year, with almost all of these deaths occuring in the developing regions of the world. Nonetheless, there is a lack of private pharmaceutical investment devoted to researching a cure for these diseases. The lack of pharmaceutical investment can be seen as a small market problem: people in developing countries cannot afford to pay for these drugs. Fortunately, certain international organizations and private foundations (e.g., the World Health Organization and the Gates Foundation) are willing to provide funding. However, "there is no consensus on how to administer the sponsorship effectively." There are currently two main types of subsidy programs (see Kremer 2001, 2002) to encourage pharmaceutical innovation: push subsidy and pull subsidy. Push programs subsidize the cost of R&D projects by providing full discretionary research grants or through sponsor co-payments. In contrast, pull programs subsidize the revenue of R&D projects through fixed-price purchase commitments, variable price purchase commitments, tax incentives, and/or extended patent protection. Clearly, the firm's price and quantity strategy could depend on the incentive program in place and the monopoly power of the firm.

Hsu and Schwartz quantitatively examine the different incentive programs using the real option valuation framework and by explicitly modelling the quality (or efficacy) of the R&D output. The quality of the drug is a key determinant of revenue as it affects demand; moreover, quality directly affects the exercise of the abandonment option. Using this framework, this paper seeks to answer five critical questions: What is the required level of monetary incentive to induce the firm to undertake the R&D to develop the vaccine? What are the expected price, quantity supplied and efficacy of the developed vaccine? What is the probability that a viable vaccine will be developed? What is the consumer surplus generated? What is the expected cost per individual successfully vaccinated?

5. Concluding remarks

This paper provides an overview of the real options approach to valuation mainly from the point of view of the author who has worked in this area for over 30 years. General development of the valuation of complex American options has allowed progress in the solution of many interesting real option problems. Two applications of the real options approach are discussed in more detail: the valuation of natural resource investments and the valuation of research and development investments.

In addition to dicussing the many opportunities provided by the real options approach to valuation, the paper also mentions some of the challenges that the approach presents. These challenges are the subject of continuing research by the author and other scholars.

REFERENCES

Brennan, M. and E.S. Schwartz (1985), "Evaluating natural resource investments," *The Journal of Business* **58**(2): 135–57.

Cortazar, G. and E.S. Schwartz (2003), "Implementing a stochastic model for oil futures prices," *Energy Economics* **25**(3): 215–38.

Hsu, J. and E.S. Schwartz (2008), "A model of R&D valuation and the design of research incentives," *Insurance: Mathematics and Economics* **43**(3): 350–67.

Longstaff, F. and E.S. Schwartz (2001), "Valuing American options by simulation: A simple least-squares approach," *The Review of Financial Studies* **14**(1): 113–47.

Miltersen, K. and E.S. Schwartz (2004), "R&D investments with competitive interactions," NBER Working Paper No. 10,258.

Moon, M. and E.S. Schwartz (2000), "Evaluation of research and development invesments," in M.J. Brennan and L. Trigeorgis, eds., *Innovation, infrastructure and strategic options*. Oxford: Oxford University Press.

Trolle A.B. and E.S. Schwartz (2009) (2009), "Unspanned stochastic volatility and the pricing of commodity derivatives," *Review of Financial Studies* **22**(11): 4423–61.

Schwartz, E.S. (1982), "The pricing of commodity-linked bonds," *The Journal of Finance* **37**(1): 525–39.

Schwartz, E.S. (2004), "Patents and R&D as real options," *Economic Notes* **33**(1): 23–54.

Schwartz, E.S. (2013), "The real options approach to valuation: Challenges and opportunities," presentation at Finance UC 4th International Conference, Pontificia Universidad Católica de Chile.

Schwartz, E.S. and A.B. Trolle (2010), "Pricing expropriation risk in natural resource contracts – A real options approach," in W. Hogan and F. Sturzenegger, eds., *The natural resources trap: Private investment without public commitment.* Cambridge, Mass.: MIT Press.

LAJE
LATIN AMERICAN JOURNAL OF ECONOMICS
VOL. 50 NO. 2 (NOV, 2013), 179–208

FINANCE, GROWTH, AND INSTITUTIONS IN LATIN AMERICA: WHAT ARE THE LINKS?*

LUISA BLANCO**

Using a panel of 16 countries during the 1961-2010 period, we find that financial development has a positive significant effect on economic growth in the long run for high-income countries but a negative significant effect for low-income countries. When studying the determinants of financial development, we find that higher financial openness and lower country risk are associated with greater financial development. The financial risk index has a positive significant effect on financial development, while the economic risk index has a negative significant effect. In addition, lower foreign debt and better socioeconomic conditions increase financial development.

JEL classification: G21, O16, O43, O54

Keywords: Financial development, economic growth, institutions, political stability, Latin America

1. INTRODUCTION

The development of financial markets in Latin American countries in the last two decades is well known. Private credit as share of GDP for the Latin American region rose from an average of 15% in the 1970-1974 period to 33% in the 2006-2010 period.[1] The substantial development of the region's financial sector has generated increased interest in studying the relationship between financial development and economic growth in Latin America.

According to Levine (1997), the financial sector performs several functions that are relevant for economic development. Financial intermediaries help in dealing with risk, which facilitates trading and diversification. The financial sector also has the ability to acquire information and monitor firms and managers, which contributes to efficient allocation of resources. Financial intermediaries also improve resource allocation

* I am thankful to Jeffrey Nugent, William Sawyer, and an anonymous referee for providing excellent comments on my work. I also received helpful comments at seminars at the University of Southern California, Texas Christian University, the Western Economic Association International and Southern Economic Association meetings. All errors are my own.

** School of Public Policy, Pepperdine University and RAND Center for Latin American Social Policy. Email: lblanco@pepperdine.edu.

1. Author's calculation with the data from the sample used in this analysis (16 countries).

doi 10.7764/LAJE.50.2.179

through pooling the savings of individuals, resulting in specialization and greater capital accumulation and productivity.[2]

Although there is a vast amount of work on the finance-growth link, there is no consensus on how financial development affects economic growth. While several theoretical and empirical analyses show that financial development leads to economic growth (Beck, Levine, and Loayza, 2000; Rajan and Zingales, 2003), some provide evidence that financial development has no significant effect on economic growth (Shan, 2005). Others argue that the effect is dependent on certain conditions (Rioja and Valev, 2004a,b) and that financial development may have a negative effect in some cases, depending on the time frame considered (Loayza and Ranciere, 2006). Thus, the study of the finance-growth link continues to be a topic of interest. There is also a growing body of work on the factors that explain financial development.

This paper studies the impact of financial development on economic growth in the short and long run and the determinants of financial development in Latin America. This analysis contributes to the literature in several ways. First, it expands on Loayza and Ranciere's (2006) study of the impact of financial development on economic growth by focusing only on the Latin American region and expanding the sample period. Second, along the lines of the work of Rioja and Valev (2004a), this analysis considers different income groups when determining the long- and short-term effect of financial development on economic growth. Third, in relation to the study of the determinants of financial development, this paper expands on the work of Chin and Ito (2006) and Baltagi *et al.* (2009) by focusing on Latin American countries, expanding the sample period, and considering other factors related to institutions and country stability as possible determinants of financial development.

This paper answers the following questions for the Latin American region: 1) What is the effect of financial development on economic growth for different time frames and across countries with varying income levels? 2) What factors lead to greater financial development?

Studying financial development in Latin America is relevant for several reasons. First, countries in Latin America share a common set of coefficients due to their shared experience, which is not necessarily the case for other regions (Grier and Tullock, 1989). Second, Latin America

2. Please refer to Levine's (1997) work for a more in-depth discussion of the individual functions performed by the financial sector. Levine also provides a good parable that provides a complete understanding of the key role that the financial sector plays in the economy.

is a natural laboratory for studying the impact and determinants of financial development because the region has experienced significant improvements in the financial sector in the last decades. Thus, there is significant variation over time. Third, there is sufficient variation in our variable of interest, financial development, across countries in the region. For example, in the 1970-1974 period, while private credit as a share of GDP is 27% for Mexico, it is only 6% for Bolivia. Then, in the 2006-2010 period, Panama shows the highest level of private credit as a share of GDP (78%), while Argentina is at the bottom (12%).

Using data for the 1961-2010 period in a panel framework for 16 Latin American countries, the main findings in relation to the impact of financial development on economic growth are the following. For the full sample, financial development has a significant positive effect on economic growth in the long run, but a significant negative effect in the short run. This finding agrees with the conclusions of Loayza and Ranciere (2006). Nonetheless, we find that countries in the region do not share the same set of coefficients, such that the long-run positive effect of financial development on economic growth only holds for the high-income group. For the low-income group, financial development has a significant negative effect in the long run. Financial development has no significant effect on economic growth in the short run for either the high- or low-income group.

In the analysis of the determinants of financial development, using 5-year average observations during the period 1985-2010, greater financial openness and lower country risk are associated with greater financial development. Financial openness seems to create the most significant benefit in those countries that are relatively closed. Of the components of the country risk index (financial, economic, and political), the financial risk index has a positive significant effect, while the economic risk index has a negative significant effect. Of the components of the financial risk index, the index related to foreign debt is positive and statistically significant (lower foreign debt as a share of GDP is associated with greater financial development). Of the components of the political risk index, only the socioeconomic conditions index has a positive significant effect on financial development at the 5% level, while the indices related to internal conflict, government stability, and investment profile are positive and marginally statistically significant (10% level). None of the components of the economic risk index show a significant effect.

This paper is organized as follows. Section 2 presents a brief review of literature on the finance-growth link and the determinants of financial development, and Section 3 describes the methodology. Sections 4 and 5 present the results and a discussion of sample issues and main findings, respectively. Section 6 concludes.

2. Literature review

2.1. The finance-growth link

While the general belief is that financial development has a positive effect on economic growth (supply-leading hypothesis), there is theoretical and empirical work indicating that this effect is non-existent and that financial development is merely a consequence of economic growth (demand-following hypothesis).[3] Financial development can be generally defined as increasing access to credit, and the positive effect of financial development on growth is derived from the effect financial development has on capital accumulation and productivity (Beck, Levine, and Loayza, 2000). With the development of the financial sector comes greater access to capital that results in more funding available for attractive investment opportunities. Greater access to capital leads to increased labor specialization and more access to new technology (Rajan and Zingales, 2003; Saint-Paul, 1992). Consequently, improvements in capital markets lead to greater economic growth.

On the other hand, there has been some questioning of the benefits derived from financial development. There are three main reasons to be skeptical about the impact of financial development on economic growth. First, there is research that supports the demand-following hypothesis, where financial development is a consequence of economic growth (Shan, 2005). Second, the impact of financial development on economic growth seems to be dependent on certain conditions. There is empirical evidence showing that the effect of financial development on growth is different across regions and among countries with different income levels, levels of financial development, and institutional frameworks (see Aghion *et al.*, 2005; Blanco, 2009; De Gregorio and Guidotti, 1995; Rioja and Valev, 2004a,b; and Shen and Lee, 2006, among others). Third, financial development can produce greater macroeconomic volatility, becoming a destabilizing force in the economy (Loayza and Ranciere, 2006).

3. Refer to Blanco (2009) and Levine (2005) for a thorough discussion of the literature on the finance-growth link. Odhiambo (2007) presents a good discussion on the supply-leading and demand-following hypotheses.

When financial development leads to volatility, it is expected that financial development will have a negative effect on economic growth. According to Loayza and Ranciere (2006), the short-run ef fect of financial development on economic growth may be negative due macroeconomic instability, but the long-run ef fect is expected to be positive. Thus, looking at the impact of financial development at different time frames is necessary.

In the Latin American context, where countries have experienced periods of volatility, distinguishing the short- and long-run effect of financial development is of special interest to policymakers. When studying the impact of financial development on economic growth, it is also important to keep in mind that financial development might have a differential impact on growth depending on specific country conditions. Some countries will be better equipped to absorb the influx of credit. It is likely that specific country characteristics, in relation to their level of development (i.e., income) could determine a country's ability to use the influx of credit productively. For this reason, studying the impact of financial development for countries with different income levels is relevant for the design of future policies related to financial markets in Latin America.

2.2. Sources of finance

In the review of the literature, the factors considered to be the main determinants of financial development are the degree of openness, institutions, and political stability. Liberalization of goods and capital markets are associated with greater financial development (Baltagi *et al.*, 2009; Chinn and Ito, 2006; Klein and Olivei, 2008). Openness to trade and capital flows have been proposed as important determinants of financial development. According to Rajan and Zingales (2003), there will be interest groups who will oppose financial development due to the competition it brings. With trade and financial liberalization, the power of those groups opposed to financial development is significantly weakened. Therefore, substantial financial reforms can take place when the power of such interest groups is diminished by openness, leading to greater financial development.

Financial liberalization is associated with the strengthening of the financial system in two ways.[4] First, as a result of financial

4. Refer to Chinn and Ito (2006) and Klein and Olivei (2008) for a comprehensive literature review of the channels through which financial liberalization leads to greater financial development.

liberalization, the entrance of foreign banks into the domestic financial sector leads to an increase in available loanable funds and efficiency. Efficiency in the financial sector increases significantly with financial liberalization since there is greater competition and more pressure to reform the financial sector. Second, Klein and Olivei (2008) argue that a virtuous cycle of greater savings and efficiency is created with increasing capital account openness because financial intermediaries are able to achieve economies of scale.

Furthermore, institutions seem to play a key role in explaining differences in financial development across countries.[5] According to Chinn and Ito (2006), there are two different categories of institutions that have been considered important determinants of financial development: 1) institutions that affect the economy as a whole, and 2) institutions that affect the financial sector.[6] In the first group, the relevant institutions are related to bureaucratic quality, law and order, and control of corruption, among others. Because these institutional factors directly affect the way business is done and relate to perceptions about the stability of the legal system, they are expected to be associated with greater levels of financial development.

The second group of institutions includes those that specifically affect the financial sector. According to Djankov *et al.* (2007), institutions that increase creditor power and access to lending information are crucial for financial development. When creditor rights are enforced, credit is likely to expand because creditors feel more protected against default. Creditors are also more likely to lend when they are able to get more information about potential lenders. Greater financial depth is expected when there is an increase in access to information on borrowers and protection for private credit institutions.

Furthermore, the stability of a specific country may significantly influence capital markets. The degree to which there is stability in a country affects investors' perceptions and consequently their willingness to invest in that country. According to Roe and Siegel (2009), a country's capacity to protect investors is related to political

5. Beck and Levine (2005) present an excellent review of the literature on the relationship between institutions and financial development.

6. Here we follow the categorization provided by Chinn and Ito (2006) to distinguish the different types of institutions. Other institutions related to the financial sector could be those that help to promote stability in the financial sector. Institutions that help promote stability are likely to be related to the design and enforcement of prudent regulations. Data about these types of institutions are unlikely to be available consistently over time for the sample used in this analysis.

stability. Thus, countries with unstable political systems offer low protection to investors.

Empirical evidence on the importance of openness and institutions as factors explaining financial development is abundant. The cross-sectional analysis by Herger *et al.* (2008) shows that trade openness has a significant effect on financial development. In a panel framework that includes only less developed countries, Baltagi *et al.* (2009) find that trade and financial openness are relevant to explaining financial development. They investigate the interactions between trade and financial openness and find that this interaction term is negative. They conclude that while financial development requires both types of openness, relatively closed economies benefit the most from opening up to trade or capital. Chinn and Ito (2006) find that at a certain institutional threshold, financial liberalization has a positive effect on financial development. Results from Klein and Olivei (2008) are along the lines of Chinn and Ito's (2006) findings. Klein and Olivei (2008) find that institutions drive the positive effect of financial liberalization on financial development, where developed countries that have better institutions obtain greater benefits from financial liberalization. The openness to trade and capital flows experienced during the process of globalization are likely to be associated with institutional reforms that significantly affect capital markets (Mishkin, 2009).

There is also empirical evidence regarding the impact of institutions and political stability on financial development. Acemoglu and Johnson (2005) find that institutions that affect all sectors of the economy have a significant direct effect on financial development. They show empirically that property rights and contracting institutions are important determinants of financial development. Beck *et al.* (2003) also find that institutions, shaped by either legal origins or initial resource endowments, have a significant effect on financial development in a sample of 70 former colonies. Andrianova *et al.* (2008) report evidence that institutions related to governance have a significant effect on financial development, where lower quality of institutions is associated with greater government ownership in the financial sector. In relation to institutions that affect capital markets, Djankov *et al.* (2007) present strong empirical evidence that creditor rights and access to lending information are important determinants of financial development. Additionally, Roe and Siegel (2009) present empirical evidence showing that political instability explains financial backwardness.

While there are several papers on the determinants of financial development, few have taken a regional approach. When studying the factors that lead to greater financial depth, it is important to focus on countries with a common historical, political, and socio-economic background. It is unlikely that the factors that explain financial development in a specific country in Asia or Africa would explain capital markets in Latin America. By taking a regional approach to the study of the sources of finance, more specific policy recommendations could be provided.

3. Methodology

3.1. Impact of financial development on economic growth

In studying the impact of financial development on economic growth in the short and long run for Latin America, this analysis follows the methodology of Loayza and Ranciere (2006) closely. Loayza and Ranciere (2006) propose using the pooled mean group (PMG) estimator developed by Pesaran *et al.* (1999).[7] For the PMG estimator, an autoregressive distributive lag (ARDL(p,q,q,...,q)) dynamic panel specification is applied. A vector error correction model (VECM) is considered under this specification, where the short-run dynamics of the variables in the system are influenced by the deviation from equilibrium. The ARDL(p,q,q,...,q) used for the PMG estimator is specified as follows:

$$y_{it} = \sum_{j=1}^{p} \lambda_{ij} y_{i,t-j} + \sum_{j=0}^{q} \delta_{ij}^{'} \mathrm{X}_{i,t-j} + \mu_i + \varepsilon_{it} \tag{1}$$

where y_{it} represents the dependent variable for $t = 1, 2,...,$T time periods, and $i = 1, 2, ..., N$ groups. $X_{i,t-j}$ is the k x 1 vector of explanatory variables (regressors) for group i, δ_{ij} are k x 1 coefficient vectors, λ_{ij} are scalars, μ_i represents the fixed effect, and ε_{it} the time varying disturbance. Equation (1) can be reparametrized in the following way and time series observations for each group are stacked

7. Refer to Loayza and Ranciere (2006) for an explanation of the appropriateness of the PMG estimator when disentangling the finance-growth link and a description of this methodology. Refer also to Blackburne and Frank (2007) for a description of the PMG estimator in Stata.

$$\Delta y_i = \phi_i y_{i,-1} + \mathrm{X}_i \beta_i + \sum_{j=1}^{p-1} \lambda_{ij}^* \Delta y_{i,-j} + \sum_{j=1}^{q-1} \Delta \mathrm{X}_{i,-j} \delta_{ij}^* + \mu_i \iota + \varepsilon_i \quad (2)$$

where y_i is a t x 1 vector of the observations of the dependent variable of the ith group, X_i is a t x k matrix of the regressors that vary across groups and time periods, and ι is a t x 1 vector of 1s. One of the main requirements of this model's specification is the existence of a long-run relationship between y_{it} and X_{it}, where the error-correcting speed of adjustment term for the long-run relationship represented by ϕ_i must be significantly negative (and no lower than -2). The long-run relationship between y_{it} and X_{it} for each group is expressed as follows:

$$y_{it} = -(\beta_i^{'} / \phi_i)\mathrm{X}_{it} + \eta_{it} \quad (3)$$

where η is a stationary process. For the long-run homogeneity assumption, the coefficients on X_i are the same across groups. Long-run coefficients of X_i are expressed as $\theta_i = -\beta_i/\phi_i$, where $\theta_i = \theta$. In the PMG estimator, while the long-run coefficients are equal across groups, the intercept, short-run coefficients, and error variances differ across countries.[8]

For the PMG estimation in this analysis, real GDP growth (first difference of the natural log of real GDP per capita) is the dependent variable and financial development (private credit in natural logs) is in the right-hand side of the equation.[9] Initial GDP per capita (natural log), government size (natural log), trade, and inflation are included as control variables.[10] A dynamic specification of the form ARDL(3,3,1,1,1,1) is used, and all variables are time-demeaned.[11] All

8. See Blackburne and Frank (2007) for a good explanation of the specification of the PMG model. Asteriou and Hall (2007) also provide a brief discussion of the PMG estimator.

9. The methodology used here, following Loayza and Ranciere (2006, p. 1055), addresses the two-way causality between financial development and economic growth. One of the conditions for validity is that "the dynamic specification of the model be sufficiently augmented so that the regressors are strictly exogenous and the resulting residual is serially uncorrelated."

10. These variables are constructed following the approach of Loayza and Ranciere (2006); refer to Table A1 for a description of how these variables were constructed.

11. Lag lengths selected based on the augmented Dickey-Fuller (ADF) regressions. The number of lags is selected in such a way that the Akaike information criterion (AIC) for the regression is minimized. This process is carried out for each panel. The PMG estimator provides consistent estimators when the variables are I(0) and I(1), so there is no need to include unit root tests in the analysis.

independent variables are entered in levels for the long-run relationships and in first difference for the short-run relationships. The ARDL form specified above includes the first and second lag of the first difference of real GDP and private credit as regressors. Annual observations between 1961 and 2010 are used for this part of the analysis. Because of the lag structure of the model, estimations will include observations between 1964 and 2010 (47 observations per country). Table 1 shows the summary statistics, and Table 1 in the appendix provides a description of the variables used and their sources.

3.2. Determinants of financial development

In this analysis, the approach taken to find out what factors explain financial development in Latin America is similar to the one used by Baltagi *et al.* (2009). The dynamic panel general method of moments (GMM) suggested by Arellano and Bond (AB, 1991) is implemented and an ARDL(p,q,q,...,q) specification is considered for the AB estimator. For the AB estimator, the first lag of the dependent variable is included in the right-hand side of the equation, which leads to endogeneity issues since the lag of the dependent variable is determined by the error term. This endogeneity problem biases the estimates provided by the general GMM. Arellano and Bond (1991) propose differencing the data to address the endogeneity of the variables on the right-hand side and control for specific country characteristics.[12] The Arellano and Bond (1991) GMM uses lagged levels of the dependent variable as instruments to address the endogeneity of the dependent variable. The model specification of the AB estimator can be expressed as:

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta \mathrm{X}_{it} \beta_i + \Delta \varepsilon_{it} \tag{4}$$

Equation (4), which represents first difference transformation and removes the constant term and individual effects, shows that the lag of the dependent variable is included as a regressor and X_{it} is the tN x k matrix of the explanatory variables. For this estimation, the instruments used are the available lags of the levels of the endogenous variables.

12. Using Arellano and Bond's GMM estimator allows us to see the effect of the factors considered as determinants of financial development when addressing endogeneity. This methodology requires us to take the first difference of the dependent variable, which can be interpreted as the growth of finance, thus enabling us to see whether the independent variables are associated with changes in financial development in Latin America.

Table 1. Impact of financial development on growth[a]

(summary statistics)

	Obs.	Mean	Std. dev.	Min.	Max.
ln(GDP per capita)	800	-0.007	0.385	-0.768	0.820
ln(Finc Dev)	800	0.038	0.545	-2.116	1.314
ln(Initial GDP per cap)	800	-0.014	0.388	-0.780	0.784
ln(Government Size)	800	-0.074	0.550	-1.139	1.100
Trade	800	0.036	0.429	-1.090	1.108
Inflation	800	0.015	0.364	-0.627	4.186

a. Annual observation, 1961-2010, 16 countries (statistics on time-demeaned data).

The methodology of Arellano and Bond (1991) is appropriate for datasets with many panels and few periods. For this reason, and to smooth out short-run fluctuations in the data, five-year average observations are considered in this part of the analysis. These five-year average observations are constructed using available data for the period from 1985 to 2010.[13] Financial development growth (the first difference of private credit as a share of GDP in natural log) is used as the dependent variable, and its first lag is entered in the right-hand side of the equation. The growth of real GDP per capita (first difference of real GDP per capita in natural log) and a dummy for the banking crisis are included as control variables.[14]

The variables of interest that are entered in the right-hand side of the equation are trade openness (natural log), financial openness, the interaction between trade and financial openness, and the country risk index.[15] The country risk index is a composite indicator of political,

13. Five-year averages are based on available data for the periods 1985-89, 1990-94, etc.

14. Note that private credit and real GDP are entered in first difference initially as we are interested in considering the relationship between the growth rates of these variables. It is also important to note that the methodology used here allows for dealing with the two-way causality between finance and growth since this estimation method "is suited to panel data, deals with a dynamic regression specification, controls for unobserved time- and country-specific effects, and accounts for some endogeneity in the explanatory variables." (Loayza and Ranciere, 2006: 1067)

15. This analysis focuses on testing empirically the effect of financial openness on financial development, which is related to the liberalization of the capital account. Financial liberalization is defined by Ranciere *et al.* (2008) as the deregulation of domestic financial markets, in addition to liberalization of the capital account. Financial openness and financial liberalization terms are used interchangeably in the literature, but it is important to make the distinction when performing empirical analyses. For example, Abiad and Mody (2005) and Abiad *et al.* (2008) construct an index of financial liberalization that focuses on financial reform and they present an analysis of the factors explaining it. Chinn and Ito's (2008) financial openness index, which is used in this analysis, is related only to liberalization of the capital account.

Table 2. Determinants of financial development[a]

(summary statistics)

	Obs.	Mean	Std. dev.	Min.	Max.
ln(Finc dev)	144	8.589	0.420	7.561	9.396
ln(GDP per capita)	144	3.030	0.576	1.525	4.407
Banking crisis	144	0.094	0.223	0.000	1.000
ln(Trade openness)	144	3.881	0.618	2.454	5.263
Financial openness	144	0.203	1.481	-1.856	2.456
Trade open*Finc open	144	1.159	5.972	-7.866	12.924
Country risk	96	61.574	10.922	29.380	79.860
Financial risk	96	31.572	8.129	8.230	42.684
Political risk	96	59.769	11.268	31.415	79.358
Economic risk	96	31.670	5.321	13.335	40.860
Foreign debt (% of GDP)	96	5.642	1.713	1.380	9.000
Government stability	96	6.804	1.780	2.500	9.768
Internal conflict	96	7.990	2.359	0.450	11.066
Investment profile	96	6.633	2.146	2.250	11.500
Socioeconomic conditions	96	5.061	1.340	1.994	7.860

a. 5-year average observations, 1970-2010, 16 countries

financial, and economic risk indices. Thus, the model is estimated by including the components of the country risk index.[16] We also estimate the model with the components of the economic, financial, and political risk indices. In the model specification shown in Equation (4), the first difference is taken from all variables to transform the equation into the difference GMM. The lagged levels of financial development growth are used to form GMM-type instruments. Table 2 shows the summary statistics for this part of the analysis, and Table A1 in the appendix provides a description of the variables.

4. Results

4.1. Financial development's impact on economic growth

Table 3 presents the estimates obtained when using the PMG estimator to determine the short- and long-run effect of financial development on economic growth for the full sample. The first two columns show the

16. Refer to Table A1 in the appendix for a description of how the country risk index is constructed and its components.

Table 3. Impact of financial development on economic growth *(pooled mean group estimator)*

Variables	All countries		High-income countries		Low-income countries	
	Coeff.	Std. error	Coeff.	Std. error	Coeff.	Std. error
Long-run coefficients						
Financial development	0.076 ***	0.015	0.079 ***	0.016	-0.223 ***	0.084
Initial GDP per capita	0.525 ***	0.109	0.488 ***	0.126	0.067	0.254
Government size	0.181 ***	0.051	0.194 ***	0.054	0.147 *	0.088
Trade	0.235 ***	0.050	0.239 ***	0.053	0.275 ***	0.085
Inflation	-0.107 ***	0.040	-0.092 **	0.040	-0.492 ***	0.151
Error-correction coefficient - ϕ	-0.108 ***	0.037	-0.153 **	0.068	-0.103 **	0.045
Short-run coefficients						
d(GDP per capita)$_{t\text{-}1}$	0.173 ***	0.045	0.211 ***	0.053	0.130 *	0.077
d(GDP per capita)$_{t\text{-}2}$	0.005	0.042	0.047	0.040	-0.022	0.072
d(Financial development)$_t$	-0.036 **	0.018	-0.032	0.021	-0.035	0.030
d(Financial development)$_{t\text{-}1}$	0.006	0.010	-0.004	0.012	0.020	0.017
d(Financial development)$_{t\text{-}2}$	-0.016	0.017	-0.024	0.026	0.003	0.021
d(Initial GDP per capita)$_t$	0.040	0.039	0.001	0.069	0.075	0.048
d(Government size)$_t$	-0.247 ***	0.031	-0.252 ***	0.044	-0.243 ***	0.049
d(Trade)$_t$	0.045 *	0.027	0.049	0.032	0.019	0.051
d(Inflation)$_t$	-0.059 ***	0.016	-0.037 **	0.015	-0.058 **	0.028
Intercept	0.004	0.012	0.032 **	0.014	-0.055	0.037
No. of countries	16		8		8	
No. of observations	752		376		376	
Log likelihood	1717		855.6		874.8	

*, **, and *** indicate significance at the 10, 5 and 1% levels, respectively. All estimations include 47 observations per country.

coefficients and the standard errors for the full sample. In this estimation, the long-run coefficients of all control variables are significant at the 1% level. The coefficients for initial GDP per capita and government size are different than expected, but trade and inflation have the expected signs. For the short-run estimates, all control variables except for initial GDP per capita and trade are statistically significant at the 5% level. Only the coefficient sign for initial GDP per capita is unexpected, but it is not statistically significant. The first lag of the dependent variable is positive and statistically significant at the 1% level.

For the full sample, financial development has a positive significant effect at the 1% level on economic growth in the long run. For the short run, financial development has a negative effect, where only its first difference is statistically significant at the 5% level. The first difference of the first and second lag of financial development have positive and negative coefficients, but they are not statistically significant. The positive and negative effect in the long and short run respectively agrees with the finding of Loayza and Ranciere (2006). The Hausman test was performed to ensure that the PMG estimates are preferred to the ones obtained from the mean group (MG) estimator, where the MG estimator fits the model separately for each group. The Hausman test provides evidence that MG estimates are preferred since it rejects the hypothesis that the difference in coefficients is not systematic for the full sample. Thus, the homogeneity restriction is rejected jointly for all parameters.

Following the approaches of Rioja and Valev (2004a) and Blanco (2009), this analysis also evaluates the possibility that the effect of financial development is different across different income groups. This is also an appropriate approach based on the Hausman test results, which suggest that the PMG is not suitable for the full sample. Based on countries' real GDP per capita in the middle of the sample period (in 1986), the sample is divided into high- and low-income countries. The countries in the high-income group are Argentina, Colombia, Costa Rica, Mexico, Panama, Peru, Uruguay, and Venezuela. The countries in the low-income group are Bolivia, Chile, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, and Paraguay.[17]

In Table 3, columns 3 and 4 present the coefficients and standard errors for the high-income group, and columns 5 and 6 show estimates for the low-income group. For the high-income group the signs and significance of most coefficients stay the same. Financial development shows a significant positive effect in the long run at the 1% level, but has no significant effect in the short run. In the low-income group,

17. After the division there are eight countries in each group, which is just enough to estimate the PMG. The case of Chile is interesting since it is classified as low-income in this study based on 1986 income levels, even though Chile today has one of the highest income levels in the region. Classifying the countries in two income groups rather than three (high-, middle-, and low-income, as Blanco (2009) does) is somewhat restrictive in this set-up, but it is necessary to maintain the properties of the PMG estimator since splitting the sample into three categories would result in a very small sample size when estimating the model for each group. Using the income levels in the middle of the sample when classifying countries provides us with a more consistent classification that is not biased by a posteriori knowledge. Interestingly, if we add Chile to the high-income group, we find from the Hausman test that we reject the null hypothesis that the difference in coefficients is not systematic (which is not the case when Chile is excluded from this group). Thus, Chile does not appear to belong to the high-income group when we use the PMG estimator, so keeping it in the low-income group seems appropriate for this study.

the significance and sign of the coefficients change dramatically. In this estimation, financial development shows a negative significant effect in the long run at the 1% level, but no significant effect on economic growth in the short run. The Hausman test was performed to ensure that the PMG estimates are preferred to the ones obtained from the MG estimator for the high- and low-income subsamples. We fail to reject the hypothesis that the difference in coefficients is not systematic for both subsamples, which leads us to conclude that the PMG estimates are preferred over the MG estimates. It is also important to note that the condition for the error-correction speed of adjustment is met in all estimations, where ϕ_i is statistically significant with a negative value greater than -2.[18]

4.2. Determinants of financial development

Tables 4 and 5 contain estimates of the model of determinants of financial development in Latin America. In Table 4, the first two columns show the coefficients and standard errors for the baseline model that includes the composite risk index, which accounts for economic, financial, and political risk. Higher values of this index represent lower risk, more stability, and a better institutional environment. All estimations in this section include time dummies, but these estimates are not included due to space considerations.

In this estimation, real GDP growth has a negative, marginally significant effect at the 10% level, which was unexpected. Banking crisis has a positive, marginally significant effect at the 10% level, and its sign was also unexpected. One possible reason for the positive sign of this coefficient is that this indicator may capture the period of time in which financial sector restructuring takes place. It is difficult to detangle the effect of the banking crisis dummy since a dummy in a five-year period could account for the pre- and post-crisis period. Trade openness has a positive sign, but it is not statistically significant, which was unexpected. Financial openness and the interaction term between financial and trade openness are significant at the 5% level. While the coefficient for financial openness has a positive sign, its interactive term with trade openness is negative. This finding agrees

18. The positive coefficient of initial GDP per capita is unexpected according to convergence theory, and this is noted in the paper. However, this should not affect the stability of the model since initial GDP per capita is entered as an exogenous variable in the model and the condition for the error-correction speed of adjustment is met in all estimations.

Table 4. Sources of financial development

(Arellano and Bond estimator)

Variables	Model 1		Model 2		Model 3	
	Coeff.	Std. error	Coeff.	Std. error	Coeff.	Std. error
d(Financial development growth)$_{t\text{-}1}$	0.130	0.114	0.046	0.124	0.006	0.130
d(GDP per capita growth)$_t$	-1.154 *	0.614	-1.001 *	0.605	-1.382 **	0.537
d(Banking crisis)$_t$	0.367 *	0.201	0.407 *	0.209	0.432 **	0.177
d(Trade openness)$_t$	0.432	0.351	0.650 *	0.358	0.159	0.395
d(Financial openness)$_t$	0.757 **	0.364	0.887 ***	0.316	0.797 **	0.355
d(Trade open)$_t$* d(Finc open)$_t$	-0.193 **	0.087	-0.220 ***	0.076	-0.199 **	0.086
d(Country risk)$_t$	0.027 *	0.015				
d(Economic risk)$_t$			-0.043 ***	0.015		
d(Financial risk)$_t$			0.041 ***	0.010		
d(Political risk)$_t$			0.013	0.011		
d(Foreign debt)$_t$					0.165 ***	0.041
No. of countries	16		16		16	
No. of observations	80		80		80	
No. of time periods	5		5		5	
Sargan test (p-value)	20.545	(0.36)	24.672	(0.17)	23.100	(0.23)
1st-order serial correl. test (*p*-value)	-2.160	(0.03)	-2.396	(0.02)	-2.262	(0.02)
2nd-order serial correl. test (*p*-value)	-1.398	(0.16)	-1.563	(0.12)	-1.491	(0.14)

Robust standard errors provided. *, **, and *** indicate significance at 10, 5 and 1% levels, respectively. Time dummies included in all estimations (estimations not included due to space considerations). Sample uses 5-year averages with available observations from 1981-2010, and all estimations include 5 observations per country.

Table 5. Sources of financial development

(Arellano and Bond estimator)

Variables	Model 1		Model 2		Model 3		Model 4	
	Coeff.	Std. error	Coeff.	Std. error	Coeff.	Std. error	Coeff.	Std. error
d(Financial development growth)$_{t-1}$	0.141	0.140	0.116	0.134	0.127	0.119	0.054	0.117
d(GDP per capita growth)$_t$	-0.844	0.539	-0.81	0.496	-0.902 *	0.483	-0.895 *	0.498
d(Banking crisis)$_t$	0.303	0.206	0.3	0.210	0.383 *	0.205	0.490 ***	0.174
d(Trade openness)$_t$	0.586 *	0.346	0.496	0.360	0.415	0.410	0.327	0.397
d(Financial openness)$_t$	0.811 **	0.342	0.966 ***	0.370	0.971 ***	0.366	0.915 **	0.372
d(Trade open)$_t$* d(Finc open)$_t$	-0.205 **	0.082	-0.235 ***	0.090	-0.240 ***	0.090	-0.226 **	0.093
d(Internal conflict)$_t$	0.082 *	0.047						
d(Government stability)$_t$			0.102 *	0.057				
d(Investment profile)$_t$					0.075 *	0.045		
d(Socioeconomic cond)$_t$							0.153 ***	0.055
No. of countries	16		16		16		16	
No. of observations	80		80		80		80	
No. of time periods	5		5		5		5	
Sargan test (p-value)	19.675	(0.414)	26.753	(0.111)	24.140	(0.191)	20.615	(0.359)
1st- order serial correl. test (p-value)	-2.278	(0.023)	-2.258	(0.024)	-2.148	(0.032)	-1.972	(0.049)
2nd- order serial correl. test (p-value)	-1.141	(0.254)	-1.354	(0.176)	-1.542	(0.123)	-1.746	(0.081)

Robust standard errors provided. *, **, and *** indicate significance at 10, 5 and 1% levels, respectively. Time dummies included in all estimations (estimations not included due to space considerations). Sample uses 5-year averages with available observations from 1981-2010, and all estimations include 5 observations per country.

with Baltagi *et al.* (2009). The negative coefficient of the interaction term implies that the effect of capital openness on financial development will be greater for relatively closed economies than for relatively open economies. The country risk index has a positive sign and is marginally significant at the 10% level.

From the estimates shown in Table 4, it is apparent that the lag of the dependent variable is not statistically significant. This raises the question of whether the dynamic model panel approach, where the lagged dependent variable is included as a regressor, is the adequate model. A lag length test provides evidence that one lag of the financial development growth indicator is the adequate number of lags.[19]

In Table 4, columns 3 and 4 (Model 2) provide the estimates obtained when the components of the country risk index (economic, financial, and political risk indices) are included. These estimates show that only the economic and financial risk indices have a significant effect on financial development at the 1% level. While the economic index has a negative sign, the financial index has a positive sign. This suggests that a decrease in financial risk (higher index value) is beneficial for financial development, but a decrease in economic risk (higher index value) is detrimental to financial development. The effect of the economic index on financial development was not predicted, as more stable economic conditions would be expected to be more conducive to expansion of the financial sector.

We explore whether the components of the three different indices have a significant effect on financial development, and estimate our model, examining each component one at a time to avoid multicollinearity issues. We estimated our model 22 times (the economic risk index has five components, as does the financial risk index, while the political risk index has 12), and include in our tables those estimations where we find that the components of the indices are statistically significant at least at the 10% level.[20] We provide a description of the variables that compose these indices and were statistically significant in our estimations in the Appendix (Table A1).[21]

19. Lag length selected using the ADF regressions, where the regression that minimizes the AIC is chosen (in a panel set-up).

20. Other estimations that include the other components of the risk indices, one at the time, are not included for space considerations but are available from the author upon request.

21. Please refer to the Political Risk Group website for a discussion of the 22 variables that compose the different risk indices (http://www.prsgroup.com/ICRG_methodology.aspx).

The economic index measures a country's economic strengths and weaknesses and is composed by indices of GDP per capita, economic growth, inflation rate, budget balance as a percentage of GDP, and current account as a percentage of GDP. We find that none of these components were statistically significant in our model.

The financial risk index indicates the ability of a country to meet its financial obligations, such as official, commercial, and trade debt obligations. This index is composed of several indices that are related to foreign debt as a percentage of GDP, foreign debt service as a percentage of exports of goods and services, current account as a percentage of exports of goods and services, net international liquidity as months of import cover, and exchange rate stability. From these indices, interestingly, only the index related to foreign debt as a percentage of GDP was statistically significant. Columns 5 and 6 in Table 4 (Model 3) show the estimates obtained when we include the index related to foreign debt as a percentage of GDP, which is positive and statistically significant at the 1% level. A decrease in foreign debt as a percentage of GDP is associated with a higher index, and consequently with a higher level of financial development.

The model specified in Equation (4) is also estimated using the 12 components of the political risk index, one at a time. The components of the political risk index, which are closely related to institutions and country stability, are the following: government stability, socioeconomic conditions, investment profile, internal conflict, external conflict, corruption, military in politics, religious tensions, law and order, ethnic tensions, democratic accountability, and bureaucratic quality. The indicators that account for institutions that affect the economy as a whole and that are included in the political risk index are corruption, law and order, and bureaucratic quality. The investment profile index is the indicator that accounts for institutions that directly affect the financial sector since it is composed of indicators related to contract viability, expropriation, profits repatriation, and payment delays. A close relationship is expected between the investment profile index and our financial development indicator since the investment profile is related to investment risk and consequently to the willingness to invest in a specific country. Thus, there is important feedback between these two indicators, and it is expected that the AB estimator will allow for estimating the independent effect of financial sector institutions on financial development.

Four components are statistically significant when the model is estimated by including each component of the political risk index one at a time; the estimations are shown in Table 5. In that table, internal conflict (columns 1 and 2), government stability (columns 3 and 4), and investment profile (columns 5 and 6) have a positive significant effect on financial development at the 10% level. The index of socioeconomic conditions is positive and statistically significant at the 1% level (Table 5, columns 7 and 8). This index measures the degree to which socioeconomic pressures related to unemployment, consumer confidence, and poverty constrain government actions or fuel social dissatisfaction.[22]

5. Discussion

In relation to the sample used in this analysis, which is restricted to 16 Latin American countries, we consider the possibility of expanding our sample to include three other countries for which there is consistent financial development data available during the period of analysis: Jamaica, Haiti, and Trinidad and Tobago. However, because we believe that these countries do not share the same historical legacy and socioeconomic background as the countries included in the main sample, we did not initially include them in the main estimations. The excluded countries are also not Spanish-speaking countries. Trinidad and Tobago is classified as a high-income economy by the World Bank, where all the other countries in our sample are classified as developing countries.[23]

Although Jamaica, Haiti, and Trinidad and Tobago generally cannot be considered developing Latin American countries, we explore their inclusion in the estimations performed in this analysis. We estimate the model specified in Equation (2), which estimates the long- and short-run effect of financial development on economic growth using the PMG and including these three countries. We find that when these countries are included in the full sample, financial development continues to have a positive effect on economic growth in the long run

22. Note that in all the estimations the Sargan test shows that the instruments used are adequate since the hypothesis that the overindentifying restrictions are valid is not rejected. The serial correlation tests also show that the idiosyncratic errors are independently and identically distributed (i.i.d.) as required for the AB estimation. In all AB estimations we also meet the conditions of rejecting first-order autocorrelation and not rejecting the second-order autocorrelation at the 5% level.

23. We refer to the latest country classification provided by the World Bank (http://data.worldbank.org/about/country-classifications/country-and-lending-groups#LAC).

at the 1% level, and the coefficient is of the same magnitude as is shown in Table 3, column 1. We also find that financial development has a significant negative effect in the short run, where the coefficient of the first difference of private credit is negative and statistically significant at the 1% level, and of the same magnitude as before. Nonetheless, we find in the Hausman test that when these countries are included in the estimation, we reject the hypothesis that the PMG estimates are appropriate, which tells us that these countries do not share the same set of coefficients as the other countries.

We also explore whether the results we found in relation to the high- and low-income groups are robust to the inclusion of these countries. We again classify countries by their 1986 income level, where Trinidad and Tobago and Jamaica are added to the high-income group and Haiti to the low-income group. The estimations obtained here are very similar to those shown in Table 3. For the high-income group, with the inclusion of these two countries, financial development also has a positive significant effect in the long run at the 1% level. For this subsample, the Hausman test tells us that the PMG is preferred over the MG since we reject the hypothesis that the difference in coefficients is not systematic. Thus, based on the Hausman test, excluding these countries from the estimation is appropriate.

For the low-income group, when Haiti is added we find that financial development no longer has a significant negative effect on economic growth in the long run as was found previously; in this estimation the long-run coefficient of private credit is positive and statistically significant. For this subsample we find that the PMG is preferred over the MG, but we find that the condition for the error-correction speed of adjustment is not met in this estimation, where ϕ_i is positive and statistically insignificant. Thus, we can also conclude here that the inclusion of Haiti in the estimation may not be appropriate.

From the estimations related to the effect of financial development in the long and short run, we can summarize the main findings as follows. First, the effect of financial development on economic growth is different across different income groups. Thus, examining the impact of financial development for the whole region may not be appropriate as countries do not share the same set of coefficients in relation to the finance-growth relationship. Second, the impact of financial development for the different subsamples is of small magnitude and varies according to the different income groups. Using the coefficients shown in Table 3, column 3, a 1% increase in private credit is associated with a 0.08% increase in economic growth in the long run for the high-income group. For the low-income group, using the coefficients shown in Table 3, column 5, a 1% increase in private credit is associated with a long-run decrease in economic growth of 0.22%. It is interesting to note that we do not find evidence of short-run effects, which goes along with Bangake and Eggoh's (2011) finding. Our analysis here supports the claim by Bangake and Eggoh (2011) that these countries should focus on implementing long-run policies.

We also consider including Jamaica, Haiti, and Trinidad and Tobago in the estimations in which we model the determinants of financial development for Latin American countries. In these estimations, we find similar results to those shown in Tables 4 and 5. The only difference is that when including these countries, only the indices related to investment profile and socioeconomic conditions are positive and significant at the 1 and 5% levels, respectively. The index related to foreign debt continues to be positive and statistically significant at the 1% level. Financial openness and its interaction with trade openness are also statistically significant in all these estimations and have the same signs as before.

From the estimations related to the determinants of financial development, we can summarize our findings in the following way. First, financial openness has a robust, positive effect on financial development, while its interaction with trade openness has a robust, negative significant effect. Financial openness seems to be the key player in explaining financial development, which may be because of the sample period used. This analysis encompasses the 1985-2010 period, during which financial markets opened up significantly in Latin America. In fact, the standard deviation for the index of financial openness is more than double the standard deviation of the trade openness indicator. Second, the indices related to foreign debt as a percentage of GDP and socioeconomic conditions seem to be the only indicators that have a positive, significant effect on financial development at least at the 5% level.

When looking at the magnitude of the effect of financial openness on financial development, and taking into consideration the interactive term with trade openness, we find that an increase in the financial openness index of 0.10 point leads to an increase in financial development of 5.64% (using coefficients in Table 4, column 1), which is of significant magnitude. In relation to the other variables that were statistically significant at the 5% level, we find that an increase in the foreign debt index (decrease in foreign debt) or an increase in the socioeconomic conditions index (decrease in socioeconomic pressures) of 1 point is associated with an increase on financial development of 17 and 15%, respectively, which is of relevant magnitude.

6. Conclusion

In the analysis of the impact of financial development on economic growth, there is one main finding: the impact of financial development on economic growth varies across the Latin American region. This analysis shows that financial development has a positive significant effect in the long run only for the high-income group. For the low-income group, empirical evidence shows that the impact of financial development on economic growth is negative in the long run.

The results obtained when the sample is separated by income groups corroborate previous findings that the effect of financial development is dependent on certain conditions. This must be taken into consideration when designing policies to promote economic growth by developing the financial sector in Latin America. Promoting the deepening of financial markets seems to be beneficial for high-income countries, but not for low-income countries. Therefore, financial reform should be a priority for those countries with relatively high income levels in Latin America, but not for all. For further research, disentangling those conditions that allow the relatively high-income group to reap the benefits of financial development in the long run is necessary. Perhaps preconditions related to institutions or a certain financial development threshold might be relevant.

In relation to the determinants of financial development in Latin America, financial openness plays a key role in the development of financial markets, where it has a robust, positive, significant effect of great magnitude. The analysis here provides evidence that financial openness is the most beneficial in terms of improving financial markets in those countries that are relatively closed. Thus, countries with trade restrictions will find that liberalizing capital accounts can lead to significant expansions of credit.

This analysis shows that country risk and some of the components of this index are important sources of financing in the region. Specifically, the component of the financial risk index related to foreign debt has a significant, positive effect on financial development. This finding tells us that a country's ability to pay its way is an important source of financing. Thus, the stability that a country achieves by being solvent seems to have important implications for financial markets, and this is a novel finding. We also find that another component of the political risk index related to stability is relevant for financial development. A higher socioeconomic conditions index means that as socioeconomic pressures related to unemployment, poverty and consumer confidence decrease, private credit is likely to increase. This finding also indicates that financial markets value the stability of government and society. From this analysis we can conclude that stability plays a key role in the development of financial markets in the Latin American region.

For further research, it will be interesting to evaluate whether there is a relationship between financial openness and institutions. Furthermore, this analysis uses an indicator of financial openness that relates to capital account openness. Future research should consider a broader indicator of financial liberalization that accounts not only for openness of the capital account but also for financial reforms and deregulation of the domestic financial market.

REFERENCES

Abiad, A. and A. Mody (2005). "Financial reform: What shakes it? What shapes it?" *American Economic Review* **95**(1): 66–88.

Abiad, A., E. Detragiache, and T. Tressel (2010), "A new database of financial reforms," IMF Staff Papers **57**(2): 281–302.

Acemoglu, D. and S. Johnson (2005), "Unbundling institutions," *Journal of Political Economy* **113**(5): 949–95.

Aghion, P., P. Howitt, and D. Mayer-Foulkes (2005), "The effect of financial development on convergence: Theory and evidence," *Quarterly Journal of Economics* **120**:173–222.

Andrianova, S., P. Demetriades, and A. Shortland (2008), "Government ownership of banks, institutions, and financial development," *Journal of Development Economics* **85**(1-2): 218–52.

Arellano, M. and S. Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies* **58**: 277–97.

Asteriou, D. and S. Hall (2007). *Applied econometrics: A modern approach using EViews and Microfit*. Basingstoke, England: Palgrave Macmillan.

Baltagi, B., P. Demetriades, and S. Hook Law (2009), "Financial development and openness: Evidence from panel data," *Journal of Development Economics* **89**(2): 285–96.

Bangake, C. and J. Eggoh (2011), "Further evidence on finance-growth causality: A panel data analysis," *Economic Systems* **35**(2): 176–88.

Beck, T., R. Levine, and N. Loayza (2000), "Finance and the sources of growth," *Journal of Financial Economics* **58**(1-2): 261–300.

Beck, T., A. Demirguc-Kunt, and R. Levine (2000), "A new database on financial development and structure," *World Bank Economic Review* **14**(3): 597–605. Retrieved in January 2013, revised version of September 2012, from http://go.worldbank.org/ X23UD9QUX0

Beck, T., A. Demirguc-Kunt, and R. Levine (2003), "Law, endowments, and finance," *Journal of Financial Economics* **70**(2): 137–81.

Beck, T. and R. Levine (2005). "Legal institutions and financial development," in C. Menard and M. Shirley, eds. *Handbook of New Institutional Economics*, Dordrecht and New York: Kluwer Academic Publishers.

Blackburne, E. and M. Frank (2007), "Estimation of nonstationary heterogeneous panels," *Stata Journal* **7**(2): 197–208.

Enders, W. (2004). *Applied econometric time series, 2nd edition*. New York: John Wiley and Sons.

Blanco, L. (2009), "The finance-growth link in Latin America," *Southern Economic Journal* **76**(1): 224–48.

Chinn, M. and H. Ito (2006), "What matters for financial development? Capital controls, institutions, and interactions," *Journal of Development Economics* **81**(1):163–92.

Chinn, M. and H. Ito (2008), "A new measure of financial openness," *Journal of Comparative Policy Analysis: Research and Practice* **10**(3): 309–22. Retrieved in January 2013, revised version of March 2012, from http://web.pdx.edu/~ito/Chinn-Ito_website.htm

De Gregorio, J. and P. Guidotti (1995), "Financial development and economic growth," *World Development* **23**: 433–48.

Djankov, S., C. McLiesh, and A. Shleifer (2007), "Private credit in 129 countries," *Journal of Financial Economics* **84**: 299–329.

Grier, K. and G. Tullock (1989), "An empirical analysis of cross-national economic growth, 1951–1980," *Journal of Monetary Economics* **24**(2): 259–76.

Heston, A., R. Summers, and B. Aten (2012). Penn World Table Version 6.3, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania. Retrieved in January 2013, revised version of November 2012, from http://pwt.econ.upenn.edu/

Herger, N., R. Hodler, and M. Lobsiger (2008), "What determines financial development? Culture, institutions or trade," *Review of World Economics* **144**(3): 558–87.

International Monetary Fund (IMF) (2013). International financial statistics. Retrieved in January 2013 from http://www.imfstatistics.org/imf/.

Klein, M. and G. Olivei (2008), "Capital account liberalization, financial depth, and economic growth," *Journal of International Money and Finance* **27**(6): 861–75.

Laeven, L. and F. Valencia (2012), "Systemic banking crises database: An update," *IMF Working Papers*: 12/163.

Levine, R. (2005), "Finance and growth: Theory and evidence," in P. Aghion and S. Durlauf, eds., *Handbook of Economic Growth,* 1A. North-Holland: Elsevier, 866–934.

Levine, R. (1997), "Financial development and economic growth: Views and agenda," *Journal of Economic Literature* **35**(2): 688–726.

Loayza, N. and R. Ranciere (2006), "Financial development, financial fragility, and growth," *Journal of Money, Credit, and Banking* **38**: 1051–76.

Mayer, T. and S. Zignago (2006). "CEPII distance data." Centre D'Etudes Prospectives et D'Informations Internationales, available at http://www.cepii.fr/anglaisgraph/bdd/distances.htm

Mishkin, F. (2009), "Globalization and financial development," *Journal of Development Economics* **89**(2): 164–69.

Odhiambo, N. (2007), "Supply-leading versus demand-following hypothesis: Empirical evidence from three SSA countries," *African Development Review* **19**: 257–80.

Pesaran, H., Y. Shin, and R. Smith (1999), "Pooled mean group estimation of dynamic heterogeneous panels," *Journal of the American Statistical Association* **94**(446): 621–34.

Political Risk Services Group (PRS) (2013). "International Country Risk Guide." Retrieved in January 2013 (purchased) from http://www.prsgroup.com/.

Rajan, R. and L. Zingales (2003), "The great reversals: The politics of financial development in the twentieth century," *Journal of Financial Economics* **69**: 5–50.

Rancière, R., A. Tornell, and F. Westermann (2008), "Financial liberalization," in S. Durlauf and L. Blume, eds., *The New Palgrave Dictionary of Economics.* Basingstoke, England: Palgrave Macmillan, Print.

Rioja, F. and N. Valev (2004a), "Finance and the sources of growth at various stages of economic development," *Economic Inquiry* **42**: 127–40.

Rioja, F. and N. Valev (2004b), "Does one size fit all? A reexamination of the finance and growth relationship," *Journal of Development Economics* **74**: 429–47.

Roe, M. and J. Siegel (2009), "Political instability: Its effects on financial development, its roots in the severity of economic inequality." Harvard University Working Paper.

Saint-Paul, G. (1992), "Technological choice, financial markets and economic development," *European Economic Review* **36**: 763–81.

Shen, C-H. and C-C. Lee (2006), "Same financial development yet different economic growth—why?" *Journal of Money, Credit and Banking* **38**: 1907–44.

Shan, J. (2005), "Does financial development 'lead' economic growth? A vector auto-regression appraisal," *Applied Economics* **37**: 1353–67.

United Nations Commodity Trade Statistics Database (UN COMTRADE) 2013. Retrieved in January 2013 from http://comtrade.un.org/.

Wooldridge, J.M. (2002), *Econometric analysis of cross section and panel data.* Cambridge: MIT Press.

APPENDIX

Sample and data description

The data used in this analysis are divided into two parts. For the first part, which focuses on determining the impact of financial development on economic growth in the short and long run, yearly observations between 1961 and 2010 are used. For the second part, which focuses on studying the determinants of financial development in Latin America, five-year average observations between 1985 and 2010 are used.

The 16 Latin American countries included in both parts of the analysis are Argentina, Bolivia, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Panama, Paraguay, Peru, Uruguay, and Venezuela. Countries were selected in the basis of data availability over a long period of time. The paper contains some discussion of estimations that include Haiti, Jamaica, and Trinidad and Tobago. These countries were not considered as part of the main sample because they do not share the characteristics common to the other countries and are not usually considered Latin American countries in regional analyses. The sample selection is based on the data available between 1960 and 2010. While there is some data for other countries not included in the sample such as Brazil and Nicaragua, the series are not available for the period of interest in this analysis.

This analysis uses the indicator of financial development most commonly used in previous work: private credit as a share of GDP. This indicator comes mainly from Beck, Demirguc-Kunt, and Levine's (2000) data on financial structure updated in September 2012. This analysis emphasizes financial development in relation to the banking sector. While studying the impact of equity markets on growth and its determinants for the Latin American region is relevant, consistent data across the region for a lengthy period of time is not available. Furthermore, financial markets in Latin America are more heavily based on the banking sector, which makes the focus on private credit as an indicator of financial development a suitable approach.

Data on real GDP per capita, population, government spending as a share of GDP, and trade openness are obtained from the Penn World Tables (Heston *et al.*, 2012). Real GDP per capita is estimated by extrapolating 1996 values in international dollars, making this indicator comparable across countries. Data on financial openness are obtained from Chinn and Ito's (2008) database, updated in March 2013, and data on inflation are obtained from the International Financial Statistics (IMF, 2013). Banking crisis data are obtained from Laeven and Valencia (2012). Country risk data are obtained from Political Risk Services Group (2013). Other data used to construct a measure of trade openness that is exogenous in the growth equation come from the United Nations Commodity Trade Statistics Database (UNCOMTRADE, 2013) and Mayer and Zignago (2006).

Table A1. Variable description and source

Financial development	Private credit as a share of GDP. Source: Beck, Demirguc-Kunt, and Levine (2000; version published in 2012).
GDP per capita	Real GDP per capita, Laspeyres constant prices. Source: Heston *et al.* (2012).
Initial GDP per capita	Initial GDP in the five-year period divided by population in the current year (time variant, different every year). Constructed using total population and real GDP, Laspeyres constant prices. Source: Heston *et al.* (2012).
Government size	Government spending as a share of GDP (from real GDP, Laspeyres constant prices). Source: Heston *et al.* (2012).
Inflation	Inflation plus 100 (in natural log). Source: Author's construction using International Financial Statistics data (IMF, 2013).
Trade	Residual of a regression of the natural log of trade openness (exports plus imports divided by GDP, from real GDP, Laspeyres constant prices) on the natural log of the area of the country, natural log of population, landlocked dummy, net oil exporter dummy, and time dummies. Source: Author's construction using data from Heston *et al.* (2012) for the trade openness indicator and population, from Mayer and Zignago (2006) for country area and landlocked dummy, and from UNCOMTRADE (2013) for construction of the net exporter oil dummy (this estimation assumes oil dummy equals zero for missing observations).
Banking crisis	Banking crisis dummy equal to 1 if a country experienced a financial crisis in that year. Source: Laeven and Valencia (2012).
Trade openness	Exports plus imports as a share of GDP (from real GDP Laspeyres constant prices). Source: Heston *et al.* (2012).
Financial openness	Index of capital account openness. Source: Chinn and Ito (2008, version published in 2012).
Country risk	Composite index of country risk. Index composed of financial, economic and political risk indices. The political risk rating contributes 50% of the composite rating, while the financial and economic risk ratings each contribute 25%.

Table A1. (continued)[a]

Political risk	Contains the following 12 components: government stability, socioeconomic conditions, investment profile, internal conflict, external conflict, corruption, military in politics, religious tensions, law and order, ethnic tensions, democratic accountability, and bureaucracy quality.
Financial risk	Composed of the following 5 components: foreign debt as a percentage of GDP, foreign debt service as a percentage of exports of goods and services, current account as a percentage of exports of goods and service, net international liquidity as months of import cover, and exchange rate stability.
Economic risk	Composed of the following 5 components: GDP per capita, real GDP growth, annual inflation rate, budget balance as a percentage of GDP, current account as a percentage of GDP.
Government stability	This indicator relates to the government's ability to carry out its declared programs and its ability to stay in office. This indicator is composed of government unity, legislative strength and popular support.
Investment profile	This indicator is related to risks to investment, and is composed of contract viability/expropriation, profits repatriation, and payment delays.
Internal conflict	Indicator related to internal political violence and its actual or potential impact on governance. It is composed of civil war/coup threat, terrorism/political violence, and civil disorder.
Socioeconomic conditions	Constructed using data on unemployment, consumer confidence and poverty to measure socioeconomic pressures at work and in society that can lead to social dissatisfaction.
Foreign debt	Index based on foreign debt as a percentage of GDP.

*The source for variables in this section is Political Risk Services (2013).

Vol. 50 No. 2 (NOV, 2013), 209–231

LATIN AMERICAN JOURNAL OF ECONOMICS

FOREIGN INVESTMENT AND WAGES: A CROWDING-OUT EFFECT IN MEXICO

Enrique L. Kato-Vidal*

The purpose of this article is to determine the impact of foreign direct investment (FDI) on a country's overall economy rather than simply the sectors receiving such investment. The strategy consisted of adopting a crowding-in/crowding-out approach to Mexico's total capital volume in the 1993-2010 period. The substitutability of foreign and local capital implies a lower-than-expected economic dynamism. Using a dynamic panel analysis, a negative relationship was found between FDI and the general wage. Throughout the analysis, firm size stands out as a key variable in explaining the impact of FDI.

JEL classification: F21, O11, C23

Keywords: FDI, wage, firm size, substitutability of capital

1. Introduction

The economic liberalization of Mexico in recent decades has induced a substantial increase in international trade as a share of GDP and has generated a significant increase in the influx of foreign capital, especially in the years following enactment of the North American Free Trade Agreement (NAFTA). A significant economic consequence of trade reform is the rise in wage inequality driven by the higher wages received by skilled labor as a result of increased foreign capital (Hanson, 2003).

The expectation on the part of policy makers is that foreign direct investment (FDI) in Latin America would have effects similar to those seen in Asia, where technological breakthroughs stimulated investment and triggered economic growth in the region, i.e., a complementarity of investments (Petri, 2012). However, the results for Mexico under the NAFTA framework do not reveal an acceleration of capital accumulation, productivity or increasing wages. In this sense, the objective of this article is to determine the contribution of FDI to overall economic activity. Therefore, a crowding-in/crowding-out perspective on total capital is adopted, and a wage equation is estimated using a dynamic panel.

* Universidad Autónoma de Querétaro and the Sistema Nacional de Investigadores, Mexico. Address: Graduate Building, School of Accounting and Administration, Universidad Autónoma de Querétaro, Cerro de Las Campanas s/n, Col. Las Campanas, C. P. 76010, Tel. (52+442) 192 1200 ext. 5273. Email: enrileo@gmail.com.

doi 10.7764/LAJE.50.2.209

This article provides evidence of the substitutability of foreign and local capital and the negative impact of FDI on overall wages in the Mexican economy. This evidence could be used to reformulate policies to strengthen the domestic market, thereby reducing the impact of the Great Recession, as Ben Bernanke has referred to the period following the 2008 crisis. Furthermore, our findings indicate that to achieve the benefits of a policy to attract foreign investment, a lower international productivity gap is required.

The paper is structured as follows: In the first section, the crowding-in (CI) and crowding-out (CO) effects are discussed in terms of the complementarity or substitutability of foreign direct investment (FDI) with local capital. In the second section, FDI is associated with per-capita income as a function of production. In the third section a static panel model is estimated to determine the effect of FDI on total capital. The fourth section contains an estimation of the impact of FDI on wages using a dynamic panel. In Section 5, the static and dynamic estimates are reconciled and discussed. Finally, the concluding remarks affirm the importance of efficiency and average firm size in explaining the crowding-out effect of capital.

2. Investment: crowding-in and crowding-out effects

In the literature, FDI is understood to confer advantages on the region that receives it. Generally, these advantages consist of information regarding external markets, technology transfer, improvement in administrative skills, and job creation, among other factors. Lipsey (2004) presents a summary. There are also numerous articles that seek to verify the existence and magnitude of these benefits. An interesting subject for Latin America is the role that FDI plays in the promotion of local investment while reducing corruption (Larrain *et al.*, 2004). The procedures used to validate the benefits of FDI vary and are based on the researcher's interests.

An approach that is useful to our purposes is to determine the relationship between foreign and local investment. There are three possibilities: a neutral effect, crowding in, and crowding out. The first occurs when one dollar of FDI increases total investment by exactly one dollar, the second effect reflects an increase in local investment whereby total investment increases by more than a dollar, and finally, the third effect occurs when total investment increases by less than one dollar for each dollar of FDI.

It is not possible to determine *a priori* which of the three effects will prevail in an economy (Agosin and Mayer, 2000). However, it is possible to identify certain determinants that could explain the final effect. In general, national investment policies and the strength of local businesses determine the impact that FDI will have on host regions. Specifically, FDI's positive impact tends to be greater when it occurs in new markets or is oriented towards foreign markets due to the provision of knowledge and technologies. However, when FDI flows into existing markets, the impact may be positive, though small, or fully negative if it shifts competition.

In Belgium, Backer and Sleuwaegen (2003) documented a negative impact for their study period. These authors argue, on the basis of their results, that long-term benefits may exist that help counteract short-term negative balances. However, competition in existing markets is not limited to customers; it also occurs in factor markets when there is competition to recruit the best workers and, occasionally, to attract investment. For certain industries, competition may center on permits for the use of natural resources, e.g., water, forests, or mineral resources. In each of these areas, multinationals compete with local businesses. When both local and multinational companies participate in complementary activities, CI effects could be generated by production linkages. In cases where competition is direct, CO effects are expected.

One study that provides evidence on the state of inter-industrial relations in Mexico describes minor production linkages between the leading and non-leading sectors of the country (Ortiz, 2007). This finding is based on the total linkage coefficients stemming from the input-output matrices available from 1950 to 1995. A more recent study that compares the major economies of Latin America identifies Colombia and Mexico as countries that produce less than one indirect job for each new direct job associated with exports (ECLAC, 2012: 142).

In addition to the lower level of integration in the production chains, two trends are evident: a) leading sectors maintain a behavioral dynamic in which they find it profitable to grow based on imports and losses in terms of trade, and b) sectors oriented towards the domestic market exhibit behavior more in accordance with standard factors of domestic integration and the pace of domestic accumulation (Ortiz, 2007). Hence, we analyze the effect that FDI exerts on total national investment.

To perform a specific estimation for Mexico, a panel analysis is conducted with information from all 32 federal divisions (states and district). We use the results of Agosin and Mayer (2000) as a reference. We begin from the simplification that the total amount of investment in an economy is the sum of domestic investment I_d and foreign investment I_f, that is, we ignore that FDI is not always greenfield investment. The time dimension was also incorporated into the investment identity, yielding $I_t = I_{d,t} + I_{f,t}$.

We assume foreign investment to be an exogenous variable. Unlike Agosin and Mayer (2000), our estimate is performed using census data in 5-year periods, which directly produces long-term coefficients. By contrast, Agosin and Mayer performed their estimate using an equation derived from a partial adjustment model with adaptive expectations to produce consistent estimates with their annual data. In particular, the long-term ratio $\hat{\beta}_{LT}$ determines the presence of CI or CO effects.

Agosin and Mayer found different results for the three regions studied over the 1970-1996 period. In Africa, there was evidence of a neutral effect, i.e., foreign investment increased total investment by a one-to-one ratio. As shown in Table 1, Asia registered a CI effect. In Latin America, the evidence confirmed a CO effect. However, subdividing the study period in Latin America altered the sign and magnitude of the coefficient $\hat{\beta}_{LT}$; values close to zero or negative were obtained. In the breakdown by country, the authors classified Mexico, as well as Brazil and Argentina, as experiencing a neutral effect and Chile and Guatemala as experiencing a CO effect.

The evidence presented in Table 1 demonstrates that when an economy receives foreign investment, total investment may increase by far less

Table 1. Positive effects of investment in Asia and offset effects in Latin America

Latin America		Asia	
1970-1996 -0.14		1970-1996 2.71	
1976-1985 -1.22	1986-1996 0.04	1976-1985 5.56	1986-1996 2.91

Source: Estimates by Agosin and Mayer (2000).
Note: There is evidence that FDI promotes local investment when $\hat{\beta}_{LT} > 1$, i.e., evidence of CI; CO evidence is obtained when $\hat{\beta}_{LT} < 1$, and finally, we say that a neutral effect persists when $\hat{\beta}_{LT} = 1$.

than the FDI or may not increase at all. In Asia there were high rates of investment and CI effect that were also accompanied by policies that selected foreign investment projects and provided support to local businesses. Such selective policies sought to ensure that FDI did not displace local companies and that multinational companies (MNCs) would contribute new technologies or new products. In contrast, in Latin America inward FDI is the policy itself, not the promotion of local development. This may result in CO effect. For example, in the case of Mexico, there is some evidence of vertical technology diffusion from FDI but no horizontal technology diffusion (López-Córdova, 2002; cited in Ito, 2010: 18).

3. Foreign direct investment and the expectation of growth

Several publications rank Mexico as one of the primary recipients of FDI. Mexico ranked sixth in the list of top host economies for FDI in 2010-2012 according to the number of times the country is mentioned as the top FDI priority by respondent transnational corporations (UNCTAD, 2010). Some articles also argue that the benefits of attracting FDI never materialized. For example, in their assessment of Mexico, Waldkirch *et al.* (2009) express disappointment at the poor results obtained from the country's economic development strategy based on attracting FDI. Meanwhile, Ito (2010: 16) concludes that FDI inflows rose rapidly in Mexico after NAFTA was signed in 1994, but states that there is no evidence of NAFTA having contributed to the convergence of productivity toward a narrower gap.

Theory predicts that GDP can grow only if there is growth in productive factors, including the level of technology. The entry of FDI into a country contributes positively to the production process via two known factors: physical capital $K(t)$ and technology $T(t)$. The standard approximation is based on a production function that explains obtaining a product flow Y at time t using the factors of capital, technology and labor $L(t)$. Thus, the conventional production function states that $Y(t) = F[K(t), L(t), T(t)]$.

We assume that the attraction of greater amounts of FDI has two objectives. The first is to ensure a greater supply of capital to achieve higher worker productivity levels while the second is to reach a new steady state with higher capital and production. Using the intensive form of the production function $y = f(k)$, in which the lowercase letters represent per capita variables, allows us to demonstrate[1] that the properties of the growth of capital $\dot{k}/k$ are immediately transferred to production growth:

$$\frac{\dot{y}}{y} = \left[k \cdot f'(k) / f(k)\right] \cdot \left(\dot{k} / k\right) \tag{1}$$

This result indicates that additional capital induces an increase in per capita income and is accomplished in cases more general than the Cobb-Douglas functional form, except that the share of capital income grows rapidly enough to more than offset the decline of $\dot{k}/k$, as the economy develops (Barro *et al.*, 2003).

Generally, in the context of economic liberalization, economic growth is seen as the result of the rate of capital accumulation (Calderon *et al.*, 2006:61). However, it is insufficient to explain the benefits of FDI in an economy merely with respect to production because national capital and foreign capital can be either complementary or substitute inputs. In Mexico, inward FDI did not prevent the total factor productivity international gap from growing in the years following the signing of NAFTA (Ito, 2010: 28). In American multinational firms, local and foreign investment decisions were complementary (intra-organizational complementarity). Nationwide, the foreign investment conducted by U.S. multinational companies reported a significant estimated coefficient of -1.855 (Desai *et al.*, 2005). This result means that an additional dollar of investment by foreign-owned firms in the United States reduces domestic investment by U.S. multinational firms by 1.9 dollars (substitution effect in U.S.).

A panel data analysis for 35 developing countries from 1970 to 2003 reveals that foreign capital has a negative effect on local capital, although not a significant one, which suggests that the process of capital accumulation by MNCs does not significantly displace local investment opportunities (Ahmad *et al.*, 2009: 30). In the following section, we analyze whether Mexico is experiencing substitutability or complementarity.

1. Let $\partial Y/\partial K = f'(k)$ denote the marginal productivity of capital. The expression in parenthesis is, under various assumptions, the share of capital, i.e., capital income's share of total income. The equation demonstrates that the relationship between $\dot{y}/y$ and $\dot{k}/k$ depends on the behavior and share of capital. In the Cobb-Douglas case, the capital share is constant and $\dot{y}/y$ mimics that of $\dot{k}/k$ (Barro *et al.*, 2003).

4. A panel estimate for Mexico

Unlike international studies, we did not perform the estimates using investment rates (e.g., the investment/GDP ratio) but rather did so directly using the capital stock variable reported by the economic census. We believe that using capital stock in our estimate will provide a more valid coefficient and will more convincingly describe the relationship between domestic and foreign capital.

To explain total capital (domestic and foreign) per worker (k), we use foreign capital per worker (k^f) as an exogenous variable, two control variables, and a random error term e_{it}:

$$k_{it} = \alpha_i + \beta_i k_{it}^f + \theta_1 X_{1it} + \theta_2 X_{2it} + e_{it} \tag{2}$$

The first control variable X_1 is average firm size, as measured by number of employees; the second control variable X_2 is the average productivity of labor. The inclusion of both variables is fully justified by standard models of rational economic agents who pursue profit maximization. The relevance of X_1 in the equation is that it allows for identification of the demand for capital k required to realize productive investment projects.

The coefficient θ_1 must be positive whenever it reports the variation in the demand for capital according to a change in the average size of an organization. In the following sections on dynamic specification, the size of the firm will have a high explanatory power. In general, we assume that the presence of small and medium enterprises increases as capital intensity per worker decreases. A measure of human capital could also be used to explain the increased participation of small businesses. When there is a higher ratio of white collar to blue collar workers, a greater proportion of small and medium enterprises are anticipated in the production structure (Álvarez *et al.*, 2001).

Additionally, we include the productivity variable X_2 to explain k given the causality between profitability and investment decisions. Among other requirements, a project must generate revenues in excess of costs to be approved. The productivity variable we employ is related to two aspects. On the one hand, productivity is positively related to revenue due to production increases and, on the other hand, it is negatively related to costs due to efficiency gains. Therefore, we expect the variable X_2 to provide information on new capital investment induced by productivity gains. We expect θ_2 to be positive. It could approach zero if marginal profits do not induce new investments.

We use a balanced panel constructed of 128 observations from the last four census years—1993, 1998, 2003, and 2008—from the 32 federal states of Mexico. Means and standard deviations indicate high levels of dispersion, except in the case of X_1 (Table A1 in the appendix). An interesting property of the sample is obtained by analyzing the correlation matrix in which the dependent variable k itself is correlated with the exogenous and control variables. An early positive result is that the latter variables are not correlated with each other (Table A2 in the appendix). This situation is desirable to achieve consistent, unbiased, normal, and efficient estimators.

Before proceeding to estimation of the panel, we performed a partial correlation analysis to examine the interaction of data, i.e., we recalculated the correlation between k and the other variables using the control variables. The control variable X_2 has a higher effect on the correlation between k and k_f. The variable X_1 does not substantially affect the results of the simple correlation. By contrast, X_2 increases the significance and the correlation coefficient between k and k_f (Table A3 in the appendix). The evidence indicates that the variable X_1 is redundant and would not further explain k. The redundancy hypothesis is contrasted in the estimation of the panel.

The estimation of the panel was performed, and different specifications were explored. A first test in the panel specification was whether fixed effects should be incorporated into the estimation. The high significance level of the test allows for rejection of the hypothesis that fixed effects are redundant (Table 2). In addition, three estimations were performed: with one of the control variables, with the two variables, and with none. The set of estimates provides different values for the parameter of interest β for the variable k^f and demonstrates that the CO effect, or potentially a neutral effect, prevails in the relationship between total and foreign capital.

To select the final specification of the panel, the redundant variable and omitted variable hypothesis tests were used to decide which control variables were to be included. The high levels of significance obtained for both tests indicated that both control variables should be incorporated in the panel estimation. The final specification is provided in Equation [2] in Table 2. Comparatively, the estimation with fixed effects provides a lower value of β than that without (0.641 vs. 0.472), although both cases yield evidence of a CO effect.

Table 2. Crowding-out effect of FDI in Mexico

	Without fixed effects	With fixed effects			
	[1]	[2]	[3]	[4]	[5]
Constant	-255.6 (-0.1)	-21321 (-7.8)	16,997 (7.0)	-29,891 (-3.9)	4,675 (1.9)
k_f	0.641 (6.1)	0.472 (6.7)	1.004 (5.6)	0.594 (9.1)	0.827 (4.6)
X_1	1112 (2.7)	5607 (10.1)		9,752 (5.4)	
X_2	0.684 (21.1)	0.679 (21.6)			0.71 (19.9)
$\bar{X}^2$	0.799	0.871	0.31	0.57	0.81
F (prob)	169.1 (0.00)	26.2 (0.00)	2.8 (0.00)	6.0 (0.00)	17.4 (0.00)
Redundant fixed effects test		3.364 (0.00)			
H_0: redundant variable			$p = 0.0000$	$p = 0.0000$	$p = 0.0000$
H_0: omitted variable			$p = 0.0000$	$p = 0.0003$	$p = 0.0000$

Note: t-values in parentheses. Cross-sectional weights were used in the estimation. Observations = 128; entities = 32. Source: Based on data from the National Institute of Statistics and Geography (INEGI) and the National Registry of Foreign Investment (RNIE), Mexico.

After demonstrating that the relationship between foreign and total capital produced a CO effect, the impact on wages was evaluated. In particular, a dynamic panel estimate was performed using quarterly data. This second exercise complements the structural results of this section and extends the information from 2008 to 2010.

5. Foreign capital and wages: a dynamic approach

There are several ways to assess the impacts of increased capital formation on wages. The approximation that we use in this study employs a profit function, as in Leamer *et al.* (2000). The assumption is that the firm's goal is to maximize the present value of profits. We assume that there are no adjustment costs or intertemporal elements in the firm's maximization problem regarding the acquisition of capital or labor services, and the firm maximizes profits at each point in time (Barro and Sala-i-Martin, 2003). The representative firm's flow of net income or profits at a given moment in time is given by:

$$\pi = F\left(K, L, T\right) - \left(r + \delta\right)K - wL \tag{3}$$

that is, gross income from product sales $F(K,L,T)$ minus the cost of the factors, comprising capital gains $(r + \delta)K$ and workers' wages wL. Solving for w, we have:

$$w = f(k) - (r + \delta)k - \pi/L \tag{4}$$

where $k \equiv K/L$ is worker capital and the function $f(k)$ equals $(K/L,1,T)$. As in Barro and Sala-i-Martin, we assume that T is constant as an implicit parameter in the definition of $f(k)$. This expression indicates that there are three important determinants of wages: production level $(y = f(k))$, capital intensity per worker, and expected profits. However, to achieve a complete description of the relationship between FDI and wages, the basic formulation should be complemented with the known results regarding the presence of multinationals.

The extensive review by Lipsey (2004: 345) categorically states that an MNC always pays more than a local private firm, which can be explained because wage levels are almost always positively related to firm size and the MNC has a larger scale of production. Thus, theory predicts that the increased presence of foreign companies should increase the demand for labor, thereby placing upward pressure on average wages. However, the debate continues. Lipsey himself (2004) has argued that foreign companies can pay higher wages without affecting local businesses. In turn, Aitken *et al.* (1996) argue that under certain conditions, the impact of FDI on wages could be zero, e.g., because the labor pool does not change, although they recognize that there could be cases in which the increase in labor demand is indeed converted into wage increases.

Empirically, there are few strategies to reliably compare the effect of FDI on wages. The alternative that we selected consists of observing changes in productivity. The accumulated evidence is that the only characteristic of the firm that seems to be important is the productivity gap, i.e., the higher the productivity gap, the lower the wage spillover (Lipsey 2004: 352). Therefore, to capture changes in productivity it is appropriate to examine not only existing businesses at a point in time but also the companies that enter and exit the market (Lipsey 2004: 354).

In particular, FDI has a crowding-out effect on the creation of new companies, which means that FDI reduces the entry of local enterprises and increases exit of these new enterprises from the market (Backer and Sleuwaegen, 2003). The crowding-out effect of FDI arises in the form of direct competition in the market for goods, but competition also occurs in the labor market. The assumption in the analysis of Backer and Sleuwaegen (2003: 71-2) is that the wage paid by multinationals (W_{MNE}) is greater than the wage paid by local businesses (w), and also may be greater than some entrepreneurs' income (I_{own}); therefore, these individuals would abandon their personal projects to obtain more income as employees of a multinational, compared to the income that they could earn as entrepreneurs: $I_{own} \equiv F(\cdot) - rK - wl < W_{MNE}$. Thus, in the short term, the presence of FDI generates negative effects on local enterprises; however, it remains to be determined whether the balance continues to be negative in the long run.

To incorporate the wage-firm size link into the analysis and capture aggregate changes, we assume that companies pay fair wages $\hat{w}$ because worker effort decreases below this level. Following Egger *et al.* (2009), the reference wage is determined by taking the geometric mean of two components. The first reflects the productivity ρ of the firm where the workers are employed. The second component is associated with the average wage income ($\bar{w}$) and the employment rate $(1 - U)$, where U is the unemployment rate. Thus, the wage in the economy can be expressed as:

$$\hat{w}\left(\rho\right) = w = \rho^{\lambda} \cdot \left[\left(1 - U\right)\bar{w}\right]^{1-\lambda} \tag{5}$$

where $\lambda \in [0,1]$ can be interpreted as a rent-sharing or justice parameter. In the case that $\lambda = 0$, all companies pay the same wage, and if $\lambda > 0$ wages would depend only on the companies. An attribute of the reference wage equation is that it incorporates both the particular conditions of the firm and the impact of unemployment on the market wage. For estimation purposes it is convenient to specify this wage equation using the frequency and availability with which the unemployment indicators are published.

5.1. Dynamic panel estimation

Section 3 demonstrated the relationship between FDI and the capital stock. In this study, we evaluate the impact of FDI on the average general wage. To perform the statistical comparison, we use a dynamic panel analysis with fixed effects. The fixed effects model is more appropriate than the random effects model for two reasons. First, the panel data estimators allow for consistent estimation of the effects of the observed explanatory variables, although our dependent variable depends on an unobserved variable correlated with the observed explanatory variables. Second, it is also likely that the typical macroeconomic panel is integrated by entities that are not random but are rather selected by the researcher (Judson *et al.*, 1999).

Dynamic panel models include the lagged dependent variable and unobserved individual effects in their specifications. These models are powerful tools that allow us to do the empirical modeling of dynamics and account for the heterogeneity of each cross-section. Dynamic panel models explicitly include variables for analyzing past behavior and invariant individual specific effects (time-invariant), thereby permitting us to better understand what factors promote behavior over time and differentiate between true dynamics and factors that vary between cross-sections and those factors that do not vary within these cross-sections, although such factors are not observable.

The basic equation to be estimated is:

$$w_{i,t} = \alpha_i + \gamma w_{i,t-k} + \beta k^f_{i,t} + \varphi_1 y_i + \varphi_2 E_i + \varphi_3 U_i + \in_{it} \tag{6}$$

where w is the logarithm of the wage, α_i is a fixed effect, k^f is the foreign capital per worker, y is the logarithm of the economic activity index, which is a proxy for the quarterly gross domestic product at the subnational level (GDP data not available), E is the logarithm of the number of microbusinesses with establishments (a permanent place of business), U is the unemployment rate $U6$, which is the official unemployment rate $U3$ plus total employed part-time for economic reasons, which permits a more accurate valuation of unemployment in Mexico, and $e_{it} \sim N(0,\sigma^2_\in)$ is a random error term; $i = 1,\ldots,N$; $t = 1,\ldots,T$, where N is the number of federal entities in the panel and T is the number of observations over time. We assume that:

$$\sigma^2_\in > 0$$
$$E(\in_{it},\in_{jt}) = 0 \quad i \neq j \ or \ t \neq s$$
$$E(\alpha_i,\in_{jt}) = 0 \quad \forall\ i,j,s$$
$$E(k^f_{it},\in_{js}) = 0 \quad \forall\ i,j,t,s$$

We used a balanced panel constructed with 992 observations from the 32 federal states of Mexico for the period 2003Q2-2010Q4. The means and standard deviations are reported in Table A4 in the appendix.

5.2. Panel unit root tests and cointegration tests

To perform the panel data analysis, we use the Levin, Lin, and Chu (LLC) and Im, Pesaran, and Shin (IPS) tests for the stationarity of the series and to verify that all series are of the same order of integration. It is worth noting that the IPS test solves the problem that the LLC test has regarding serial correlation of the treatment of heterogeneity between units in a dynamic panel. Once the order of integration has been defined, the Pedroni tests are applied to account for heterogeneity using specific parameters that vary between individual units in the sample, which is more realistic than assuming that the cointegrating vectors are identical across panel units. The seven Pedroni tests propose non-cointegration as a null hypothesis. The first four tests focus on the within dimension, and the three remaining tests are based on the between dimension. To reject the null hypothesis of absence of cointegration, the calculated statistics must be less than the tabulated critical values.

The results of the LLC and IPS panel unit root tests show that the null hypothesis of unit roots in level for the panel data on wages [$log(w)$] and cumulative FDI inflows per worker (k^f) cannot be rejected. Nevertheless, the hypothesis is rejected when the series are in first differences (Table A5 in the appendix). These results indicate that the variables are non-stationary in level and stationary in first-differences. Therefore, we can implement a test for panel cointegration in $\log(w)$ and k^f. Regarding the complementary variables, $log(y)$, $log(E)$, and $U6$ were also found to be stationary in first-differences, although there is weak evidence that $log(y)$ and $U6$ can be non-stationary series under certain combinations of trend and intercept in the specification of the unit root test.

Both within-group and between-group tests were performed to verify cointegration in the panel data. In general, the results of the Pedroni cointegration tests allow for rejection of the null hypothesis of no cointegration at the 1% significance level under the assumption of no deterministic intercept or trend. The tests produce inconclusive results regarding cointegration when an intercept or trend is included (Table A6 in the appendix). Nevertheless, the evidence is sufficient to assert that there is cointegration between $log(w)$, k^f, $log(y)$, $log(E)$, and $U6$. The presence of a long-term relationship in the panel reveals the impact of FDI in determining wages in the country, even in the presence of the control variables.

The estimated coefficients in the dynamic panel reveal a negative effect of the ratio of foreign capital per worker (k^f) and the unemployment rate on wages. In contrast, product growth (y) and microbusinesses (E) positively affect wages (Table 3). Specifications [1] and [2] were used to obtain values that would allow for a comparison with the estimated coefficients produced by the full equation [3]. In the first specification, only k^f was included as a regressor, and the results indicate that it allows us to explain the variability of wages. In fact, an increase of $1,000 US dollars (3/8 of the average k^f) reduces wages by -0.50% in the short term and by -8.5% in the long term.

The results of the full specification [3] are believed to be more plausible than those obtained in [1]. A $1,000 increase in k^f decreases wages by 1.2% in the short term and by 3.7% in the long term. In line with standard models, the unemployment rate is highly significant. Unlike in Pissarides (2009), a broader measure of unemployment was used ($U6$ instead of $U3$) that best describes the reality in Mexico. Comparatively, the estimates from [3] are similar to the international estimates of the wages of changers or movers that Pissarides presents (2009: 1357-8).

The final specification [3] achieves its estimated adjustment between wages, foreign capital, and the unemployment rate by incorporating a key variable that Lipsey (2004: 354) terms the productivity gap and suggests can be operationalized by recording firm entry and exit. In our case, due to data availability, we used the number of microenterprises with establishments (E); this variable was highly significant. We believe that the number of microenterprises is fundamental for the analysis because it reflects changes in the production structure. This result is important because a greater number of microenterprises means a smaller relative average firm size and is expected to decrease the capital intensity per worker (Álvarez *et al.*, 2001).

Empirically, we find that an increase in microenterprises has a positive effect on wages, which is consistent with the negative effect of k^f on wages. Quantitatively, a variation in the number of microenterprises generates an impact of similar magnitude to that of product growth. This finding indicates the importance of supplementing the attraction of foreign investment with the emergence of a base of local companies such that the combined effect protects the purchasing power of wages. This evidence from the Mexican case contradicts the theoretical prediction advanced by Backer and Sleuwaegen (2003).

Table 3. Wage equation: Negative effect of k^f

Dependent variable $log(w)$	[1]		[2]		[3]	
	Short-term	Long-term	Short-term	Long-term	Short-term	Long-term
$log(w_{i,i-1})$	0.65 (22.3)		0.65 (22.5)		0.46 (9.8)	
$log(w_{i,i-2})$	0.29 (10.5)		0.29 (10.9)		0.21 (5.7)	
k^f	-4.97 x 10^{-6} (-3.3)	-8.48 x 10^{-5}	-1.75 x 10^{-7} (-0.1)	-3.02 x 10-6	-1.20 x 10-5 (-2.8)	-3.68 x 10-5
$U6$			-0.003 (-2.2)	-0.060	-0.01 (-3.7)	-0.027
$log(E)$					0.146 (3.7)	0.449
$log(y)$					0.152 (3.7)	0.466
J-statistic	242.4		245.8		134.4	
Instrument rank	80		80		62	
Sample	2001Q1-2010Q4		2001Q1-2010Q4		2003Q2-2010Q4	
Periods included	40		40		31	
Observations	1280		1280		992	

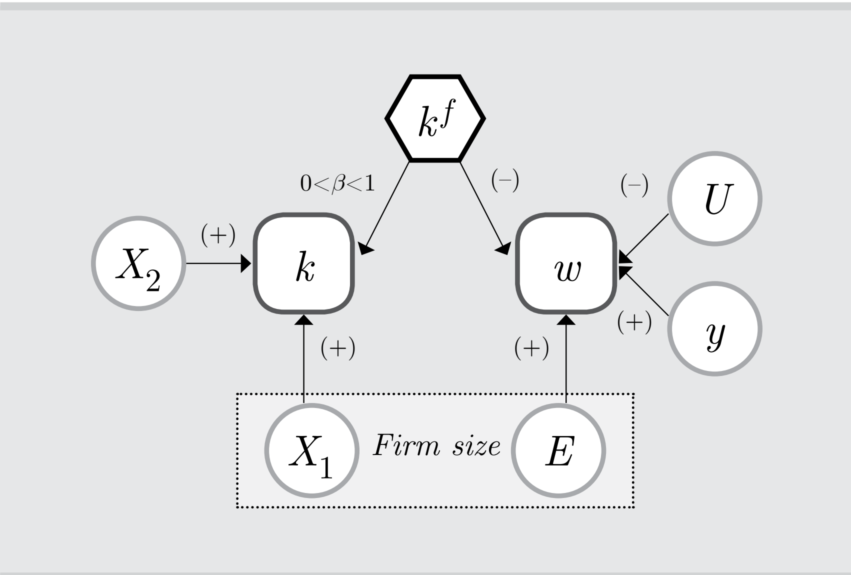
Source: Author's calculations.
Note: t-values in parentheses; entities = 32; method: panel generalized method of moments. Cross-section fixed (orthogonal deviations).

Instead, partial support is found in the evidence on Mexico reported in Dussel (2007) that the impact of FDI varied by state; in certain cases, there were increases in wages, and in others, the impact of FDI was negative. The dynamic panel methodology employed in this section and the specification that we used allows us to conclude convincingly that foreign capital has a negative impact on wages.

6. Reconciling static and dynamic estimates

This section integrates the estimations reported in sections 3 and 4 that were obtained in two panel estimates for Mexico, one static and another dynamic, into a joint explanation. The main reason for linking these estimates is the desire to understand the implications that the CO effect of FDI on capital has for wages in Mexico. In Figure 1, all variables employed in the analysis are listed with the signs of their estimated coefficients.

Figure 1. Two adverse effects of k^f: capital crowd out and wage inhibition



Note: X_1 is the average firm size as measured by the number of workers, X_2 is the average productivity of labor, and E is the logarithm of the number of microenterprises.

The figure presents the static estimation on the left and the dynamic equation on the right. The productivity indicators appear at the extremes. On the left side, the productivity variable is reported directly and has the expected positive sign, i.e., more productivity induces greater investment and capital accumulation. On the right, productivity appears implicitly through the relationship between production and unemployment. This effect is known as Okun's law, which, while theoretically contested, remains empirically useful (Blinder, 1997).

The relationship found between production and wages was highly inelastic. We believe that this relationship is due to the CO effect of FDI on capital. The theoretical prediction establishes that changes in capital are positively related to changes in income. However, the CO effect implies that capital accumulation is lower compared to what would be found with less substitutability between foreign and local capital. In this analysis, we omitted other explanations regarding the wage-setting policies that are important in the concentrated markets of Mexico.

At the base of the figure, the size of the firm is shown to be a key indicator of the interaction between the substitutability of foreign and domestic capital and their impacts on wages. The results of the panel allow us to assume that the attraction of foreign capital implies a reorganization of the production structure and thus the importance of incorporating firm size indicators in the estimates. The theoretical prediction states that the higher the productivity gap between local and foreign companies, the lower the positive spillover of foreign investment in the recipient economy. The evidence obtained seems to indicate that the gap is sufficiently wide that a meaningful amount of profits derived from the presence of foreign companies fails to be observed in the system.

Instead, we find that general wages have increased as a result of an increasing number of microenterprises. Firm size within Mexico tends to decrease because the representative size of microenterprises is smaller than the average size of firms in the economy. This finding may seem counterintuitive, given that a decrease in the average size of the firm should be associated with wage decreases, rather than wage increases, as is actually the case. A plausible explanation is that investment in microenterprises has a multiplier effect on spending in the economy and thus generates a high capital acceleration effect.

7. Concluding remarks

In this article, the effects of foreign direct investment were estimated for the entire Mexican economy beginning in 1993, the year that economic liberalization rebounded as a result of NAFTA. Using panel data on the 32 federal entities, the negative impacts of FDI on total capital and on general wages in the country were estimated. The negative effect of FDI could be due to a productivity gap wide enough to prevent FDI from having positive effects.

The panel estimates demonstrate that productivity or efficiency indicators are significant in explaining the interaction of FDI with the variables of interest. Additionally, firm size plays an important role in the calibration of the equations. In fact, we find coefficients of similar magnitude for the number of microenterprises and production volume, which reveals that the emergence of new enterprises does not have a minor or insignificant impact, as one might imagine, but rather is of a magnitude comparable to that of economic growth.

The evidence presented contrasts with the results of other studies that analyze the sectors receiving FDI and find an increase in wages, especially for skilled workers; however, the literature on Mexico documents regional wage dispersion due to differential access to FDI and a lack of wage convergence in Mexico compared to the U.S. In this sense, the results obtained here are consistent with the view that FDI does not positively affect the economy as a whole.

In general, the recommended policy is one that attracts FDI while seeking complementarities with the national economic structure, to avoid competition in the goods market that displaces (crowds out) local investment to the detriment of global efficiency. Such a policy should also avoid the unnecessary increase in the price of inputs, such as credit, rents in the commercial and industrial sectors, and skilled labor, which are necessary to create a productive platform in countries with low entrepreneurial development.

REFERENCES

Aitken, B., A. Harrison, and R.E. Lipsey (1996), "Wages and foreign ownership: A comparative study of Mexico, Venezuela, and the United States," *Journal of International Economics* **40**: 345–71.

Agosin, M.R. and R. Mayer (2000), "Foreign direct investment in developing countries: Does it crowd in domestic investment?", United Nations Conference on Trade and Development Discussion Paper 146, in UNCTAD's *World Investment Report 1999*.

Ahmad, E. and A. Malik (2009), "Financial sector development and economic growth: An empirical analysis of developing countries," *Journal of Economic Cooperation and Development* **30**(1): 17–40.

Álvarez E., R. and G. Crespi T. (2001), "La importancia relativa de las pequeñas y medianas mmpresas: Un análisis de sus determinantes en la industria manufacturera" (The relative importance of small and medium enterprises: An analysis of its determinants in the manufacturing industry), *Latin American Journal of Economics/Cuadernos de Economía* **38**(115): 347–65.

Barro, R.J. and X. Sala-i-Martin (2003), *Economic Growth*. Cambridge: MIT Press.

Blinder, A.S. (1997), "Is there a core of practical macroeconomics that we should all believe?" *The American Economic Review* **87**(2): 240–43.

Calderón, C. and R. Fuentes (2006), "Complementarities between institutions and openness in economic development: Evidence for a panel of countries," *Latin American Journal of Economics/Cuadernos de Economía* **43**: 49–80.

Economic Commission for Latin America and the Caribbean - ECLAC (2012). "Panorama de la inserción internacional de América Latina y el Caribe. Crisis duradera en el centro y nuevas oportunidades para las economías en desarrollo" (Overview of the integration of Latin America and the Caribbean: Lasting crisis in the center and new opportunities for developing economies). Santiago: ECLAC – United Nations.

World Bank (2008), "The growth report: Strategies for sustained growth and inclusive development." Commission on Growth and Development, Washington, D.C.

De Backer, K. and L. Sleuwaegen (2003), "Does foreign direct investment crowd out domestic entrepreneurship?" *Review of Industrial Organization* **22**: 67–84.

Desai, M.A., C.F. Foley, and J.R. Hines Jr. (2005), "Foreign direct investment and the domestic capital stock," *American Economic Review* **95**(2): 33–8.

Durand, C. (2005), "Los límites de la inversión extranjera directa (FDI) como fuente de ideas para el crecimiento de las economías en desarrollo" (The limits of foreign direct investment (FDI) as a source of ideas for the growth of developing economies), *Revista Problemas del Desarrollo* **36**(140), Instituto de Investigaciones Económicas, UNAM, Mexico.

Dussel Peters, E. (ed.) (2007), *Inversión extranjera directa en México: desempeño y potencial. Una perspectiva macro, mesa, micro y territorial* (Foreign direct investment in Mexico: performance and potential. A macro, meso, micro, and territorial perspective). Mexico: Siglo XXI Editores.

Egger, H. and U. Kreickemeier (2009), “Firm heterogeneity and the labor market effects of trade liberalisation,” *International Economic Review* **50**(1): 187–216.

Judson, R.A. and A.L. Owen (1999), “Estimating dynamic panel data models: A guide for macroeconomists,” *Economics Letters* **65**(1): 9–15.

Hanson, G.H. (2003), “What has happened to wages in Mexico since NAFTA? Implications for hemispheric free trade,” NBER Working Paper 9563.

Ito, T. (2010), “NAFTA and productivity convergence between Mexico and the U.S.,” *Latin American Journal of Economics/Cuadernos de Economía* **47**: 15–55.

Larraín, F.B. and J. Tavares (2004), “Does foreign direct investment decrease corruption?”*Latin American Journal of Economics/Cuadernos de Economía* **41**: 217–230.

Leamer, E.E. and C.F. Thornberg (2000), “Effort and wages: A new look at the interindustry wage differentials,” R. C. Feenstra, ed., *The impact of international trade on wages*. Chicago: University of Chicago Press.

Lipsey, R.E. (2004), “Home- and host-country effects of foreign direct investment,” in R. E. Baldwin and L. A. Winters, eds., *Challenges to Globalization: Analyzing the Economics*. Chicago: University of Chicago Press.

Ortiz, E. (2007), *Políticas de cambio estructural en la economía mexicana: Evaluación y perspectivas para un nuevo proyecto de nación* (Structural change policies in the Mexican economy). Mexico: UAM-X.

Petri, P.A. (2012), “The determinants of bilateral FDI: Is Asia different?” *Journal of Asian Economics* **23**(3): 201–09.

Pissarides, C.A. (2009), “The unemployment volatility puzzle: Is wage stickiness the answer?” *Econometrica* **77**(5): 1339–69.

United Nations Conference on Trade and Development - UNCTAD (2010), *World Development Report 2010*. New York and Geneva: United Nations.

Waldkircha, A., P. Nunnenkamp, and J.E. Alatorre Bremonta (2009), “Employment effects of FDI in Mexico’s non-maquiladora manufacturing,” *Journal of Development Studies* **45**(7): 1165–83.

APPENDIX A

Table A1. Panel descriptive statistics, 1993-2008

	k	k_f	X_1	X_2
Mean	20,349	3,339	4.9	18,200
Std. dev.	23,648	6,130	1.8	32,846

Source: Author's calculations using data from the National Institute of Statistics and Geography (INEGI) and the National Registry of Foreign Investment (RNIE), Mexico.
Note: X_1 is the average firm size as measured by the number of workers and X_2 is the average labor productivity. The figures are expressed in dollars, except X_1, which denotes the number of people.

Table A2. Correlation matrix, 1993-2008

	k	k_f	X_1	X_2
k	1.000			
	—			
k_f	0.171	1.000		
	(0.054)	—		
X_1	0.213	0.067	1.000	
	(0.016)	(0.455)	—	
X_2	0.867	0.020	0.115	1.000
	(0.000)	(0.819)	(0.196)	—

Source: Author's calculations using data from INEGI and RNIE.
Note: P-values in parentheses.

Table A3. Partial correlation with k, 1993-2008

	Simple correlation	Partial analysis controlling for		
	k	$k \mid X_1$	$k \mid X_2$	$k \mid (X_1, X_2)$
k_f	0.171	0.161	0.308	0.302
	(0.054)	(0.071)	(0.0004)	(0.0006)
X_1	0.213	—	0.229	—
	(0.016)	—	(0.010)	—
X_2	0.867	0.868	—	—
	(0.000)	(0.000)	—	—

Source: Author's calculations using data from INEGI and RNIE.
Note: P-values in parentheses.

Table A4. Panel descriptive statistics, 2003-2010

	log(w)	k^f	L	log(y)	log(E)	$U6$
Mean	8.381	2,665	1,309,425	4.70	11.40	10.2
Std. dev.	0.238	5,105	1,141,859	0.15	0.78	2.8

Source: Author's calculations using information from the National Institute of Statistics and Geography (INEGI) and the National Registry of Foreign Investment (RNIE), Mexico.
$k^f = K^f$ / L in dollars. K^f is the cumulative flow of FDI in dollars; L is the number of people employed.

Table A5. Unit root tests

(probability in parentheses)

Level case				1st differenced case			
		Common unit root	Individual unit root			Common unit root	Individual unit root
Variable	Situation	LLC	IPS	Variable	Situation	LLC	IPS
log(w)	a)	-5.417 (0.000)	-3.756 (0.000)	log(w)	a)	-24.318 (0.000)	-27.739 (0.000)
	b)	-1.283 (0.100)	1.666 (0.952)		b)	-25.105 (0.000)	-28.606 (0.000)
k_f	a)	-0.689 (0.245)	5.486 (1.000)	k_f	a)	-17.045 (0.000)	-17.257 (0.000)
	b)	0.962 (0.832)	1.993 (0.977)		b)	-16.324 (0.000)	-15.603 (0.000)
log(y)	a)	-3.4828 (0.0002)	-2.8204 (0.002)	log(y)	a)	-17.882 (0.000)	-20.261 (0.000)
	b)	-4.980 (0.000)	-2.237 (0.0127)		b)	-14.765 (0.000)	-17.390 (0.000)
	c)	3.324 (1.000)	20.755 (1.000)		c)	-26.080 (0.000)	655.556 (0.000)
log(E)	a)	-0.918 (0.179)	1.883 (0.970)	log(E)	a)	-21.511 (0.000)	-30.404 (0.000)
	b)	-1.743 (0.041)	-5.763 (0.000)		b)	-18.520 (0.000)	-28.123 (0.000)
	c)	6.785 (1.000)	4.948 (1.000)		c)	-33.806 (0.000)	1152.520 (0.000)
$U6$	a)	-1.438 (0.075)	-1.061 (0.145)	$U6$	a)	-22.671 (0.000)	-27.219 (0.000)
	b)	-2.422 (0.008)	-4.624 (0.000)		b)	-20.243 (0.000)	-24.853 (0.000)
	c)	3.044 (0.999)	13.875 (1.000)		c)	-32.189 (0.000)	1030.690 (0.000)

Source: Author's calculations using data from INEGI and the RNIE, Mexico.
a) Individual intercept, b) individual intercept and trend, c) none.

Table A6. Pedroni residual cointegration test

Null hypothesis: no cointegration

	Trend assumption					
	No deterministic intercept or trend		**Individual intercept + no deterministic trend**		**Deterministic intercept and trend**	
Within-dimension (panel)	**Statistic**	**Prob.**	**Statistic**	**Prob.**	**Statistic**	**Prob.**
v-Stat	-1.83	0.967	0.90	0.185	-1.025	0.847
ρ-Stat	-4.92	0.000	-3.78	0.0001	-2.154	0.016
PP-Stat	-9.23	0.000	-8.24	0.000	-9.152	0.000
ADF-Stat	-3.06	0.001	-0.35	0.363	-1.021	0.154
Between-dimension (group)			Statistic	Prob.	Statistic	Prob.
Group ρ -Stat	-3.53	0.0002	-1.91	0.028	-0.145	0.442
Group PP-Stat	-11.15	0.000	-9.06	0.000	-10.514	0.000
Group ADF-Stat	-3.49	0.0002	-0.75	0.227	-1.341	0.090

Source: Author's calculations using data from INEGI and RNIE, Mexico.
Series: $\log(w)$, k_f, U6, $\log(E)$, $\log(y)$.
At the top, calculated values of within-dimension statistics are shown that combine (pool) the self-regressive coefficients of the different federal entities for the unit root tests of the estimated residuals. At the bottom of the table, the estimated values of between-dimension statistics are shown that average the estimated coefficients for each federal entity.

Vol. 50 No. 2 (NOV, 2013), 233–255

LAJE
LATIN AMERICAN JOURNAL OF ECONOMICS

THE RELATIONSHIP BETWEEN ENERGY CONSUMPTION AND GDP: EVIDENCE FROM A PANEL OF 10 LATIN AMERICAN COUNTRIES*

Jacobo Campo**
Viviana Sarmiento***

We estimate the elasticity of the long-run relationship between energy consumption and GDP for 10 countries in Latin America from 1971 to 2007. We employ Pedroni's (1999, 2004) panel cointegration test to determine if such a long-run relationship exists. Westerlund's (2006) cointegration test for panel data is used to estimate the slopes of the long-run relationship variables. These findings provide empirical guidance for policies to promote energy conservation and efficiency. Cointegration between the two variables is found to exist in both directions. This paper discusses the energy dependence of some countries and describes potential implementation of energy conservation policies in others.

JEL classification: C32, O40, Q43

Keywords: energy consumption, panel stationarity, panel cointegration, Latin America

1. Introduction

In the three decades since Kraft and Kraft's (1978) seminal study, economists and other researchers have studied the relationship between energy consumption and GDP from different perspectives and using diverse methodologies. Their methods have ranged from descriptive time series analysis to cointegration applications with panel data. This paper presents empirical evidence on the long-run relationship between energy consumption and real GDP through the application of a cointegration panel test and estimation methods in both directions, i.e., energy consumption to GDP and GDP to energy consumption.

* We would like to thank the referees and editors Raimundo Soto and Felipe Zurita for their valuable suggestions. We are also grateful to José Fernando Moreno Gutierrez, and the seminar participants at *Banco de la República*, *Departamento Nacional de Planeación* (DNP), *Unidad de Planeación Minero-Energética*, *Instituto Nacional Politécnico de México* and *Universidad Católica de Colombia*.

** Researcher, Professor and Group Head of Finance and Political Economy, *Facultad de Economía*, Universidad Católica de Colombia, Cra 13. # 47 – 49, Bogotá, Colombia. Tel: +57 (1) 3277300. Email: jacampo@ucatolica.edu.co or jacobo.campo@gmail.com.

*** At the time this article was written, she was a research assistant at the *Facultad de Economía*, *Universidad Católica de Colombia*. She is now a research assistant at CAMACOL in Bogotá, Colombia. Email: lvsarmiento48@ucatolica.edu.co or vivianasarmientog@gmail.com.

doi 10.7764/LAJE.50.2.233

Governments are interested in studying this relationship because of the international goal of curbing the increase in global temperature to a maximum of 2ºC in the context of global warming. To achieve this goal, it has become necessary to assess the impacts of policies that promote energy conservation and efficiency on national GDP and economic growth. According to the International Energy Agency (IEA), "80% of emissions from the energy sector that were planned for 2020 have already been reached and 40% of CO^2 emissions from OECD countries." This accelerated trend in global emissions represents a step backwards in the battle against global warming. However, some policy makers are concerned that conservative energy efficiency policies, such as those designed and formulated according to researchers' suggestions, may have disastrous results on their countries' economic activity.

The first purpose of this paper is to estimate the relationship between energy consumption and GDP for 10 countries in Latin America. Secondly, we compare these estimates at a regional level and compare them with estimates obtained in other studies. We do this by using recent panel data methods, including developments in the second generation of cointegrated panel data, which accounts for structural breaks and cross-sectional dependence in the long-run relationship. For this method of analysis, we apply unit root tests according to the methodology proposed by Westerlund (2006).

This paper is organized as follows. The second section presents a brief literature review on the topic of energy consumption and GDP. Section 3 presents the methodology used in this study and the data analyzed in our paper. Section 4 presents the estimates and results of our model and compares these results with estimates obtained in previous works. The fifth and final section concludes.

2. Energy consumption and GDP: A brief literature review

Ozturk (2010), Squalli (2007) and Magazzino (2011) provide four hypotheses about the direction of causality between energy consumption and GDP. The first is the hypothesis of neutrality, which holds that there is no causality (in either direction) between these two variables. The second is the energy conservation hypothesis, which holds that there is evidence of unidirectional causality from GDP growth to energy consumption. Under the third hypothesis, which is known as the growth hypothesis, energy consumption drives GDP growth. The fourth hypothesis is the feedback hypothesis, which suggests a bidirectional causal relationship between energy consumption and GDP growth.

Several authors have studied the relationship between GDP and energy consumption and have found varying results. For example, Kraft and Kraft (1978) found unidirectional causality from GDP to energy consumption in the United States for the 1947-1974 period. Abosedra and Baghestani (1991) proved the assertion put forth by Kraft and Kraft (1978) by using the standard test of Granger causality. However, Akarca and Long (1980) argue that the results of Kraft and Kraft (1978) are spurious because they found no evidence of causality when the time period is bounded by two-year intervals. Yu and Hwang (1984) and Yu and Choi (1985) also found no causal relationship between energy consumption and GDP, although they used several different methods.

Erol and Yu (1987) found bidirectional causality between energy consumption and real GDP in Japan and Italy, a one-way causal relationship in East Germany, and a neutral relationship in France, the United Kingdom and Canada using the Granger causality test. Sims (1972) and Hwang and Gum (1992) found evidence of bidirectional causality between energy consumption and GDP in Taiwan. Fatai (2002) found no relationship between energy consumption and GDP in New Zealand.

In a study by Al-Iriani (2005) for the six countries comprising the Gulf Cooperation Council (Kuwait, Oman, Saudi Arabia, Bahrain, United Arab Emirates, and Qatar), the results indicate that there is unidirectional causality from GDP to energy consumption; Soytas *et al.* (2001) found evidence of unidirectional causality from energy consumption to GDP in Turkey using a model of cointegration and vector error correction analysis.

Narayan *et al.* (2010) examine the long-run elasticities of the impacts of energy consumption on GDP in addition to the impacts of GDP growth on energy consumption for 93 countries during the time period from 1980 to 2006. They apply unit root tests and the cointegration test of Pedroni (1999, 2004) to calculate long-run elasticities between energy consumption and GDP and GDP and energy consumption. Lee (2005) estimates elasticities based on a capital-driven production function, finding a significant coefficient in the direction of energy consumption to GDP. Later in this paper, we compare the results obtained from our model with the estimates obtained by other researchers in the field.

3. Methodology and data

This paper analyzes information on energy consumption and GDP in 10 countries for the time period between 1971 and 2007. The methodology for estimating long-run elasticities in energy consumption and GDP relationship is given by Westerlund (2006). In this case, we propose estimating the following relationships to find the (β_i) for each of the following equations. We will use the index $i = 1,2,3,\ldots,N$ to denote countries, and $t = 1,2,3,\ldots,T$ to denote time. Consider the following panel regressions:

$$GDP_{it} = \alpha_i + \beta_i EC_{it} + \varepsilon_{it} \tag{1}$$

$$EC_{it} = \alpha_i + \beta_i GDP_{it} + \varepsilon_{it} \tag{2}$$

GDP_{it} is the real GDP for each country (i) during the period (t), and EC_{it} is the energy consumption (both variables are scaled in per-capita units). Note that (β_i) is a country-specific slope (elasticity in this case) that is assumed to be constant over the period from 1971 to 2007, and (ε_{it}) is the term error.

To avoid spurious regressions, we test the stationarity of the series, which requires the use of panel unit root tests. Next, if the series are not stationary, i.e., if the series have unit roots, then we test whether there is a long-run relationship between the relevant variables. Then we establish the following procedure. First, we apply a panel unit root test to determine the integration order of each variable. Second, we apply the panel cointegration test used by Pedroni (1999, 2004) to determine whether there is a long-run relationship between the model's variables. Finally, we apply the panel cointegration test used by Westerlund (2006) to determine the existence of a long-run relationship between variables, taking into account the cross-sectional dependence and possible structural breaks in the long-run relationship. As a result of the last test, we estimate the long-run coefficients (elasticities) of the variables.

3.1. Panel unit root tests

We begin by considering the integration order of the energy consumption and real GDP series using panel unit root tests, following methods used by IPS (2003), Levin, Lin, and Chu (2002), Bruiting (2000), Maddala and Wu (1999) (Fisher-type ADF), Choi (2001) (Fisher-type PP), and Hadri (2000). The advantage of the methods used by Maddala and Wu (1999) and IPS (2003) is that these papers relax the assumption of homogeneity. Choi (2001) models cross-sectional dependence by common factors, considering a homogeneous AR(1) model.

Panel unit root tests are theoretically grounded in a time series approach. Both theory and literature suggest that panel data unit root tests offer advantages over time series data, primarily because panel data combine cross-sectional data units and time series, providing a greater number of degrees of freedom and improving statistical efficiency. Additionally, this approach successfully mitigates the problem of bias that is caused by unobserved heterogeneity in the regression.

3.2. Panel cointegration tests

Having established that the two variables are I(1), i.e., that the variables contain a panel unit root, we next test whether there is a long-run relationship between the variables. First, we follow Pedroni (1999, 2004) by applying a heterogeneous panel cointegration test (first-generation test). Then, we follow Westerlund (2006) by applying a heterogeneous panel cointegration test with multiple structural breaks (second-generation test).

Westerlund (2006) proposes a cointegration test for the null cointegration hypothesis, based on a Lagrange multiplier (LM). This test allows for multiple structural breaks in both the level and trend of a panel regression. Following Westerlund (2006), the panel LM test statistic for this particular hypothesis is defined as follows:

$$Z\left(M\right)=\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M_i+1}\sum_{t=T_{ij-1}+1}^{T_{ij}}\frac{S_{it}}{\left(T_{ij}-T_{ij-1}\right)^2\hat{\sigma}_i^2} \tag{3}$$

In this equation, $S_{it}=\sum_{s=T_{ij-1}+1}^{t}\hat{\varepsilon}_{it}$, and $\hat{\varepsilon}_{it}$ is the regression error obtained by using the fully modified OLS (FMOLS) estimator. Westerlund (2006) shows that statistic $Z(M)$, which is standardized by its mean and standard deviation, has an asymptotic, standard normal distribution under the null hypothesis. Using the procedure from Bai and Perron (2003), Westerlund (2006) estimates the number of breaks and the locations of the breaks in the sample for each country. Westerlund (2006) accounts for the impact of cross-country dependence and suggests using the bootstrap method. In this case, we use Davison and Hinkley's (1997) block method of bootstrapping. The advantage of this procedure is that the breaks are determined endogenously, whereas in other approaches the breaks are chosen exogenously.

The empirical analysis is based on a panel of 10 countries (Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay and Venezuela) in Latin America during the 1971 -2007 period. We use a series of per-capita energy consumption and per-capita GDP data from the World Bank. Table 1 presents descriptive statistics of the series for each country.

The per-capita energy consumption series for the 10 countries studied follows a normal distribution, except for Peru. It should be noted that Venezuela has the highest average per-capita power consumption in the region (2276.667 kWh), with a standard deviation of 584.2418 kWh. Bolivia reports the lowest average in the region with 311.0732 kWh consumed per person annually, and a standard deviation of 95.23 kWh (See Table 1).

On the other hand, the real per-capita GDP series data follow a normal distribution for all countries except Paraguay. However, the inclusion of structural breaks endogenously estimated and the Westerlund (2006) test allow for modeling behavior in the series that is not normally distributed. Argentina has a real per-capita GDP that is higher than the regional average at USD 7,121.8 with a standard deviation of USD 749.552. In contrast, Paraguay has an average per-capita GDP of USD 1,284.95. This is the lowest figure amongst all the countries, with a standard deviation of USD 195.7014 (see Table 2).

Figure 1 presents graphs of the variables for each of the countries in the sample, both in logarithms and in first differences. In general, these graphs show that in some periods, the variables appear to have strong relationships, but in other periods the directionality of the relationships is reversed. These stylized facts give us an idea of the close relationship between energy consumption and GDP in the Latin American countries under study. Moreover, the graphs in Figure 1 show comovements between the growth rates of the two sets in each of the countries.

Table 1. Descriptive statistics: real GDP

(in per-capita runs)

Country	Mean	Median	Maximum	Minimum	Std. dvs.	Kurtosis	Jarque-Bera	Probability
Argentina	7121.80	7002.48	9359.59	5606.86	749.55	4.07	4.65	0.098
Bolivia	2637.92	2676.59	3049.45	2211.73	253.74	1.94	1.77	0.412
Brazil	3395.72	3509.45	4290.47	2162.34	438.08	3.96	4.97	0.083
Chile	3466.17	3010.49	6077.35	1889.51	1329.65	1.76	3.91	0.141
Colombia	2247.99	2236.45	3082.79	1538.74	387.29	2.15	1.15	0.563
Ecuador	1319.12	1315.13	1680.46	959.73	141.88	4.77	4.82	0.090
Paraguay	1284.95	1346.53	1486.95	797.83	195.70	3.49	11.02	0.004
Peru	2102.74	2101.48	2692.17	1620.56	234.64	3.27	0.30	0.860
Uruguay	5815.09	5212.62	8060.65	4407.32	990.22	2.14	1.91	0.385
Venezuela	5333.07	5175.67	6521.48	3966.50	634.14	2.31	1.30	0.523

Source: Authors' calculations.

Table 2. Descriptive statistics: energy consumption (in per-capita runs)

Country	Mean	Median	Maximum	Minimum	Std. dvs.	Kurtosis	Jarque-Bera	Probability
Argentina	1537.34	1352.62	2658.67	870.26	495.37	2.38	3.56	0.168
Bolivia	311.07	275.44	514.88	176.12	95.23	2.00	2.80	0.247
Brazil	1356.85	1451.40	2170.68	495.96	492.59	1.96	2.21	0.330
Chile	1570.30	1218.02	3318.20	743.35	836.55	2.20	5.02	0.081
Colombia	746.37	810.13	976.75	401.46	175.56	2.09	4.18	0.123
Ecuador	471.50	449.53	788.12	146.81	187.86	1.96	1.84	0.398
Paraguay	495.66	443.80	958.27	81.28	300.89	1.42	3.95	0.139
Peru	582.13	554.46	960.80	398.33	133.60	3.64	7.51	0.023
Uruguay	1341.61	1246.15	2196.51	686.88	472.98	1.71	2.97	0.226
Venezuela	2276.67	2480.28	3098.44	1113.05	584.24	2.37	4.67	0.097

Source: Authors' calculations.

Figure 1. Energy consumption and GDP and first differences

(in logarithms)

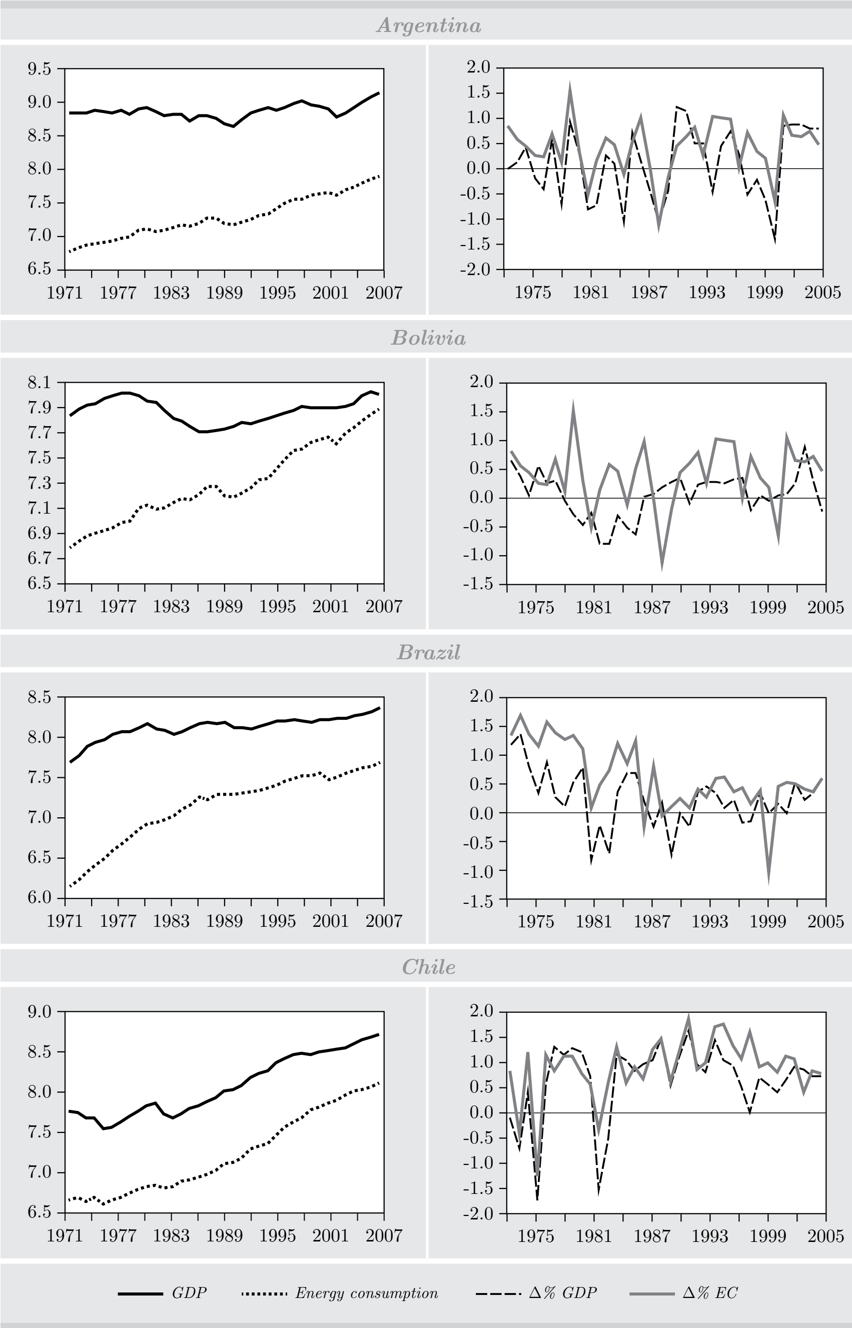


Figure 1. (continued)

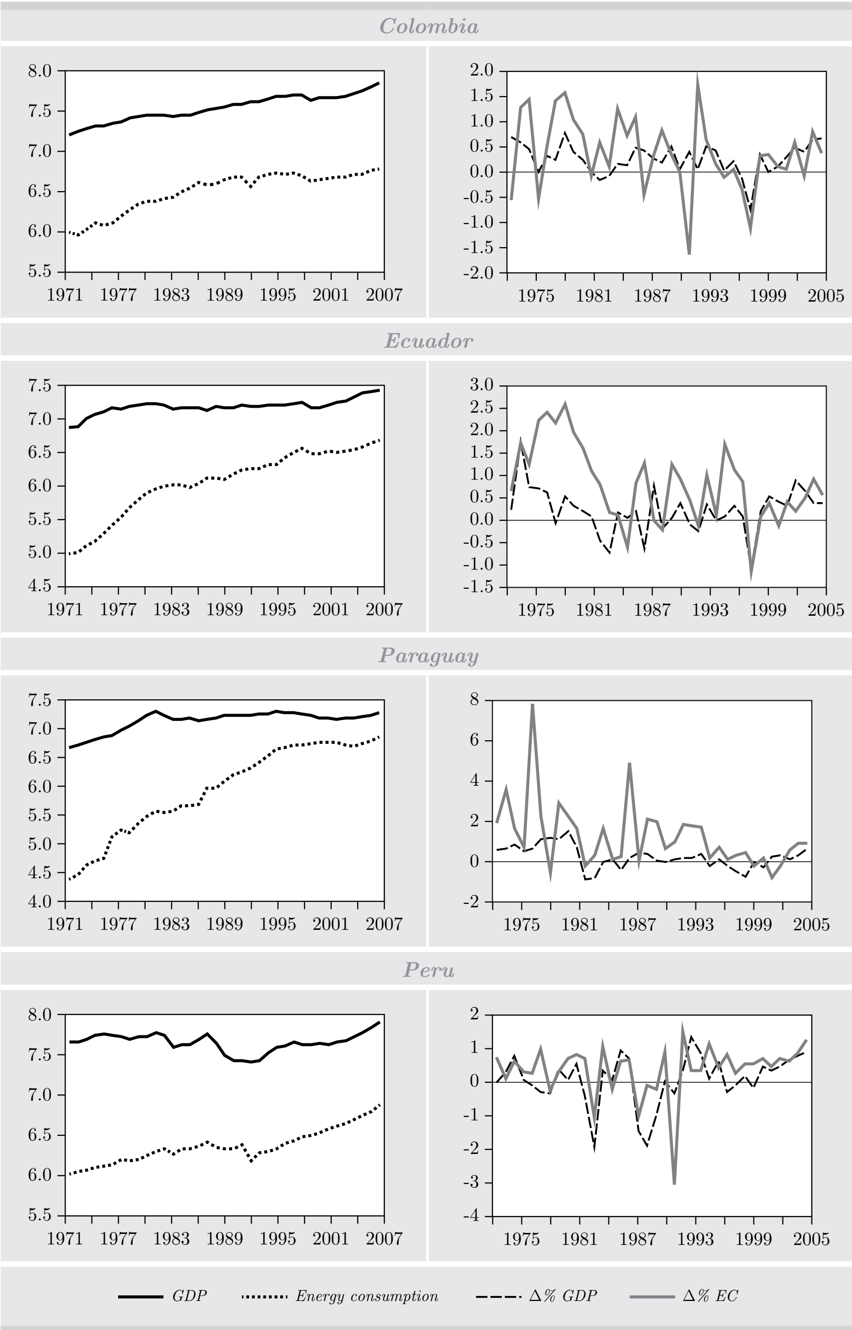
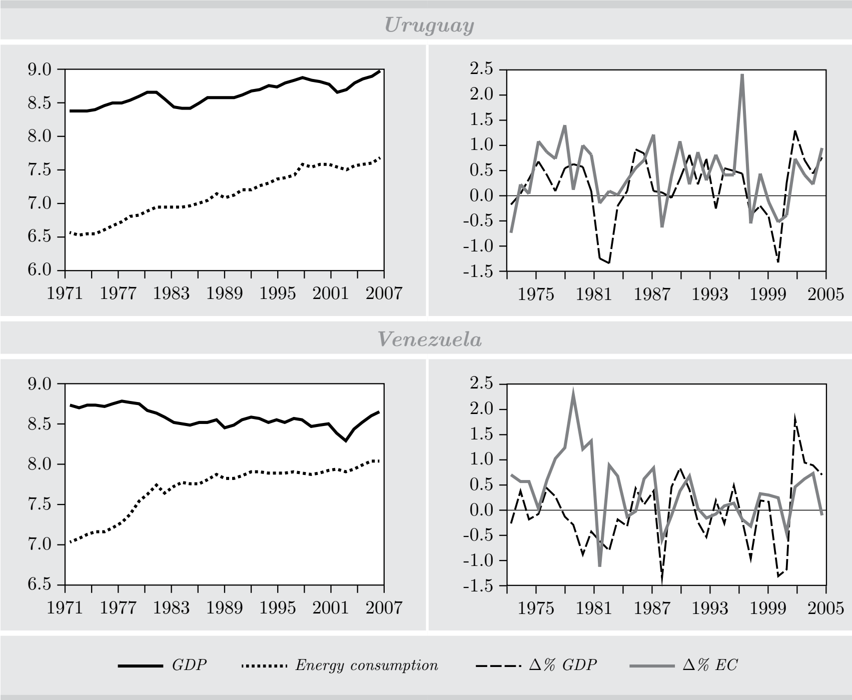


Figure 1. (continued)



Source: World Bank, 2011.

4. Estimations and results

In this section, we present the results obtained by using the procedure explained in Section 3. Specifically, we discuss the results of panel unit root tests, the results of the panel cointegration tests, and the estimated long-run elasticities for each country and for the regional subsamples of countries.

4.1. Panel unit root tests

The results of the panel unit root tests are reported in Table 3 and Table 4. In Table 3, the results of the unit root analysis suggest that for almost all panel unit root tests, at the 5% significance level, the logarithm of energy consumption and the logarithm of real GDP are non-stationary for the panel. For the two variables, the null hypothesis of a unit root cannot be rejected at these levels. For IPS,

Table 3. Panel unit root tests

(levels)

Test	Energy consumption		GDP	
	Statistic	Probability	Statistic	Probability
Im, Pesaran, and Shin (W-statistic)	1.063	0.856	-0.081	0.467
ADF - Fisher (χ^2)	12.474	0.898	25.527	0.182
PP - Fisher (χ^2)	11.803	0.923	21.564	0.364
Levin, Lin & Chu (t-statistic)	-1.88**	0.030	-0.809	0.209
Breitung (t-statistic)	1.497	0.933	-2.056**	0.019
Hadri (Z-statistic)	8.069***	0.000	4.918***	0.000

Notes: ***(**) denote statistical significance at the 1% (5%) level. The optimal lag length was selected automatically using the Schwarz criterion.

Table 4. Panel unit root tests

(first difference)

Test	Δ(Energy consumption)		ΔGDP	
	Statistic	Probability	Statistic	Probability
Im, Pesaran, and Shin (W-statistic)	-6.7339***	0.000	-8.016***	0.000
ADF - Fisher (χ^2)	85.3998***	0.000	101.49***	0.000
PP - Fisher (χ^2)	168.42***	0.000	102.68***	0.000
Levin, Lin & Chu (t-statistic)	-5.333***	0.000	-8.243***	0.000
Breitung (t-statistic)	-6.253***	0.000	-6.098***	0.000
Hadri (Z-statistic)	4.911***	0.000	1.021	0.153

Notes: ***(**) denote statistical significance at the 1% (5%) level. The optimal lag length was selected automatically using the Schwarz criterion.

ADF (Fisher type), PP (Fisher type), Bruiting, and Hadri tests, the null of stationarity is rejected at the 1% significance level.

Table 4 presents the results of the unit root tests performed on the series in first differences. The purpose of this test is to verify the existence of additional unit roots, thus determining the order of integration of the series. The results indicate that the difference series are stationary, which implies that the series in levels are integrated by one order, i.e., are I(1).

4.2. Panel cointegration test

Having established that the two variables are I(1), i.e., that they contain a panel unit root, we proceed to test whether there is a long-run relationship between the two variables. First, we use the Pedroni (1999, 2004) heterogeneous panel cointegration test, and then the Westerlund (2006) heterogeneous panel cointegration test with multiple structural breaks.

Table 5 reports the panel cointegration estimation results. We reject the null hypothesis of no cointegration only for the panel v-statistic. Furthermore, for the relationship between GDP and energy consumption (GDP-EC), none of the statistics allow us to reject the null hypothesis of no cointegration. This finding may have occurred because this test can lead researchers to mistakenly fail to reject the null hypothesis because the test excludes the possible existence of cross-country dependence and the possible structural breaks in the cointegration relationship. For this reason, we choose to apply the Westerlund (2006) test to control for these factors.

Table 5. Pedroni (1999, 2004) panel cointegration test

Within-dimension	**EC → GDP**		**GDP → EC**	
	Statistic	***p*-value**	**Statistic**	***p*-value**
Panel v-statistic	2.721***	0.003	-1.807	0.965
Panel Phillips-Perron type ρ-statistic	-0.549	0.292	1.168	0.879
Panel Phillips-Perron type t-statistic	-0.316	0.376	1.047	0.852
Panel ADF type t-statistic	-1.734**	0.041	0.808	0.791
Between-dimension				
Group Phillips-Perron type ρ-statistic	0.508	0.694	1.847	0.968
Group Phillips-Perron type t-statistic	0.406	0.658	1.653	0.951
Group ADF type t-statistic	-1.281	0.100	1.349	0.911

Notes: ***(**) denote statistical significance at the 1% (5%) level. The optimal lag length was selected automatically using the Schwarz criterion.

Table 6 reports results from the Westerlund (2006)[1] heterogeneous panel test with multiple breaks, which accommodates cross-sectional

1. We are grateful to Professor Westerlund for kindly providing the GAUSS code, which we used to estimate these results using R-Project software.

dependence using the block bootstrapping method developed by Davison and Hinckley (1997). Using a block size of 5 over 1,000 replications, the bootstrap p-value for the null hypothesis of cointegration in both cases is 0.98 and 0.93 respectively, indicating rejection of the null hypothesis at any conventional level of significance.

Table 6. Westerlund (2006) panel cointegration test

About the method		
Estimation method	FMOLS	
Maximun number of breaks	5	
Bootstrap method	Davison and Hinkley Blocks	
Block type	Blocks with geometric distribution	
Block size	5	
Bootstrap replications	1,000	
	EC → GDP	**GDP → EC**
LM statistic with bootstrapping	10,671	7,357
The p-value bootstrap	0.986	0.938

These results provide empirical evidence about the existence of a bidirectional causality, since when estimating long-run relationships through cointegration tests, the magnitude (positive or negative) of the coefficients represents the direction of the causality (Granger, 1983; Engle and Granger, 1987).

Table 7 presents the estimated breaks that are present in the long-run relationship between energy consumption and GDP. It also presents specific breakdowns for the relationship of energy consumption to GDP and GDP to energy consumption. These results suggest that there is strong relationship instability between energy consumption and GDP in Latin American countries during the study period. Note that the Westerlund test detected 35 breaks in the panel for the energy consumption-GDP relationship and detected 36 breaks in the GDP-energy consumption relationship.

The estimations of the long-run elasticities are reported in Table 8. The first column lists the countries, the second column shows the long-run elasticity of energy consumption's impact on GDP, accompanied by its standard deviation, and the last column reports the long-run elasticity of GDP's impact on energy consumption and its standard deviation. For each of the 10 Latin American countries and for the

Table 7. Estimated breaks

Matrix structural breaks (energy consumption → GDP)						
Country	**Breaks**	**B1**	**B2**	**B3**	**B4**	**B5**
Argentina	3	1988	1994	2002	—	—
Bolivia	5	1975	1981	1987	1997	2002
Brazil	4	1980	1985	1994	2000	—
Chile	3	1986	1993	2000	—	—
Colombia	2	1976	1986	—	—	—
Ecuador	3	1975	1980	1995	2002	—
Paraguay	4	1975	1986	1994	—	—
Peru	2	1981	1991	—	—	—
Uruguay	5	1976	1981	1987	1992	1997
Venezuela	4	1976	1981	1988	1997	—
Matrix structural breaks (GDP → energy consumption)						
Argentina	3	1987	1994	2002	—	—
Bolivia	3	1977	1986	2000	—	—
Brazil	3	1975	1982	2000	—	—
Chile	4	1975	1981	1995	2002	—
Colombia	4	1979	1985	1997	2002	—
Ecuador	3	1981	1995	2002	—	—
Paraguay	5	1975	1981	1987	1996	2001
Peru	4	1982	1988	1993	2002	—
Uruguay	4	1979	1984	1996	2001	—
Venezuela	3	1978	1983	2001	—	—

Note: Westerlund's structural breaks (2006) are estimated using the Bai and Perron (2003) procedure with a maximum number of five breaks for each country.

regional subsample (panel), the long-run elasticities of the impact of energy consumption on real GDP indicate a bidirectional, long-run relationship between real GDP and energy consumption. Moreover, a 1 percent increase in energy consumption increases real GDP by 0.59% across the entire panel; a 1 percent increase in real GDP increases energy consumption by 0.59% across the entire panel. As mentioned previously, these results represent bidirectional causality, because when long-run relationships are estimated using cointegration tests, the magnitude (positive or negative) of the coefficients represents the direction of the causality (Granger, 1983; Engle and Granger, 1987). For example, Lee and Chang (2007) explore the direction of causality between energy consumption and GDP through a dynamic panel in a sample of developed and developing countries.

Table 8. Panel cointegration slope estimates

(long-run elasticities)

Country	$GDP_{it} = \alpha_i + \beta_i EC_{it} + \varepsilon_{it}$		$EC_{it} = \alpha_i + \beta_i GDP_{it} + \varepsilon_{it}$	
	Elasticity (slope)	**S.D.**	**Elasticity (slope)**	**S.D.**
Argentina	1.433	0.108	0.533	0.039
Bolivia	0.214	0.073	0.18	0.141
Brazil	1.236	0.002	0.189	0.052
Chile	1.141	0.002	0.4	0.058
Colombia	0.154	0.068	0.815	0.002
Ecuador	0.204	0.031	0.739	0.003
Paraguay	0.115	0.027	0.693	0.004
Peru	0.379	0.180	0.786	0.003
Uruguay	0.367	0.094	0.783	0.004
Venezuela	0.655	0.232	0.82	0.003
Regional	**0.590**	**0.082**	**0.594**	**0.031**

Note: Regional slope is an average of all 10 countries.

On a per-country basis, the cointegration results show that for all countries the slopes have a positive sign and are statistically significant. In other words, energy consumption has a positive impact on GDP. In the cases of Argentina, Brazil and Chile, energy consumption exerts a positive and elastic effect on real GDP, whereas in the cases of Colombia and Paraguay energy consumption has a positive but smaller effect on real GDP.

Regarding the relationship between real GDP and energy consumption, these findings show that for all countries, GDP has a positive effect on energy consumption. In the cases of Colombia and Venezuela, a 1% increase in real GDP increases energy consumption by 0.82%. This result demonstrates compliance with the feedback hypothesis, which holds that there is a bidirectional relationship between these two variables. Other studies such as Narayan *et al.* (2010) and Lee (2005) also estimated long-run elasticities for some of the countries included in our sample. Table 9 presents these authors' results. We observe that the results of Narayan *et al.* (2010) are similar to those obtained in this paper for Argentina and Venezuela. However, our results differ from the estimates shown in Table 9, particularly for Colombia, but these results are not statistically significant. Finally, the results of Lee (2005) are similar to our estimates, but differ in the case of Venezuela.

Table 9. Other results of long-run elasticities

Country	Narayan et al (2010)		Lee (2005)
	EC → GDP	GDP → EC	EC → GDP
	Elasticity	Elasticity	Elasticity
Argentina	1.23**	0.43**	0.84**
Bolivia	1.95**	0.25**	—
Brazil	0.29	0.26	—
Chile	0.72**	1.34**	0.83**
Colombia	-0.7	-0.11	1.53**
Ecuador	1.05**	0.44**	—
Paraguay	—	—	—
Peru	1.42**	0.44**	0.96**
Uruguay	1.72**	0.27**	—
Venezuela	0.85**	1.09**	0.58**
Regional	0.91**	0.43**	—

Note: ** denotes statistical significance at the 5% level.

The results shown in Table 8 can be understood by analyzing the evolution of productive economic structures of South American countries, i.e., how the shares of their primary, secondary, and tertiary sectors changed during the sample period. Table 10 shows the value added for each sector of the economy as a percentage of GDP for the 10 countries included in the study. We can see that in recent decades the primary sector represents a smaller share of a nation's GDP and that secondary and tertiary sectors represent larger GDP shares. Argentina, Brazil and Chile have the largest elasticities reported in Table 8. This signifies that these countries' economies are goods-intensive in the secondary and tertiary sectors (industry and services). In the cases of Paraguay and Colombia, countries with lower elasticities, we can see that during the 1970s and the past decade, both countries reduced the share of their primary sectors and increased participation of their tertiary sectors.

In particular, Table 10 demonstrates that countries with a high GDP share generated in the primary sector (in relative terms) in the 1970s and a low share in the 1990s should have low elasticities in the relationship between energy consumption and GDP. Conversely, countries whose GDP share from the primary sector has declined significantly should show high elasticities in the relationship between energy consumption

Table 10. Value added by economy sector

Country	1970 - 1979. (% of GDP)			1980 - 1989. (% of GDP)			1990 - 1999. (% of GDP)			2000 - 2009. (% of GDP)		
	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary
Argentina	9.18	45.97	44.85	8.16	39.92	51.92	5.94	29.93	64.13	8.64	32.56	58.80
Bolivia	20.66	34.11	45.23	18.99	34.04	46.97	16.36	32.04	51.60	14.37	32.42	53.21
Brazil	12.69	39.33	47.98	10.45	44.55	45.01	6.87	32.64	60.49	6.12	27.83	66.05
Chile	7.62	39.97	52.40	7.59	38.65	53.77	8.08	38.12	53.80	4.62	41.53	53.85
Colombia	24.35	30.09	45.57	18.42	34.67	46.91	15.10	32.62	52.28	8.33	32.18	59.49
Ecuador	19.11	33.33	47.56	13.88	39.28	46.84	15.06	29.60	55.34	10.39	31.89	57.73
Paraguay	34.18	21.87	43.96	27.86	24.06	48.08	21.89	23.85	54.27	19.90	21.40	58.70
Peru	15.78	34.53	49.69	9.96	32.60	57.44	8.81	29.41	61.78	7.52	33.23	59.25
Uruguay	—	—	—	12.99	34.78	52.23	8.09	29.67	62.24	9.79	25.82	64.39
Venezuela	5.15	44.17	50.68	5.87	48.22	45.92	5.23	49.83	44.94	4.25	51.73	44.02

Source: World Bank data, authors' calculations.

and GDP. One exception to this behavior is Argentina, which saw a small change in relative terms.

5. Conclusions

It is important to understand the effects of increased energy consumption on GDP so that economic policy makers can predict the impacts of implementing energy policies on a country's GDP. Our evidence reflects the existence of panel stationarity for Latin American countries, and our panel attests to bidirectional causality between energy consumption and GDP in all sample countries.

The literature investigates the impact of energy consumption on GDP for many countries using different techniques and methodologies. The results of these studies show that different methodologies lead to confusing and contradictory conclusions about this relationship. This paper estimates the elasticity of the long-run relationship of energy consumption-GDP and GDP-energy consumption for 10 countries in Latin America during the period from 1971 to 2007. We employ Pedroni's (1999, 2004) panel cointegration test to determine if a long-run relationship exists between the variables in equations (1) and (2). By using a cointegration test for panel data developed by Westerlund (2006), which accounts for possible cross-sectional dependence between countries and any existing structural breaks in the long-run relationship, we identify the long-run elasticities. In the sections above, we provide empirical evidence about policy maker's abilities to design and implement programs to promote energy conservation and efficiency. For example, a 1% increase in energy consumption increases real GDP by 0.59% across the entire panel, while a 1% increase in real GDP increases energy consumption by 0.59% across the entire panel.

In this case, because there is a long-run relationship between energy consumption and GDP, it is understood that in the long run energy generates economic growth for Latin American countries. In the cases of Bolivia, Colombia, Ecuador, Paraguay and even Peru and Uruguay, the elasticity of energy consumption is low (below the regional average). In these countries, policy makers could implement energy conservation programs with low negative impacts in the short run. However, if there is truth to the feedback hypothesis, which suggests that energy consumption and GDP are interrelated and complementary over time in a bidirectional, causal relationship, then policies that promote the energy efficiency do not negatively affect GDP. In addition, according to the results of our panel stationarity tests, if shocks in energy consumption and GDP are temporary, stabilization policies will power long-lasting effects in the countries of Latin America. Finally, the result of our cointegration test suggests that energy consumption and GDP are endogenous variables in Latin American countries at the rate of the bidirectionality of causality. Another interesting result is that the methodology is better than those used previously, in the sense that it reflects the presence of structural breaks, controls endogeneity and includes the presence of cross-correlation between the countries concerned.

Countries such as Argentina, Brazil and Chile are energy-dependent, which means that policies that seek to conserve energy in the long run would have disastrous results on their economic growth. Additionally, this dependence on the part of some Latin American countries indicates the need to diversify energy sources, since those countries must weigh the need for sustainable economic growth against the environmental costs associated with excessive energy consumption.

Although it is difficult to make definitive conclusions about the energy policy of Latin American countries based on the empirical results presented in this paper, these findings serve to explain certain tools that can be used in conjunction with other studies in the decision-making process.

Future research could include variables such as physical capital, human capital and labor to estimate the long-run elasticities, following the methods of Mankiw, Romer and Weil (1992). This procedure would account for the fact that these factors of production are just as important as energy consumption. In addition, future research could extend the analysis to short-term relationships with a VEC model, as this model provides evidence that the series are cointegrated. Additionally, future research could evaluate energy efficiency policies not on the basis of energy conservation measures, but rather from the perspective of the efficiency of energy use in production processes.

REFERENCES

Akarca, A.T. and T.V. Long (1980), "On the relationship between energy and GNP: A reexamination," *Journal of Energy Development* **5**: 326–31.

Abosedra, S. and H. Baghestani (1991), "New evidence on the causal relationship between U.S. energy consumption and gross national product," *Journal of Energy and Development* **14**: 285–92.

Al-Iriani, M.A. (2006), "Energy GDP relationship revisited: An example from GCC countries using panel causality," *Energy Policy* **34**: 3342–50.

Breitung, J. (2000), "The local power of some unit root tests for panel data," *Advances in Econometrics* **15**: 161–77.

Bai, J. and P. Perron (2003), "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics* **18**: 1–22.

Cheng, B.S. (1995), "An investigation of cointegration and causality between energy consumption and economic growth," *Journal of Energy Development* **21**: 73–84.

Choi, I. (2001), "Unit root test for panel data," *Journal of International Money and Finance* **20**: 249–72.

Davison, A. and D. Hinckley (1997), *Bootstrap methods and their application.* Cambridge: Cambridge University Press.

Dichey, D. and W. Fuller (1979), "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association* **74**: 427–31.

Engle, R.F. and C. Granger (1987), "Co-integration and error-correction: Representation, estimation and testing," *Econometrica* **55**(2): 251–76.

Erol, U. and E.S.H. Yu (1987), "On the causal relationship between energy and income for industrialized countries," *Journal of Energy Development* **13**: 113–22.

Fatai, K., L. Oxley, and F. Scrimgeour (2002), "Energy consumption and employment in New Zealand: Searching for causality," paper presented at NZAE Conference, Wellington, June 26-28, 2002.

Ghosh, S. (2002), "Electricity consumption and economic growth in India," *Energy Policy* **30**: 125–29.

Glasure, Y. and A. Lee (1997), "Cointegration, error correction and relationship between GDP and energy: The case of South Korea and Singapore," *Resource and Energy Economics* **20**: 17–25.

Granger, C. and P. Newbold (1974), "Spurious regressions in econometrics," *Journal of Econometrics* **2**: 111–20.

Granger, C. (1969), "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica* **37**: 424–38.

Hadri, K. (2000), "Testing for stationarity in heterogeneous panel data," *Econometric Journal* **3**: 148–61.

Hwang, D. and B. Gum (1992), "The causal relationship between energy and GNP: The case of Taiwan," *Journal of Energy and Development* **16**(2): 219–26.

Im, K., M. Pesaran, and Y. Shin (2003), "Testing for unit roots in heterogeneous panels," *Journal of Econometrics* **115**: 53–74.

Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin (1992), "Testing the null hypothesis of stationarity against the alternative of unit root," *Journal of Econometrics* **54**(1-3): 159–78.

Kraft, J. and A. Kraft (1978), "On the relationship between energy and GNP," *Journal of Energy and Development* **3**(2): 401–03.

Lee, C. (2005), "Energy consumption and GDP in developing countries: A cointegrated panel analysis," *Energy Economics* **27**: 415–27.

Lee, C. and C. Chang (2007), "Energy consumption and GDP revisited: A panel analysis of developed and developing countries," *Energy Economics* **29**: 1206–23.

Lee, C. and C. Chang (2008), "Energy consumption and economic growth in Asian economies: A more comprehensive analysis using panel data," *Resource and Energy Economics* **30**: 50–65.

Lorde, T. and K. Guy (2009), "Energy consumption and economic growth in Latin America and the Caribbean: A panel cointegration approach," Central Bank of Barbados Research Department.

Johansen, S. (1988), "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control* **12**(2-3): 231–54.

Johansen, S. (1991), "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," *Econometrica* **59**(6): 1551–80.

Maddala, G. and S. Wu (1999), "A comparative study of unit root test with panel data and a new simple test," *Oxford Bulletin of Economics and Statistics* **61**: 631–52.

Magazzino, C. (2011), "Energy consumption and aggregate income in Italy: Cointegration and causality analysis," *Munich Personal RePEc Archive (MPRA)*, No. 28494.

Masih, A., and R. Masih (1996), "Energy consumption and real income temporal causality: Results from a multi-country study based on cointegration and error-correction modeling techniques", *Energy Economics* **18**: 165–83.

Pedroni, P. (1999), "Critical values for cointegration tests in heterogeneous panels with multiple regressors," *Oxford Bulletin of Economics and Statistics*, Special Issue 0305-9049.

Pedroni, P. (2004), "Panel cointegration: Asymptotic and finite sample propierties of pooled time series with an application to the PPP hypothesis: New results," *Econometric Theory* **20**: 597–627.

Squalli, J. (2007), "Electricity consumption and economic growth: Bounds and causality analyses of OPEC members," *Energy Economics* **29**: 1192–1205.

Soytas, U., R. Sari, and O. Ozdemir (2001), "Energy consumption and GDP relations in Turkey: A cointegration and vector error correction analysis," *Economics and business in transition: facilitating competitiveness and change in the global environment proceedings*, Global Business and Technology Association: 838–44 .

Yu, E.S.H. and B.K. Hwang (1984), "The relationship between energy and GNP: Further results," *Energy Economics* **6**: 186–90.

Yu, E.S.H. and J.Y. Choi (1985), "The causal relationship between energy and GNP: An international comparision," *Journal of Energy and Development* **10**: 249–72.

Westerlund, J. (2006), "Testing for panel cointegration with multiple structural breaks," *Oxford Bulletin of Economics and Statistics* **68**: 10132.

LATIN AMERICAN JOURNAL OF ECONOMICS

Vol. 50 No. 2 (Nov, 2013), 257–287

Income and Wealth Distribution with Physical and Human Capital Accumulation: Extending the Uzawa-Lucas Model to a Heterogeneous Households Economy*

Wei-Bin Zhang**

This paper proposes a dynamic economic model with wealth accumulation and human capital accumulation with endogenous education. In addition to learning by education like in the Uzawa-Lucas model, we also consider Arrow's learning by producing and Zhang's learning by consuming (creative learning) in the human capital accumulation equation. We simulate the model to demonstrate the existence of equilibrium points and motion of the dynamic system. We also examine how effects of changes in the propensity to receive education, the population, the propensity to save, and the education sector's total productivity will alter the paths of the economic dynamics.

JEL classification: O41, I2

Keywords: learning by producing, learning by consuming, learning by education, wealth and income distribution, heterogeneous households

1. Introduction

Dynamic interdependence between economic growth and human capital is currently a central topic in economic theory and empirical research. It has become evident that capital accumulation is not sufficient for explaining why countries grow differently, as posited by neoclassical growth theory. As Easterlin (1981) observed, in 1850 there were few people outside northwestern Europe and North America who had any formal education, and the spread of formal schooling seems to have preceded the beginning of modern economic growth. In modern economies, human capital is a key determinant of economic growth[1] and there have been many studies on the dynamic interdependence

* The author is grateful to the anonymous referee and editor Felipe Zurita for their constructive comments and suggestions, and acknowledges financial support from the Grants-in-Aid for Scientific Research (C), Project No. 25380246, Japan Society for the Promotion of Science.

** Professor, Ritsumeikan Asia Pacific University, 1-1 Jumonjibaru, Beppu-Shi, Oita-ken, 874-8577, Japan. Tel: 0977-78-1010; email: wbz1@apu.ac.jp. 1. For instance, see Hanushek and Kimko (2000), Barro (2001), Krueger and Lindahl (2001), and Castelló-Climent and Hidalgo-Cabrillana (2012).

doi 10.7764/LAJE.50.2.257

between education and economic growth in literature of economic theory and empirical economic studies.

Estimating the impact of education on earnings has been the focus of numerous empirical studies since Mincer (1974) published his seminal work. He finds that for white males not working on farms, an additional year of education raises income by about 7%. Other studies (Tilak, 1989) have shown that the spread of education can substantially reduce inequality within countries. Could *et al.* (2001) build a model to provide insights into the evolution of wage inequality within and between industries and education groups in the recent decades. The model shows that increasing randomness is the primary source of inequality growth among uneducated workers, but inequality growth among educated workers is determined more by changes in composition and return to ability (which is closely related to education). Tselios (2008) studies the relationship between income and educational inequalities in the regions of the European Union, using the European Community Household Panel data survey for 94 regions over the period 1995-2000. The research findings suggest a positive relationship between income and educational inequalities. Fleisher *et al.* (2011) examine the role of education in worker productivity and firms' total factor productivity on the basis of firm-level data from China. The study shows that an additional year of schooling raises marginal product by 30.1 percent, and the CEO's education increases TFP for foreign-invested firms.

The return is also closely related to ownership. For instance, the effect of schooling on productivity is highest in foreign-invested firms. One significant conclusion is that market mechanisms contribute to more efficient use of human capital within firms. Zhu (2011) studies individual heterogeneity in returns to education in China from 1995-2002, finding heterogeneous effects both within and between gender groups. In Zhu's study, heterogeneity in schooling returns falls from 1995 to 2002 for both genders in urban China, although their rates of education return have increased substantially. One reason for the narrowing heterogeneity is a better-functioning and increasingly integrated urban labor market in China.

The literature on endogenous knowledge and economic growth has expanded since Romer (1986) re-examined issues of endogenous technological change and economic growth in his 1986 paper.[2] But it

2. See Lucas (1988), Grossman and Helpman (1991), and Aghion and Howitt (1998).

is the work of Lucas (1988) that has created substantial interest in formal modeling of education and economic growth among economists. The first formal dynamic growth model with education was actually proposed by Uzawa (1965). In the Uzawa-Lucas model and many of its extensions and generalizations, it is implicitly assumed that all skills and human capital are developed by formal schooling. However, much of human capital may be accumulated through family and other social and economic activities. For instance, the human capital of a graduate student from a wealthy family in the U.S. may be quite different from the human capital of a graduate student from a middle-class family in India. By ignoring non-school factors, we may misunderstand the role of formal education in economic development. Another issue is described by Chen and Chevalier (2008): "Making and exploiting an investment in human capital requires individuals to sacrifice not only consumption, but also leisure. When estimating the returns to education, existing studies typically weigh the monetary costs of schooling (tuition and forgone wages) against increased wages, neglecting the associated labor/leisure tradeoff." The purpose of this study is to introduce other sources of learning into the Uzawa-Lucas two-sector growth model.

Another key purpose of this study is to introduce heterogeneous households into the two-sector growth model with education. Different households have different propensities to save and obtain education, as well as different abilities to absorb knowledge and increase human capital through education, learning by doing and learning by consuming. Most of the extensions and generalizations of the Uzawa-Lucas model are limited to a single representative household. However, there are models of endogenous human capital with heterogeneous households. For instance, Galor and Zeira (1993) propose a model to study the relationship between growth and inequality with human capital as the driving force of economic growth. A main conclusion of their study is that in the presence of credit constraints on human capital investments, high initial inequality may reduce long-run growth, while redistribution may increase the growth rate. Maoz and Moav (1999) build a similar model and show that the impact of income redistribution is situation-dependent. In another model by Galor and Moav (2004), it is demonstrated that in the early stage of modern economic development, high inequality encourages growth as the rich have a higher propensity to save, whereas at later stages high inequality may discourage growth as human capital becomes increasingly important and high inequality may be an impediment to human capital accumulation.

Fender and Wang (2003) build an overlapping-generations model with endogenous education choice with and without credit constraints. In their model, credit constraints are associated with lower education and a lower rate of interest. Laitner (2000) examines the dynamics of earnings within education groups and overall productivity using a model with endogenous human capital and a distribution of natural abilities. In a model of education where the distribution of abilities is the source of heterogeneity, Cardak (2004) shows that private education results in higher incomes and less income inequality than the public education model. Erosa *et al.* (2010) build a model of endogenous human capital accumulation with education to explain the variation in per-capita income across countries. Further literature can be found in the studies cited.

A main deviation of our approach from the previous models is that we derive demand for education in an alternative to the typical Ramsey approach. This allows us to explicitly derive the differential equations of the economic system and simulate transition processes. Our model is built on the three main growth models–Solow's one-sector growth model, Arrow's learning-by-doing model, and the Uzawa-Lucas growth model with education–in the growth literature. The main mechanisms of economic growth in these three models are integrated into a single framework with heterogeneous households. Our model is also based on the growth model with heterogeneous groups by Zhang (2009). In Zhang's model, human capital is fixed. The synthesis of the three growth models within a single framework is still analytically tractable because we propose an alternative approach to consumer behavior.

The paper is organized as follows. Section 2 introduces the basic model of wealth accumulation and human capital accumulation. Section 3 examines the dynamic properties of the model and simulates the model with three types of households. Section 4 contains our comparative dynamic analysis with regard to certain parameters and Section 5 contains concluding remarks.

2. The basic model

In our model, the economy has one production sector and one education sector. Most aspects of the production sector are similar to the standard one-sector growth model.[3] It is assumed that there is only one (durable) good in the economy under consideration. Households own assets in the economy and distribute their incomes to consume and save. Firms use inputs such as labor with varied levels of human capital, different kinds of capital, knowledge and natural resources to produce material goods or services. Exchanges take place in perfectly competitive markets. Factor markets work well; factors are inelastically supplied and the available factors are fully utilized at every moment. Saving is undertaken only by households. All firm earnings are distributed in the form of payments to factors of production, labor, managerial skill and capital ownership. Each group has a fixed population, $\bar{N}_j$, $(j = 1,\ldots,J)$. Let prices be measured in terms of the commodity and the price of the commodity be unitary. We denote wage and interest rates by $w_j(t)$ and $r(t)$ respectively. We use $H_j(t)$ to stand for group j's level of human capital.

The total capital stock $K(t)$ is allocated between the two sectors. We use $N_i(t)$ and $K_i(t)$ to stand for the labor force and capital stocks employed by the production sector, and $N_e(t)$ and $K_e(t)$ for the labor force and capital stocks employed by the education sector. We use $T_j(t)$ and $T_{je}(t)$ to stand for, respectively, the work time and study time of a typical worker in group j. As full employment of labor and capital is assumed, we have:

$$K_i(t) + K_e(t) = K(t), \quad N_i(t) + N_e(t) = N(t) \tag{1}$$

Where $N(t)$ is the total qualified labor supply defined by:

$$N(t) = \sum_{j=1}^{J} T_j(t) H_j^{m_j}(t) \bar{N}_j . \tag{2}$$

We rewrite (1) as follows:

$$n_i(t)k_i(t) + n_e(t)k_e(t) = k(t), \quad n_i(t) + n_e(t) = 1 \tag{3}$$

in which:

$$k_j(t) \equiv \frac{K_j(t)}{N_j(t)}, \quad n_j(t) \equiv \frac{N_j(t)}{N(t)}, \quad k(t) \equiv \frac{K(t)}{N(t)}, \quad j = i, e .$$

3. See Burmeister and Dobell (1970), Azariadis (1993), and Barro and Sala-i-Martin (1995).

2.1. The production sector

We assume that production is to combine the labor force $N_i(t)$ and physical capital $K_i(t)$. We use the conventional production function to describe a relationship between inputs and output. The function $F_i(t)$ is specified as:

$$F_i(t) = A_i K_i^{\alpha_i}(t) N_i^{\beta_i}(t), \quad A_i, \alpha_i, \beta_i > 0, \quad \alpha_i + \beta_i = 1.$$

Markets are competitive, so labor and capital earn their marginal products and firms earn zero profit. The rate of interest and wage rate are determined by markets. For any individual firm, $r(t)$ and $w_j(t)$ are given at each point in time. The production sector chooses the two variables $K_i(t)$ and $N_i(t)$ to maximize its profit. The marginal conditions are given by:

$$r(t) + \delta_k = \alpha_i A_i k_i^{-\beta_i}(t), \quad w_j(t) = \beta_i A_i H_j^{m_j}(t) k_i^{\alpha_i}(t), \tag{4}$$

where δ_k is the depreciation rate of physical capital.

2.2. Education sector

We assume that the education sector is also characterized by perfect competition. Here, we do not consider any government financial support for education. Students pay the education fee $p(t)$ per unit of time. The education sector pays teachers and capital at market rates. The total educational service is measured by the total education time received by the population. We specify the production function of the education sector as follows:

$$F_e(t) = A_e K_e^{\alpha_e}(t) N_e^{\beta_e}(t), \quad \alpha_e, \beta_e > 0, \quad \alpha_e + \beta_e = 1, \tag{5}$$

Where A_e, a_e and β_e are positive parameters. The marginal conditions for the education sector are:

$$r + \delta_k = \frac{\alpha_e p F_e}{K_e} = \alpha_e A_e p k_e^{-\beta_e}, \quad w_j = \beta_e A_e p H_j^{m_j} k_e^{\alpha_e}. \tag{6}$$

We see that the demand for labor from the education sector increases with the price and level of human capital and decreases with the wage rate.

2.3. Consumer behavior and wealth dynamics

Consumers make decisions about levels of consumption of education, services and commodities as well as how much to save, and family plays a particularly important role in decisions about investment in education. There are different models for decisions about education in a family. For instance, in Becker (1981), parents and children share a single unified utility function. In Cox (1987), the family is a nexus for transactions–the old lend to their children who repay them with care during old age. There is also a range between these two formulations in which parents place varying amounts of weight on the income, consumption or human capital of their children.[4] In this study, we follow Zhang (2007) in modeling choice of education time. The preference over current and future consumption is reflected in the consumer's preference structure over education, consumption and saving. Let $\bar{k}_j(t)$ stand for the per-capita wealth of group j. We have $\bar{k}_j(t) = \bar{K}_j(t)/\bar{N}_j$. Per-capita current income from the interest payment $r(t)\bar{k}_j(t)$ and the wage payment $T_j(t)w_j(t)$ is given by:

$$y_j(t) = r(t)\bar{k}_j(t) + T_j(t)w_j(t).$$

We call $y_j(t)$ the current income in the sense that it comes from consumers' payment for human capital and effort and consumers' current earnings from wealth ownership. The sum of money that consumers use for consumption, saving, and education are not necessarily equal to temporary income because consumers can sell wealth to pay, for instance, for current consumption if temporary income is not sufficient for buying food and travelling the country. The total value of wealth that consumers can sell to purchase goods and save is equal to $\bar{k}_j(t)$. Here, we assume that selling and buying wealth can be conducted instantaneously without any transaction cost. The per-capita disposable income is given by:

$$\hat{y}_j(t) = y_j(t) + \bar{k}_j(t) = (1 + r(t))\bar{k}_j(t) + T_j(t)w_j(t). \qquad (7)$$

4. See Behrman *et al.* (1982), Fernandez and Rogerson (1998), and Banerjee (2004).

Disposable income is used for saving, consumption, and education. It should be noted that the value $\bar{k}_j(t)$, (i.e., $p(t)\bar{k}_j(t)$ with $p(t)=1$), in the above equation is a flow variable. Under the assumption that wealth can be sold instantaneously without any transaction cost, we may consider $\bar{k}_j(t)$ as the amount of income that the consumer obtains at time t by selling all of his wealth. Hence, at time t the consumer has the total income amount $\hat{y}_j(t)$ to distribute among saving, consumption and education. In the growth literature, for instance in the Solow model, saving is proportional to current income, $y_j(t)$ while in this study saving is chosen by maximizing the utility subject to the budget constraint.

At each point in time, a consumer would distribute the total available budget among saving $s_j(t)$, consumption of goods $c_j(t)$ and education $p_j(t)\,T_{je}(t)$. The budget constraint is given by:

$$c_j(t) + s_j(t) + p_j(t)\,T_{je}(t) = \hat{y}_j(t) = (1 + r(t))\bar{k}_j(t) + T_j(t)\,w_j(t). \quad (8)$$

The consumer is faced with the following time constraint:

$$T_j(t) + T_{je}(t) = T_0,$$

Where T_0 is the total available time for work and study. Substituting this function into (8) yields:

$$c_j(t) + s_j(t) + (p_j(t) + w_j(t))\,T_{je}(t) = \bar{y}_j(t) \equiv (1 + r(t))\bar{k}_j(t) + T_0 w_j(t). \quad (9)$$

We now introduce a utility function for analyzing household behavior. Here, we consider that education has two kinds of returns. As education raises labor productivity, its effect is reflected in higher wages. As Lazear (1977: 570) describes: "...education is simply a normal consumption good and that, like all other normal goods, an increase in wealth will produce an increase in the amount of schooling purchased. Increased incomes are associated with higher schooling attainment as the simple result of an income effect." Education also results in direct pleasure, greater knowledge, higher social status and so on.[5] The relative importance of these returns may vary across different types of education with different individuals. This study introduces education time as a normal good into the utility function. In our model, at each point in time consumers have three variables to choose: level of consumption, level of saving, and education time.

5. See, for instance, Heckman (1976), Lazear (1977), and Malchow-Møller *et al.* (2011).

We assume that consumers' utility function is a function of level of goods $c_j(t)$, level of saving $s_j(t)$, and education level $T_{je}(t)$ as follows:

$$U(t) = c^{\xi_{j0}}(t) s^{\lambda_{j0}}(t) T_e^{\eta_{j0}}(t), \quad \xi_{j0}, \lambda_{j0}, \eta_{j0} > 0, \tag{10}$$

where ζ_{j0} is called the propensity to consume, λ_{j0} is the propensity to own wealth, and η_{j0} is the propensity to obtain education. This utility function is applied to different economic problems. A detailed explanation of the approach and its applications to different problems of economic dynamics are provided in Zhang (2005, 2009). As discussed by Zhang, this utility function overcomes the problem of the Solow model in which household behavior is modeled without micro-economic foundation. Another approach to household behavior in growth theory is the so-called Ramsey approach. In this approach, the household's preferences are expressed by an instantaneous utility function, $u(c(t))$ where $c(t)$ is the flow of consumption per person, and a discount rate for utility, denoted by ρ. Assume that each household maximizes utility U as given by:

$$U = \int_0^\infty u(c(t)) \; e^{-\rho t} \, dt, \quad c(t) \ge 0, \quad t \ge 0.$$

The household makes the decision subject to a lifetime budget constraint. This type of utility formulation means that the household's utility at time 0 is a weighted sum of all future flows of utility. The parameter ρ (≥ 0) is defined as the rate of time preference. A positive value of ρ means that utilities are valued less the later they are received.

There are two assumptions involved in the Ramsey model. The first is that utility is additional over time. Intuitively it is not reasonable to add happiness over time. It is well known in utility theory that when we use the utility function to describe consumer behavior, an arbitrary increasing transformation of the function would result in identical maximization of the consumer at each point in time. Obviously, the above formulation will not result in identical behavior if U is subjected to arbitrarily different increasing transformations at different times. The second implication of the above formation is that the parameter ρ is meaningless if utility is not additional over time.

It should be noted that Ramsey considered the meanings of this parameter from an ethical perspective. He interpreted the agent as a social planner, rather than a household. The planner chooses consumption and saving for current and future generations. Ramsey assumed $\rho = 0$ and considered $\rho > 0$ "ethically indefensible" (Ramsey, 1928). In fact, as shown in a recent survey on studies of estimating individuals' discount rates by Frederick *et al.* (2002), the rates differ dramatically across studies and within studies across individuals.[6] There is no convergence toward an agreed-on rate of impatience. As observed by Frederick *et al.* (2002), "The [discounted utility] model, which continues to be widely used by economists, has little empirical support. Even its developers–Samuelson, who originally proposed the model, and Koopmans, who provided the first axiomatic derivation–had concerns about its descriptive realism, and it was never empirically validated as the appropriate model for intertemporal choice. ... [D]eveloping descriptively adequate models of interremporal choice will not be easy." This study is based the alternative approach to interremporal choice proposed by Zhang, which does not use the concept of the discounted rate. It should be noted that another analytical advantage of Zhang's approach is that the dimension of resulted dynamics is lower than in Ramsey. As demonstrated later, this makes the analysis of behavior much easier.

For the representative consumer, wage rate $w_j(t)$ and rate of interest $r(t)$ are given in markets and wealth $\bar{k}_j(t)$ is predetermined before decision-making. Maximizing $U_j(t)$ subject to (9) yields:

$$c_j = \xi_j \, \bar{y}_j , \;\; s_j = \lambda_j \, \bar{y}_j , \;\; \left(p + w_j\right) T_{je} = \eta_j \, \bar{y}_j , \tag{11}$$

where:

$$\xi_{j0} \equiv \rho_j \, \xi_{j0} , \;\; \lambda_{j0} \equiv \rho_j \, \lambda_{j0} , \;\; \eta_{j0}$$

$$\equiv \rho_j \, \eta_{j0} , \;\; \rho_j \equiv \frac{1}{\xi_{j0} + \lambda_{j0} + \eta_{j0}} .$$

Demand for education is given by $T_{je} = \eta_j \bar{y}_j \, / \, (p_j + w_j)$ Demand for education decreases with the price of education and the wage rate and increases in $\bar{y}_j$. An increase in the propensity to obtain education increases education time when the other conditions are fixed. In this dynamic system, because any factor is related to all the

6. With regard to some limitations of the traditional Ramsey approach to household behavior, we refer to Attanasio and Weber (2010).

other factors over time, it is difficult to see how one factor affects any other variable over time. We will demonstrate complicated interactions by simulation.

According to the definitions of $s_j(t)$, the wealth accumulation of the representative household in group j is given by:

$$\dot{\bar{k}}_j(t) = s_j(t) - \bar{k}_j(t). \tag{12}$$

This equation simply states that the change in wealth is equal to savings minus dissavings.

2.4. Dynamics of human capital

As empirically tested by Aakvik *et al.* (2010), different forms of learning have different human capital accumulation effects. We assume that there are three sources of improvement of human capital: education, learning by producing, and learning by leisure. Arrow (1962) first introduced learning by doing into growth theory; Uzawa (1965) took into account trade-offs between investment in education and capital accumulation, and Zhang (2007) introduced the impact of consumption on human capital accumulation (via so-called creative leisure) into growth theory. In fact, Arrow (1962: 172) recognizes the necessity of extending his idea to include other sources of human capital accumulation: "It has been assumed here that learning takes place only as a by-product of ordinary production. In fact, society has created institutions, education and research, whose purpose it is to enable learning to take place more rapidly. A fuller model would take account of these as additional variables."

We propose that human capital dynamics is given by:

$$\dot{H}_j = \frac{\upsilon_{je} F_e^{a_{je}} \left(H_j^{m_j} T_{je} \bar{N}_j \right)^{b_{je}}}{H_j^{\pi_{je}} \bar{N}_j} + \frac{\upsilon_{ji} F_i^{a_{ji}}}{H_j^{\pi_{ji}} \bar{N}_j} + \frac{\upsilon_{jh} C_j^{a_{jh}}}{H_j^{\pi_h} \bar{N}_j} - \delta_{jh} H_j, \tag{13}$$

where δ_{jh} (>0) is the depreciation rate of human capital and υ_{je}, υ_{ji}, υ_{jh}, a_{je}, b_{je}, a_{ji}, and a_{jh} are non-negative parameters. The signs of the parameters π_{je}, π_{ji}, and π_{jh} are not specified as they may be either negative or positive.

The above equation is a synthesis and generalization of Arrow's, Uzawa's, and Zhang's ideas about human capital accumulation.

The term $v_{je}\, F_e^{a_{je}}(H_j^{m_j} T_{je} \bar{N}_j)^{b_{je}} / H_j^{\pi_{je}} \bar{N}_j$ describes the contribution to human capital improvement of education. Human capital tends to increase with an increase in the level of education service, F_e, and in the (qualified) total study time, $H_j^{m_j} T_{je} \bar{N}_j$. The population $\bar{N}_j$ in the denominator measures the contribution in per-capita terms. The term $H^{\pi_{je}}$ indicates that as the level of human capital of the population increases, it may be more difficult (in the case of π_{je} being large) or easier (in the case of π_{je} being small) to accumulate more human capital via formal education. The term $\bar{N}_j$ in the denominator term measures the contribution in per-capita terms.

We take into account the effects of learning by producing in human capital accumulation through the term $v_{ji}\, F_i^{a_{ji}} / H_j^{\pi_{ji}}$. This term implies that the contribution of the production sector to human capital improvement is positively related to its production scale F_{ji} and is dependent on the level of human capital. The term $H^{\pi_{ji}}$ takes account of returns-to-scale effects in human capital accumulation. The case of $\pi_{ji} > (<)\ 0$ implies that as human capital is increased it is more difficult (easier) to further improve the level of human capital. We account for learning by consuming through the term $v_{jh}\, C_j^{a_{jh}} / H_j^{\pi_{jh}} \bar{N}_j$. This term, introduced into the human capital accumulation equation by Zhang (2007) can be interpreted in a similar fashion as the term for learning by producing. In contemporary (in particular, developed) economies, human capital is evidently closely related to leisure activities such as club activities, traveling to different parts of the world, playing computer games, watching TV, and playing sports. Playing recreational games and using mobile phones to communicate with friends enable people to accumulate skills for operating computers. Traveling raises awareness of differences in, for instance, geography and cultures. People born in poor and rich countries obviously have different levels of human capital due to differences in living conditions. Neither Arrow's learning by doing nor Uzawa's formal education take into account this source of human capital accumulation.

It should be noted that in the literature on education and economic growth, it is assumed that human capital evolves according to the following equation (see Barro and Sala-i-Martin, 1995):

$$\dot{H}(t) = H^{\eta}(t) G(T_e(t)),$$

where the function G is increasing as the effort rises with $G(0) = 0$. In the case of $\eta < 1$ there is diminishing return to the human capital accumulation. This formation is due to Lucas (1988). As $\dot{H}/H < H^{\eta - 1}G(1)$, we conclude that the growth rate of human capital must eventually tend to zero no matter how much effort is devoted to accumulating human capital. Uzawa's model may be considered a special case of the Lucas model with $\gamma = 0$, $U(c) = c$ and the assumption that the right-hand side of the above equation is linear in the effort. It seems reasonable to consider diminishing returns in human capital accumulation: people accumulate it rapidly early in life, then less rapidly, then not at all–as though each additional percentage increment were harder to gain than the preceding one. Solow adapts the Uzawa formation to the following form:

$$\dot{H}(t) = H(t)\kappa T_e(t).$$

This is a special case of the above equation. The new formation implies that if no effort is devoted to human capital accumulation, then $\dot{H}(0) = 0$ (human capital does not vary as time passes; this results from depreciation of human capital being ignored); if all effort is devoted to human capital accumulation, then $g_H(t) = \kappa$ (human capital grows at its maximum rate; this results from the assumption of potentially unlimited growth of human capital). Between the two extremes, there is no diminishing return to the stock $H(t)$. Achieving a given percentage increase in $H(t)$ requires the same effort. As remarked by Solow (2000), the above formulation is very far from a plausible relationship. If we consider the above equation as a production function for new human capital (i.e., $H(t)$), and if the inputs consist of already accumulated human capital and study time, then this production function is homogenous of degree two. It has strong increasing returns to scale and constant returns to $H(t)$ itself. It can be seen that our approach is more general than the traditional formation with regard to education. Moreover, we also treat teaching as a significant factor in human capital accumulation. Efforts in teaching are neglected in the Uzawa-Lucas model.

For the education sector, the demand and supply balances at any point in time:

$$\sum_{j=1}^{J} T_{je}\, \bar{N}_j = F_{je}(t). \tag{14}$$

The total capital stocks employed by the world are equal to the wealth owned by the world. That is,

$$K(t) = K_i(t) + K_e(t) = \sum_{j=1}^{J} \bar{k}_j(t)\bar{N}_j. \tag{15}$$

World production is equal to world consumption and world net savings. That is,

$$C(t) + S(t) - K(t) + \delta_k K(t) = F_i(t), \tag{16}$$

where:

$$C(t) \equiv \sum_{j=1}^{J} c_j(t)\bar{N}_j, \quad S(t) \equiv \sum_{j=1}^{J} s_j(t)\bar{N}_j.$$

It is straightforward to show that this equation can be derived from the other equations in the system. We have thus built the dynamic model. We now examine the dynamics of the model.

3. The dynamics and their properties

Because the system has heterogeneous households, the dynamics are highly dimensional. The following lemma shows that the motion of the economy is expressed by $2J$ dimensional differential equations.

Lemma 1. The dynamics of the economy are governed by the following $2J$ dimensional differential equation system with $k_{1i}(t)$, $\{\bar{k}_j(t)\}$, $(H_j(t))$, as the variables, where $\{\bar{k}_j\} \equiv (\bar{k}_2,\ldots,\bar{k}_J)$ and $(H_j) \equiv (H_1,\ldots,H_J)$:

$$\dot{k}_{1i} = \bar{\Lambda}_1\left(k_{1i}, \left(H_j\right), \left\{\bar{k}_j\right\}\right),$$

$$\dot{\bar{k}}_j = \bar{\Lambda}_j\left(k_{1i}, \left(H_j\right), \left\{\bar{k}_j\right\}\right), \quad j = 2, ..., J,$$

$$\dot{H}_j = \Lambda_j\left(k_{1i}, \left(H_j\right), \left\{\bar{k}_j\right\}\right), \quad j = 1, ..., J,$$

in which $\bar{\Lambda}_j$ and Λ_j are unique functions of $k_{1i}(t)$, $\{\bar{k}_j(t)\}$ and $(H_j(t))$ at any point in time, as defined in the appendix. For any given positive values of $k_{1i}(t)$, $\{\bar{k}_j(t)\}$ and $(H_j(t))$ at any point in time, the other variables are uniquely determined by the following procedure: k_{ji} by (A1) $\rightarrow$ k_{je} by (A3) $\rightarrow$ r and w_j by (A2) $\rightarrow$ p by (A4) $\rightarrow$ $\bar{k}_1$ by (A12) K_j by (A12) $\rightarrow$ k_j by (A9) $\rightarrow$ T_j by (A10) $\rightarrow$ T_{je} by (A8) $\rightarrow$ $N_j = T_j\bar{N}_j \rightarrow$ n_{ji} and n_{je} by (A5) $\rightarrow$ $N_{ji} = n_{ji}N_j$ and $N_{je} = n_{je}N_j$ $\rightarrow$ $K_{ji} = k_{ji}N_{ji}$ and $K_{je} = k_{je}N_{je} \rightarrow F_{ji}$ by (2) $\rightarrow$ F_{je} by (12) $\rightarrow$ $\bar{y}_j$ by (7) $\rightarrow$ c_j and s_j by (9).

We have the dynamic equations for the economy with any number of household types. The system is nonlinear and of high dimension. It is difficult to generally analyze behavior of the system. To illustrate motion of the system, we specify the parameters as follows:

$$\begin{pmatrix} N_1 \\ N_2 \\ N_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 20 \\ 30 \end{pmatrix}, \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.4 \\ 0.3 \end{pmatrix}, \begin{pmatrix} \xi_{10} \\ \xi_{20} \\ \xi_{30} \end{pmatrix} = \begin{pmatrix} 0.12 \\ 0.18 \\ 0.2 \end{pmatrix}, \begin{pmatrix} \lambda_{10} \\ \lambda_{20} \\ \lambda_{30} \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.75 \\ 0.7 \end{pmatrix},$$

$$\begin{pmatrix} b_{1h} \\ b_{2h} \\ b_{3h} \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.35 \\ 0.4 \end{pmatrix}, \begin{pmatrix} \pi_{1e} \\ \pi_{2e} \\ \pi_{3e} \end{pmatrix} = \begin{pmatrix} -0.2 \\ -0.15 \\ -0.1 \end{pmatrix}, \begin{pmatrix} \pi_{1i} \\ \pi_{2i} \\ \pi_{3i} \end{pmatrix} = \begin{pmatrix} 0.7 \\ 0.75 \\ 0.8 \end{pmatrix}, \begin{pmatrix} \pi_{1h} \\ \pi_{2h} \\ \pi_{3h} \end{pmatrix} = \begin{pmatrix} 0.1 \\ 0.15 \\ 0.2 \end{pmatrix},$$

$$\begin{pmatrix} a_{1e} \\ a_{2e} \\ a_{3e} \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.4 \\ 0.45 \end{pmatrix}, \begin{pmatrix} b_{1e} \\ b_{2e} \\ b_{3e} \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.55 \\ 0.6 \end{pmatrix}, \begin{pmatrix} a_{1i} \\ a_{2i} \\ a_{3i} \end{pmatrix} = \begin{pmatrix} 0.4 \\ 0.45 \\ 0.5 \end{pmatrix}, \begin{pmatrix} a_{1h} \\ a_{2h} \\ a_{3h} \end{pmatrix} = \begin{pmatrix} 0.1 \\ 0.15 \\ 0.2 \end{pmatrix}, \quad (17)$$

$$\begin{pmatrix} \eta_{10} \\ \eta_{20} \\ \eta_{30} \end{pmatrix} = \begin{pmatrix} 0.015 \\ 0.010 \\ 0.008 \end{pmatrix}, \begin{pmatrix} v_{1e} \\ v_{2e} \\ v_{3e} \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.7 \\ 0.5 \end{pmatrix}, \begin{pmatrix} v_{1i} \\ v_{2i} \\ v_{3i} \end{pmatrix} = \begin{pmatrix} 2.5 \\ 2 \\ 1.7 \end{pmatrix}, \begin{pmatrix} v_{1h} \\ v_{2h} \\ v_{3h} \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.6 \\ 0.5 \end{pmatrix},$$

$$A_i = 0.9,\ A_e = 0.8,$$
$$\alpha_i = 0.32,\ \alpha_e = 0.37,\ T_0 = 1,\ \delta_k = 0.05,\ \delta_{1h} = 0.04,$$
$$\delta_{2h} = 0.05,\ \delta_{3h} = 0.06,$$

The populations of groups 1, 2 and 3 are respectively 10, 30 and 60. Group 3 has the largest population. The total productivities of the industrial sector and the education sector are respectively 0.9 and 0.8. The utilization efficiency parameters of groups 1, 2, and 3, m_j are respectively 0.5, 0.4 and 0.3. Group 1 utilizes human capital mostly effectively; group 2 is next and group 3 uses it least effectively. We call the three groups respectively rich, middle, and poor class (RC, MC, PC). We specify the values of the parameters, a_j, in the Cobb-Douglas production functions as approximately equal to 0.3.[7] The RC's learning by doing parameter, v_{1i}, is the highest. The returns to scale parameters in learning by doing, π_{ji}, are all positive, which implies that knowledge exhibits decreasing returns to scale in learning by doing. The depreciation rates of physical capital and knowledge

7. The value is often used in empirical studies, such as Abel and Bernanke (1998).

are specified around 0.05. The RC's propensity to save is 0.8 and the PC's propensity to save is 0.7 The value of the MC's propensity is between the other groups. The RC's propensity to obtain education is highest among the three classes; the PC has the lowest propensity to obtain education.

Under (17) we find that the system has a unique equilibrium. The equilibrium values are listed in (18). The national industrial output is 292.25 and the interest rate is about 3.9 percent. The human capital levels of the RC, MC and PC are respectively 22.62, 6.63 and 3.64. The wage rates of the three groups are respectively 22.62, 6.63 and 3.64. The RC has the highest human capital as well as the highest wage rate. The RC also spends much more time in education than the other two groups. According to Dur and Glazer (2008), rich people tend to attend college at a higher rate than poor people, as rich people obtain more benefits from the consumption content of education. The PC spends the longest amount of time working. The RC's consumption level and wealth are also highest.

$$\begin{aligned}
&r = 0.039,\ p = 1.026,\ k = 5.67,\ N = 189.09,\ K = 1071.22,\\
&F_i = 292.25,\ F_e = 4.09, N_i = 186.61,\ N_e = 2.48,\ K_i = 1053.74,\\
&K_e = 17.48,\ k_i = 5.65,\ k_e = 7.07,\ f_i = 1.56,\ f_e = 1.65,\\
&H_1 = 22.62,\ H_2 = 6.63,\ H_3 = 3.64,\ w_1 = 5.07,\ w_2 = 2.27,\\
&w_3 = 1.57,\ \bar{k}_1 = 38.96,\ \bar{k}_2 = 10.57,\ \bar{k}_3 = 6.07,\ T_{1e} = 0.12,\\
&T_{2e} = 0.04,\ T_{3e} = 0.03,\ c_1 = 5.84,\ c_2 = 2.54,\ c_3 = 1.74.
\end{aligned} \tag{18}$$

It is straightforward to calculate the six eigenvalues as follows:

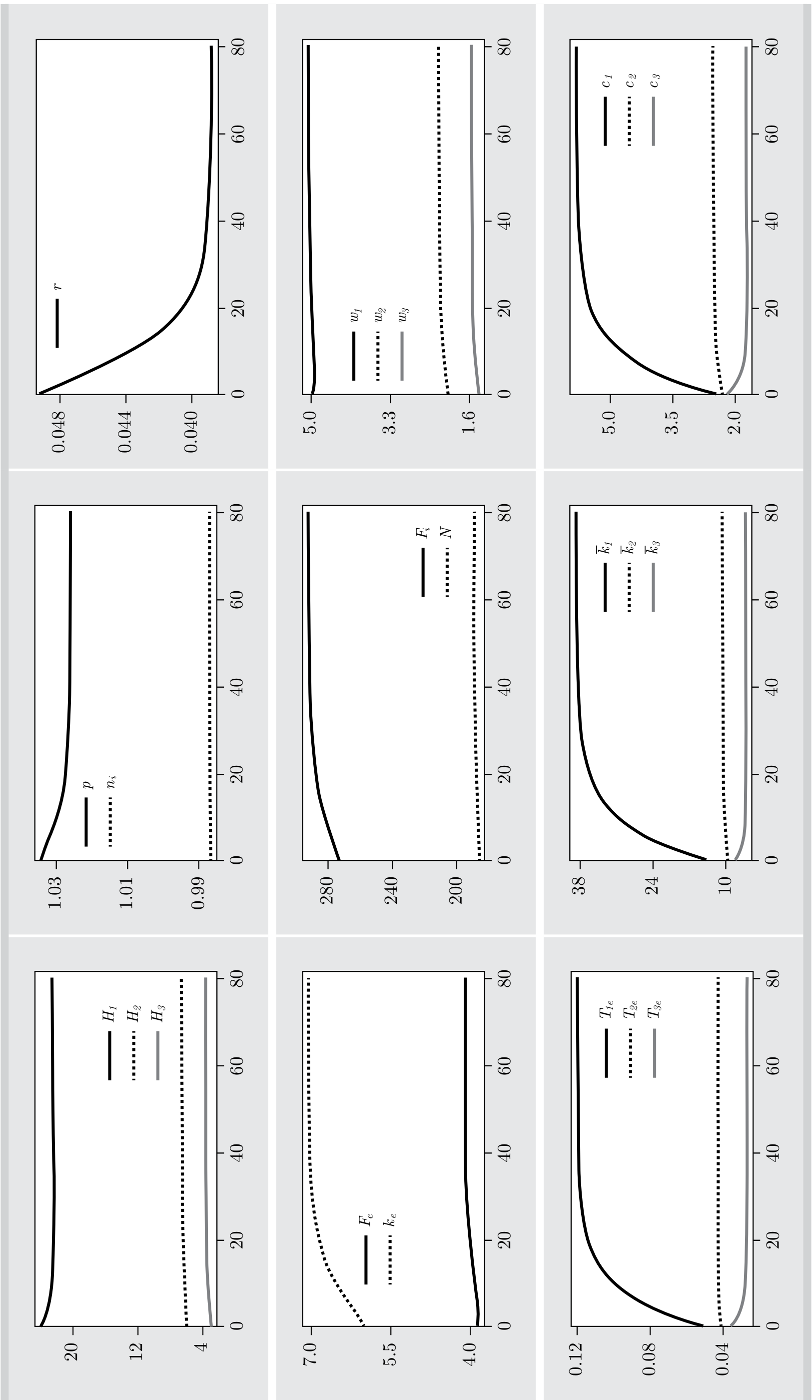
$$-0.20, -0.18, -0.13, -0.09, -0.06, -0.03.$$

As all the eigenvalues are negative, we see that the equilibrium is locally stable. We start with different initial states close to the equilibrium point and find that the system approaches the equilibrium point. In Figure 1, we plot the motion of the system with the following initial conditions:

$$\begin{aligned}
&\bar{k}_1(0) = 11,\ \bar{k}_2(0) = 4.5,\ \bar{k}_3(0) = 2,\ H_1(0) = 3.5,\ H_2(0) = 1.9,\\
&H_3(0) = 0.5.
\end{aligned} \tag{19}$$

The system approaches its equilibrium point in the long term.

Figure 1. Motion of some variables

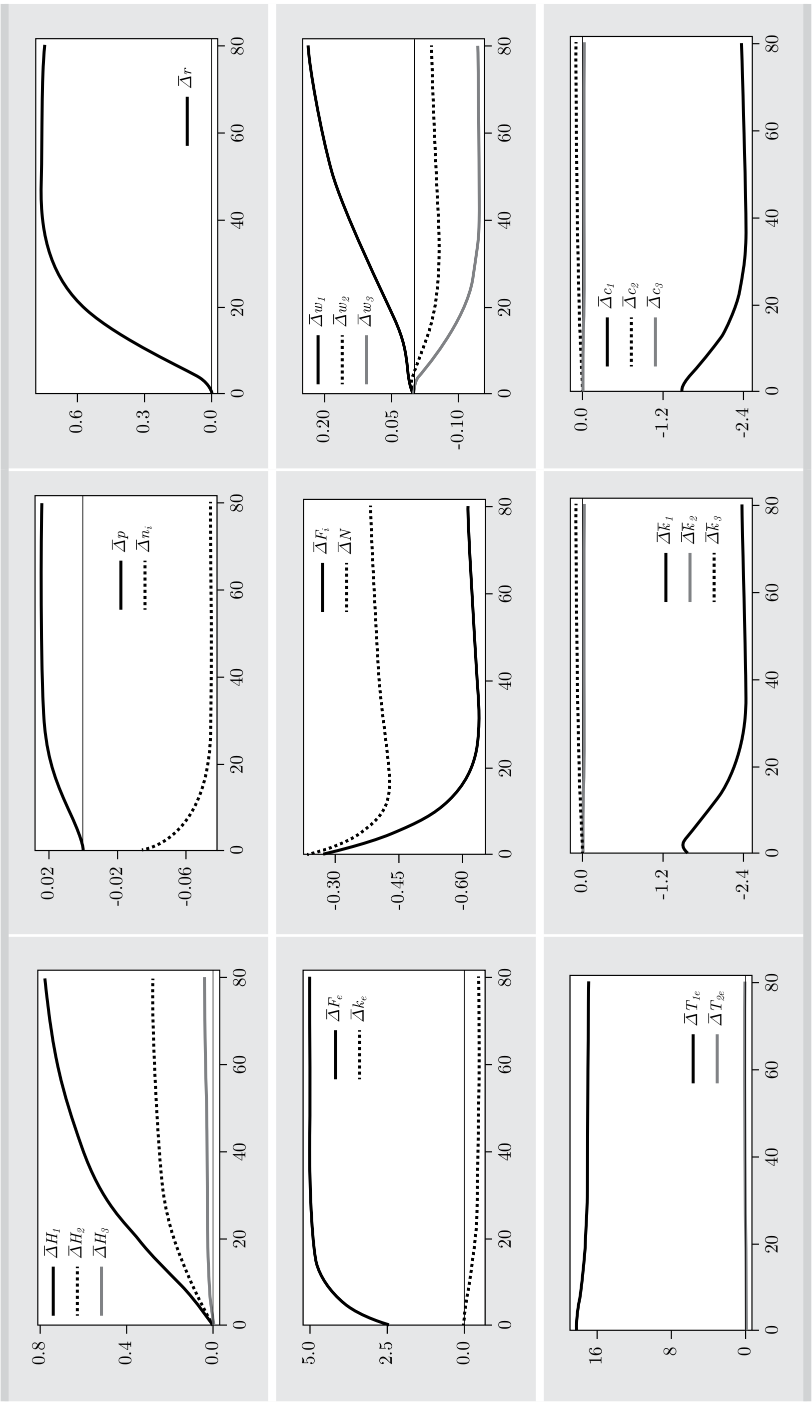


4. Comparative dynamic analysis

In simulating the motion of the dynamic system, it is important to ask questions such as how a change in one group's propensity to save or to obtain education affects the economy and each group's wealth and consumption. First, we examine the case that all the parameters, except the RC's propensity to obtain education, η_{10}, are the same as in (17). We increase the propensity in the following way: η_{10}: 0.015 $\Rightarrow$ 0.018. The simulation results are plotted in Figure 2. In the plots, a variable $\bar{\Delta} x_j(t)$ stands for the change rate of the variable, $x_j(t)$, as a percentage, due to changes in the parameter value. From Figure 2 we see that as the RC increases the propensity to obtain education, the RC's level of human capital is increased. To examine the process of the change in the human capital, we have to observe how all the variables in the system react to the parameter shift. Initially, as the rich class increases its preference for formal education, rich people would spend more time on schooling (with the other conditions remaining the same). Hence, the human capital of the rich class will be increased. The rise in human capital increases the total labor force, but the fall in work time reduces the labor force. As shown in Figure 2, the total labor force is reduced. Hence, the total output of the industrial sector falls. As the demand for education is increased, the education fee is increased and output of the education sector is increased. The labor share of the education sector is increased.

Although the RC increases its education time, the education time of the other two classes is slightly affected. The other two classes' human capital levels are increased, but the PC's human capital is much less affected. The wage rate of the RC increases, but the wage rates of the other two classes fall. As the education hours of the MC and PC are slightly affected, we can conclude that the returns from schooling in terms of wage rate are reduced for the two classes. As the RC puts more resources into schooling, wealth and consumption are reduced. It should be noted that in a model with education and inequality developed by Nakajima and Nakamura (2009), they find that the educational expenditure of rich households could prevent poor households from escaping poverty. Their model attempts to explain the possible effects of high prices of education on growth and inequality in countries such as Japan, Korea, and the U.S. The basic insight from the model is that as rich households demand higher

Figure 2. A rise in the rich class's propensity to obtain education

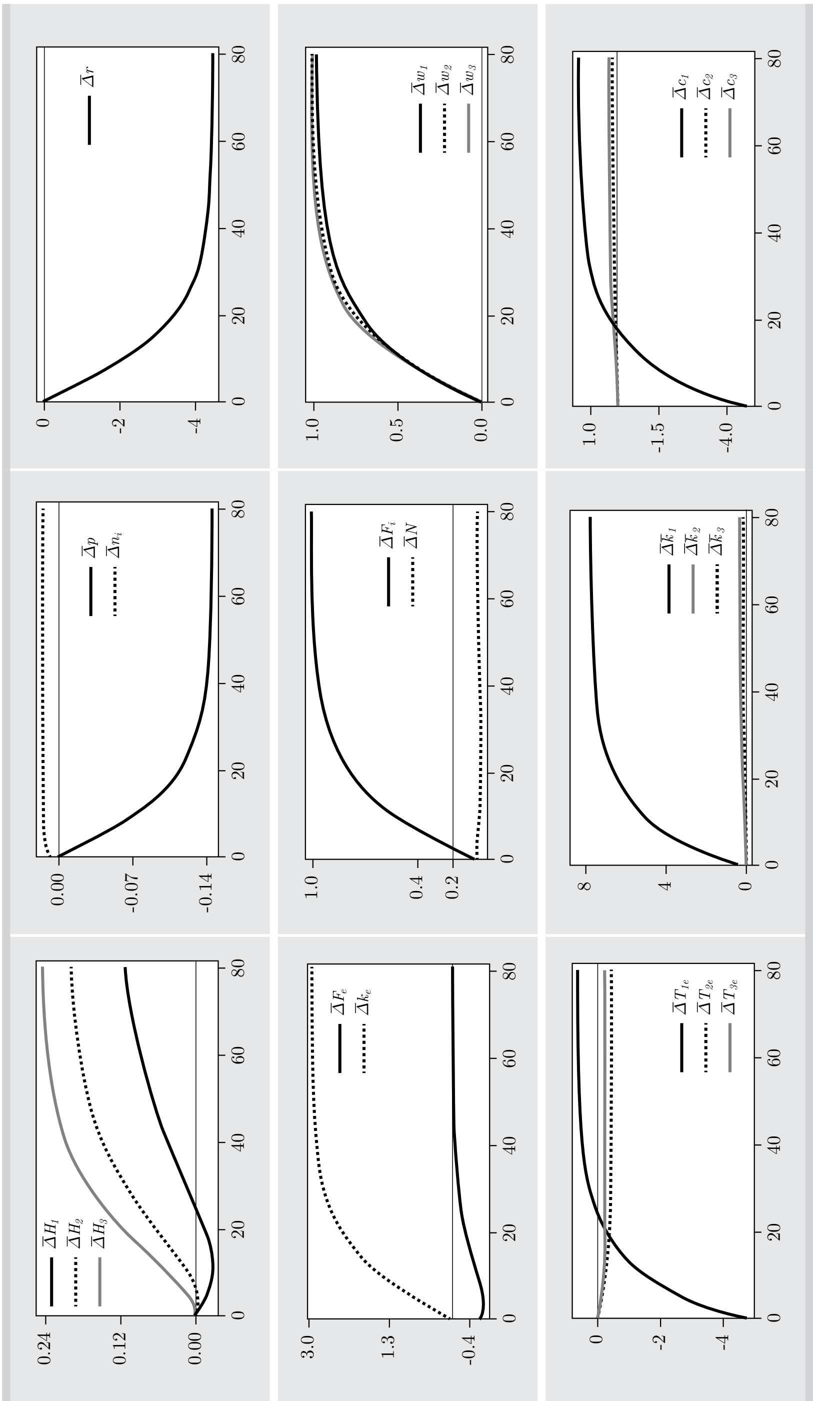


education, the price is expected to rise, excluding the poor from higher education. This also leads to greater inequality between the rich and poor in the long term. Our model predicts similar effects in a country with heterogeneous households.

We now increase the RC's propensity to save in the following way: λ_{10}: $0.8 \Rightarrow 0.85$. The simulation results are plotted in Figure 3. As the RC increases the propensity to save, its wealth per capita rises. The society's total capital is increased. As more capital is accumulated in the society, the interest rate falls. The increased capital results in higher wage rates. As wage rates rise, the opportunity costs of education are increased, which initially results in the reduction of education time for the three classes. As the RC accumulates more wealth and the RC's wage rate is increased, we see that the RC increases its education time in the long term. As the RC reduces education time and puts away more money for future consumption, its level of human capital falls initially. But in the long term the RC's human capital is increased as the RC devotes more time to education, becomes more effective through learning by consuming (due to a higher consumption) and learning by producing (due to increased output level of the industrial sector).

The other two classes' consumption levels are increased but only slightly. The RC's wealth is increased significantly, but the other two classes' wealth is only slightly affected. The total labor input is slightly affected. Because the RC cares more about wealth accumulation, we see that the share of labor force in the education sector is reduced and that of the industrial sector is increased. It is interesting to note that the human capital levels of the MC and the PC are increased more than the RC and the education fee is also reduced; the MC and PC don't experience increases in wealth and consumption. We see that a rise in the RC's propensity to save tends to enlarge the gaps with the other two classes in terms of wealth and consumption levels, but tends to reduce gaps for human capital among the classes. It should be noted that in this study, we fix propensities. In reality, as extremely rich people have "too much" money to spend, the propensity to save tends to increase in the long term. It is also worth noting that the propensity to save in our model is different from that defined in neoclassical growth theory. In our model, the propensity to save is equal to the share of disposable income saved for the future, while in the Solow model the propensity to save is equal to the share of current income saved

Figure 3. A rise in the rich class's propensity to save



for the future. Hence, in the Solow model the extent to which people accumulate wealth has a much weaker impact on saving behavior than in our model when the wealth per capita is extremely high. It can be seen that our model predicts enlarged income gaps between the rich and the poor over time, even though the human capital levels may not vary much between the two groups.

It has been observed that the effect of population growth varies with the level of economic development and can be positive for some developed economies. Theoretical models of human capital predict situation-dependent interactions between population and economic growth.[8] We now increase the RC's population in the following way: $\bar{N}_1$: $10 \Rightarrow 15$. The simulation results are plotted in Figure 4. As the RC population increases, its human capital and wage rate falls over time. The schooling time, consumption level and wealth per capita of the RC rise initially and fall in the long term. The price of education falls, which benefits the MC and PC. As the output of the industrial sector is increased as a consequence of the population increase, the MC and PC learn more effectively through learning by producing. The human capital levels of both the MC and the PC are increased. In the long term, the per-capita consumption and wealth levels of the MC and the PC are increased. We see that the MC and PC benefit in the long term. We thus conclude that an increase in the RC's population reduces gaps in income, wealth and consumption between the rich and the poor in the long term.

Another important question is what will happen to different people and the national economy if the total productivity of the education sector is increased. We increase the total productivity in the following way: A_e: $0.8 \Rightarrow 0.9$. The simulation results are plotted in Figure 5. The rise in productivity increases the human capital of all groups and reduces the price of education. As all the classes' human capital levels are increased and the differences in these increases are not large, the distribution of the total labor force is slightly affected. The two sectors increase their output levels in the long term. The education time, wage rates, wealth and consumption levels of all the groups are increased in the long term. We thus conclude that an improvement in the efficiency of the educational system will benefit different people in the long term.

8. See Ehlich and Lui (1997), Galor and Weil (1999), and Boucekkine et al. (2002).

Figure 4. A rise in the rich class's population

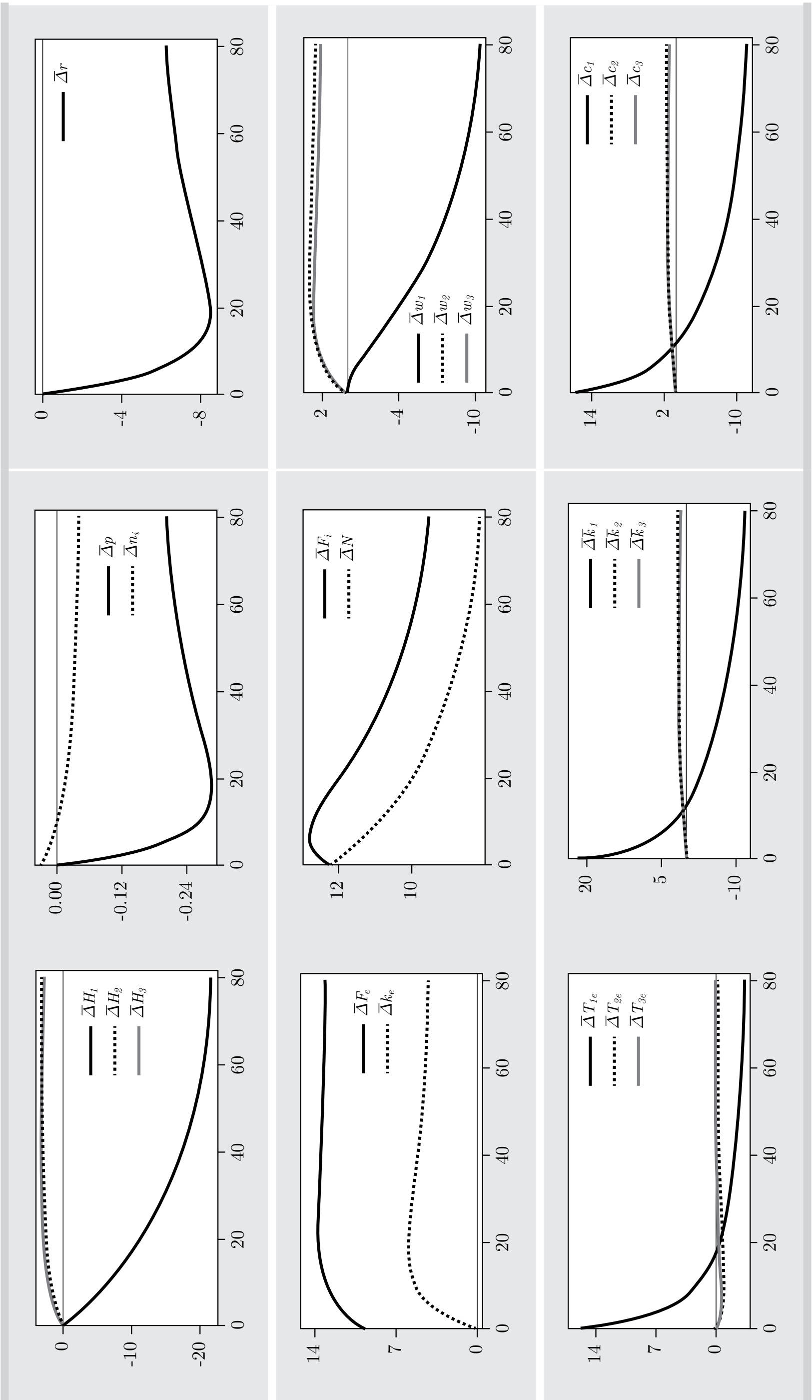
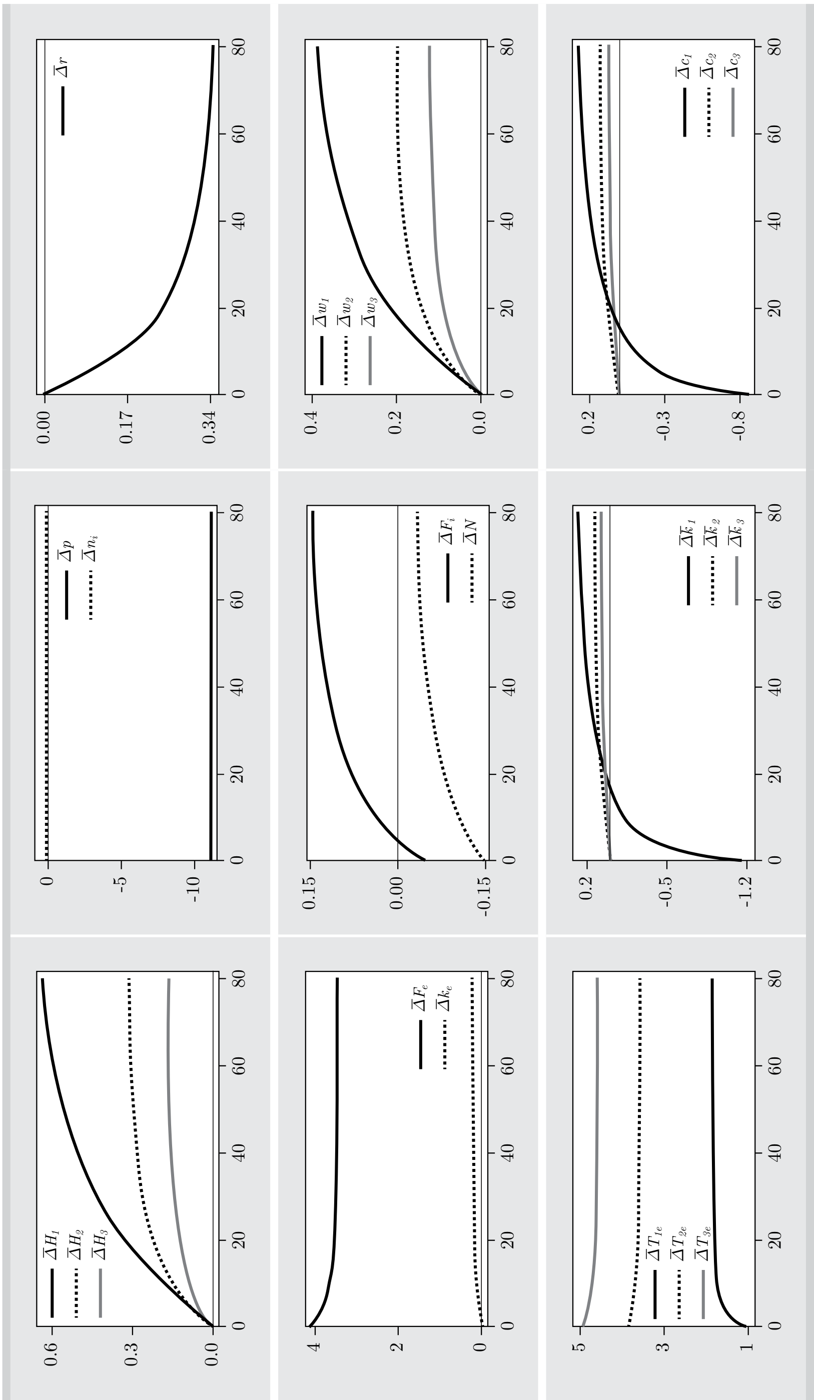


Figure 5. A rise in the education sector's total productivity



5. Concluding remarks

This paper proposes a growth model of heterogeneous households with wealth accumulation and human capital accumulation. The economic system consists of one production sector and one education sector. We consider three ways of improving human capital: learning by producing, learning by education, and learning by consuming. The model describes a dynamic interdependence among wealth accumulation, human capital accumulation, and division of labor under perfect competition. We simulated the model of three groups to demonstrate the existence of equilibrium points and motion in the dynamic system. We also examined the effects of changes in some parameters on the motion of the system.

The model could be extended in several directions. For instance, as mentioned by Bertocchi and Spagat (2004), the specific structure of the educational system is largely unexplored. Education may be privately or publicly provided. An education system may also be founded on a hierarchical differentiation between vocational and general education. As demonstrated in Bertocchi and Spagat, the educational structure varies over time. We could introduce some kind of government intervention in education into the model. In this study, we don't consider public provision or subsidy of education. In the literature on education and economic growth, many models with heterogeneous households are proposed to address issues related to taxation, education policy, distribution of income and wealth, and economic growth. For instance, Bénabou (2002) studies the effects of progressive income taxes and education finance in a dynamic heterogeneous-agent model. As shown in Glomm and Kaganovich (2008), public provision may have a significant impact on growth and inequality across households. They examine the relationship between growth and inequality with public education in the context of an overlapping generations economy with heterogeneous agents. The government collects a tax on labor income to finance education. The model predicts that government spending on education reduces income inequality. Another study by Dur and Glazer (2008) shows that rich people tend to attend college at a higher rate than poor people, as rich people obtain more benefits from the consumption content of education. They conclude that to make sure that colleges attract the most competent students and not simply the richest, colleges should charge rich students higher tuition fees.

APPENDIX

Proof of lemma 1

From (4) and (6), we obtain:

$$\frac{K_e}{N_e} = \alpha \frac{K_i}{N_i}, \quad \text{i.e.,} \quad k_e = \alpha k_i, \tag{A1}$$

where $\alpha \equiv \alpha_e \beta_i / \alpha_i \beta_e$ ($\neq 1$ assumed). From (A1), (4) and (6), we obtain:

$$p = \frac{\alpha^{\beta_e} \alpha_i A_i}{\alpha_e A_e} k_i^{\beta}, \tag{A2}$$

where $\beta \equiv \beta_e \beta_i - \beta_i$. From (A1) and (1), we solve the labor distribution as functions of k_i and k:

$$n_i = \frac{\alpha k_i - k}{(\alpha - 1) k_i}, \quad n_e = \frac{k - k_i}{(\alpha - 1) k_i}. \tag{A3}$$

From $(p + w_j) T_{je} = \eta_j \overline{y}_j$ in (11) and the definition of $\overline{y}_j$ we have:

$$T_{je} = \phi_{jp} \overline{k}_j + \phi_{j0}, \tag{A4}$$

where:

$$\phi_{jp}(k_i, H_j) \equiv \frac{(1 + r)\eta_j}{p + w_j}, \quad \phi_{j0}(k_i, H_j) \equiv \frac{\eta_j T_0 w_j}{p + w_j}.$$

From (A4) and (14):

$$\phi \sum_{j=1}^{J} (\phi_{jp} \overline{k}_j + \phi_{j0}) \overline{N}_j = N(k - k_i), \tag{A5}$$

where we also use $F_e = A_e k_e^{\alpha_e} n_e N$, (A3), $k_e = \alpha k_i$, and:

$$\phi(k_i) \equiv \frac{(\alpha - 1) k_i^{\beta_e}}{\alpha^{\alpha_e} A_e}.$$

From $T_j + T_{je} = T_0$ and (A4), we have:

$$T_j = T_0 - \phi_{jp} \bar{k}_j - \phi_{j0} . \tag{A6}$$

From (A6), we have:

$$N = \sum_{j=1}^{J} \left(T_0 - \phi_{jp} \bar{k}_j - \phi_{j0} \right) H_j^{m_j} \bar{N}_j . \tag{A7}$$

From (2) and $K = kN$, we have:

$$K = k \left(\sum_{j=1}^{J} \left(T_0 - \phi_{jp} \bar{k}_j - \phi_{j0} \right) H_j^{m_j} \bar{N}_j \right). \tag{A8}$$

From (A8) and (15), we have:

$$k N = k \left(\sum_{j=1}^{J} \left(T_0 - \phi_{jp} \bar{k}_j - \phi_{j0} \right) H_j^{m_j} \bar{N}_j \right) = \sum_{j=1}^{J} \bar{k}_j \bar{N}_j . \tag{A9}$$

Insert (A9) and (A7) in (A5):

$$\bar{k}_1 = \Lambda_k \left(k_i , \left(H_j \right), \left\{ \bar{k}_j \right\} \right) \equiv \frac{\Lambda_0 - \phi \phi_{10} - \left(T_0 - \phi_{10} \right) k_i H_1^{m_1}}{\phi \phi_{1p} - 1 - \phi_{1p} k_i H_1^{m_1}}, \tag{A10}$$

where:

$$\Lambda_0 \left(k_i , \left(H_j \right), \left\{ \bar{k}_j \right\} \right) \equiv \frac{1}{\bar{N}_1} \sum_{j=2}^{J} \Big[\bar{k}_j - \left(T_0 - \phi_{jp} \bar{k}_j - \phi_{j0} \right) k_i H_j^{m_j}$$

$$- \phi \left(\phi_{jp} \bar{k}_j + \phi_{j0} \right) \Big] \bar{N}_j ,$$

where $(H_j) \equiv (H_1,\dots,H_J)$ and $\{\bar{k}_j\} \equiv (\bar{k}_2,\dots,\bar{k}_J)$. It is straightforward to confirm that all the variables can be expressed as functions of k_i, (H_j) and $\{\bar{k}_j\}$ by the following procedure: k_e by (A1) $\rightarrow r$ and w_j by (4) $\rightarrow p$ by (A2) $\rightarrow \bar{k}_1$ by (A10) $\rightarrow k$ by (A9) $\rightarrow K$ by (A8) $\rightarrow T_j$ by (A6) $\rightarrow T_{je}$ by (A4) $\rightarrow N$ by (A7) $\rightarrow n_i$ and n_e by (A3) $\rightarrow N_i = n_i N$ and $N_e = n_e N$ $\rightarrow K_i = k_i N_i$ and $K_e = k_e N_e \rightarrow F_i$ by (4) $\rightarrow F_e$ by (5) $\rightarrow \bar{y}_j$ by (9) $\rightarrow$ c_j and s_j by (11). From this procedure and (11), we have:

$$\dot{H}_j = \Lambda_j\left(k_{1i}, \left(H_j\right), \left\{\bar{k}_j\right\}\right) \equiv \frac{\upsilon_{je} F_e^{a_{je}} \left(H_j^{m_j} T_{je} \bar{N}_j\right)^{b_{je}}}{H_j^{\pi_{je}} \bar{N}_j} + \frac{\upsilon_{ji} F_i^{a_{ji}}}{H_j^{\pi_{ji}} \bar{N}_j} + \frac{\upsilon_{jh} C_j^{a_{jh}}}{H_j^{\pi_h} \bar{N}_j} - \delta_{jh} H_j . \tag{A11}$$

Here, we don't provide explicit expressions of the functions as they are tedious. Substituting $\bar{y}_j = (1 + r)\bar{k}_j + T_0 w_j$ in $s_j = \lambda_j \bar{y}_j$ yields:

$$s_j = \left(1 + r\right)\lambda_j \bar{k}_j + \lambda_j T_0 w_j . \tag{A12}$$

Substituting (A12) in (12), we have:

$$\dot{\bar{k}}_1 = \lambda_1 T_0 w_1 - R\left(k_i, H_1\right)\bar{k}_1 , \tag{A13}$$

$$\dot{\bar{k}}_j = \bar{\Lambda}_j\left(k_i, \left(H_j\right), \left\{\bar{k}_j\right\}\right) \equiv \lambda_j T_0 w_j - \left(1 - \lambda_j - \lambda_j r\right)\bar{k}_j , \quad j = 2, ..., J , \tag{A14}$$

in which $R(k_i, H_1) \equiv 1 - \lambda_1 - \lambda_1 r$. Taking derivatives of equation (A10) with respect to t yields:

$$\dot{\bar{k}}_1 = \frac{\partial \Lambda_k}{\partial k_i}\dot{k}_i + \sum_{j=1}^{J} \Lambda_j \frac{\partial \Lambda_k}{\partial H_j} + \sum_{j=2}^{J} \bar{\Lambda}_j \frac{\partial \Lambda_k}{\partial \bar{k}_j}, \tag{A15}$$

where we use (A11) and (A14). Equaling the right-hand sides of equations (A13) and (A15), we get:

$$\dot{k}_i = \bar{\Lambda}_1\left(k_i, \left(H_j\right), \left\{\bar{k}_j\right\}\right) \equiv \begin{bmatrix} \lambda_1 T_0 w_1 - R \Lambda_k \\ - \sum_{j=1}^{J} \Lambda_j \frac{\partial \Lambda_k}{\partial H_j} - \sum_{j=2}^{J} \bar{\Lambda}_j \frac{\partial \Lambda_k}{\partial \bar{k}_j} \end{bmatrix} \left(\frac{\partial \Lambda_k}{\partial k_i}\right)^{-1} . \tag{A16}$$

In summary, we have proved Lemma 1.

■

REFERENCES

Aakvik, A., K.G. Salvanes, and K. Vaage (2010), "Measuring heterogeneity in the returns to education using an education reform," *European Economic Review* **54**: 453–500.

Abel, A.B. and B.S. Bernanke (1998), *Macroeconomics, 3rd edition.* New York: Addison-Wesley.

Aghion, P. and P. Howitt (1998), *Endogenous growth theory.* Cambridge: The MIT Press.

Arrow, K.J. (1962), "The economic implications of learning by doing." *Review of Economic Studies* **29**: 155–73.

Attanasio, O.P. and G. Weber (2010), "Consumption and saving: Models of intertemporal allocation and their implications for public policy," NBER Working Paper 15756.

Azariadis, C. (1993), *Intertemporal macroeconomics.* Oxford: Blackwell.

Banerjee, A.V. (2004), "Educational policy and the economics of the family," *Journal of Development Economics* **74**: 3–32.

Barro, R.J. (2001), "Human capital and growth," *American Economic Review, Papers and Proceedings* **91**: 12–17.

Barro, R.J. and X. Sala-i-Martin (1995), *Economic growth.* New York: McGraw-Hill, Inc.

Becker, G. (1981), *A treatise on the family.* Cambridge: Harvard University Press.

Behrman, J., R. Pollak, and P. Taubman (1982), "Parental preferences and provision for progeny," *Journal of Political Economy* **90**: 52–73.

Bénabou, R. (2002), "Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency?" *Econometrica* **70**: 481–517.

Bertocchi, G. and M. Spagat (2004), "The evolution of modern educational systems: Technical vs. general education, distributional conflict, and Growth," *Journal of Development Economics* **73**: 558–82.

Boucekkine, R., D. de la Croix, and O. Licandro (2002), «Vintage human capital: Demographic trends and endogenous growth," *Journal of Economic Growth* 104: 340–75.

Burmeister, E. and A.R. Dobell (1970), *Mathematical theories of economic growth.* London: Collier Macmillan Publishers.

Cardak, B.A. (2004), "Education choice, endogenous growth and income distribution," *Economica* **71**: 57–81.

Castelló-Climent, A. and A. Hidalgo-Cabrillana (2012), "The role of education quality and quantity in the process of economic development," *Economics of Education Review* (forthcoming).

Chen, M.K. and J.A. Chevalier (2008), "The taste for leisure, career choice, and the returns to education," *Economics Letters* **99**: 353–56.

Could, E.D., O. Moav, and B.A. Weinberg (2001), “Precautionary demand for education, inequality, and technological progress,” *Journal of Economic Growth* **6**: 285–315.

Cox, D. (1987), “Motives for private income transfers,” *Journal of Political Economy* **95**: 508–46.

Dur, R. and A. Glazer (2008), “Subsidizing enjoyable education,” *Labour Economics* **15**: 1023–39.

Easterlin, R. (1981), “Why isn’t the whole world developed?” *Journal of Economic History* **41**: 1–19.

Ehrlich, I. and F. Lui (1997), “The problem of population and growth: A review of the literature from Malthus to contemporary models of endogenous population and endogenous growth,” *Journal of Economic Dynamics and Control* **21**: 205–4.

Erosa, A., T. Koreshkova, and D. Restuccia (2010), "How important is human capital? A quantitative theory assessment of world income inequality," *The Review of Economic Studies* **77**: 1421–49.

Fender, J. P. and Wang (2003), "Educational policy in a credit constrained economy with skill heterogeneity," *International Economic Review* **44**: 939-64.

Fernandez, R. and R. Rogerson (1998), “Public education and income distribution: A dynamic quantitative evaluation of education finance reform,” *American Economic Review* **88**: 813–33.

Fleisher, B., Y.F. Hu, H.Z. Li, and S.H. Kim (2011), “Economic transition, higher education and worker productivity in China,” *Journal of Development Economics* **94**: 86–94.

Frederick, S., G. Loewenstein, and T. O’Donoghue (2002), “Time discontinuing and time preference: A critical review,” *Journal of Economic Literature* **40**: 351–401.

Galor, O. and O. Moav (2004), “From physical to human capital accumulation: Inequality in the process of development,” *Review of Economic Studies* **71**: 1001–26.

Galor, O. and D. Weil (1999), “From Malthusian stagnation to modern growth,” *American Economic Review* **89**: 150–54.

Galor, O. and J. Zeira (1993), “Income distribution and macroeconomics,” *Review of Economic Studies* **60**: 35–52.

Glomm, G. and M. Kaganovich (2008), “Social security, public education and the growth-inequality relationship,” *European Economic Review* **52**: 1009–34.

Grossman, G.M. and E. Helpman (1991), *Innovation and Growth in the Global Economy.* Cambridge: The MIT Press.

Hanushek, E. and D. Kimko (2000), “Schooling, labor-force quality and the growth of nations,” *American Economic Review* **90**: 1194–204.

Heckman, J.J. (1976), “A life-cycle model of earnings, learning, and consumption,” *Journal of Political Economy* **84**: S11–44.

Krueger, A.B. and M. Lindahl (2001), "Education for growth: Why and for whom," *Journal of Economic Literature* **39**: 1101–36.

Laitner, J. (2000), "Earnings within education groups and overall productivity growth," *Journal of Political Economy* **108**: 807–32.

Lazear, E. (1977), "Education: Consumption or production," *Journal of Political Economy* **85**: 569–97.

Lucas, R.E. (1988), "On the mechanics of economic development," *Journal of Monetary Economics* **22**: 3–42.

Malchow-Møller, N., S. Nielsen, and J.R. Skaksen (2011), "Taxes, tuition fees, and education for pleasure," *Journal of Public Economic Theory* **13**: 189–215.

Maoz, Y.D. and O. Moav (1999), "Intergenerational mobility and the process of development," *Economic Journal* **109**: 677–97.

Mincer, J. (1974), *Schooling, experience and earnings.* New York: Columbia University Press.

Nakajima, T. and H. Nakamura (2009), "The price of education and inequality," *Economics Letters* **105**: 183–85.

Romer, P.M. (1986), "Increasing returns and long-run growth," *Journal of Political Economy* **94**: 1002–37.

Solow, R. (2000), *Growth theory – An exposition.* New York: Oxford University Press.

Tilak, J.B.C. (1989), *Education and Its Relation to Economic Growth, Poverty and Income Distribution: Past Evidence and Future Analysis.* Washington, D.C.: World Bank.

Tselios, V. (2008), "Income and educational inequalities in the regions of the European Union: Geographical spillovers under welfare state restrictions," *Papers in Regional Science* **87**: 403–30.

Uzawa, H. (1965), "Optimal technical change in an aggregative model of economic growth," *International Economic Review* **6**: 18–31.

Zhang, W.B. (2005), *Economic Growth Theory.* London: Ashgate.

Zhang, W.B. (2007), "Economic growth with learning by producing, learning by education, and learning by consuming," *Interdisciplinary Description of Complex Systems* **5**: 21–38.

Zhang, W.B. (2009), "A discrete one-sector growth model with endogenous labor supply and heterogeneous groups," *Discrete Dynamics in Nature and Society* 2009.

Zhu, R. (2011), "Individual heterogeneity in returns to education in urban China during 1995-2002," *Economics Letters* **113**: 84–7.

Vol. 50 No. 2 (Nov, 2013), 289–326

LAJE
Latin American Journal of Economics

POLARIZATION AND THE MIDDLE CLASS IN URUGUAY*

Fernando Borraz**
Nicolás González***
Máximo Rossi****

Some approaches to measuring the middle class are based on an arbitrary definition such as income quartiles or the poverty line. Foster and Wolfson have recently developed a methodology without arbitrariness. We apply this tool and a complementary method–the relative distribution approach–to analyze the evolution of the middle class and polarization in Uruguay during the 1994-2004 and 2004-2010 periods. During the first period, characterized by increasing income inequality, the middle class declines and income polarization increases. In the second period, which includes the recovery from the 2002 downturn, we find that the middle class increases and polarization decreases.

JEL Classification: D3, D6, I3

Keywords: income polarization, bipolarization, middle class, inequality, social policies

1. Introduction

Barriers to investment are usually related to capital taxes, legal restrictions (i.e., red tape encountered in the creation of a firm), investment restraints, other taxes, etc. However, Azzimonti (2011) finds that "highly polarized societies tend to grow at a lower rate and converge to lower levels of income per capita in the long run." Easterly (2001) concludes that "A higher share of income for the middle class and lower ethnic divisions are empirically associated with higher income and higher growth. These associations are robust to a number of alternative controls." There is a great deal of research that analyzes the relationship between the size of the middle class and economic performance and political and social attitudes. Loaiza *et al.* (2012) find that the size of the middle class increases social spending

* The views expressed here are those of the authors and do not reflect necessarily the views of the *Banco Central del Uruguay*. We are grateful to Marisa Bucheli, Andrea Vigorito, and Marcel Vaillant for useful comments. All errors are our responsibility.

** *Banco Central del Uruguay* and Economics Department, FCS-UDELAR. Email: fborraz@bcu.gub.uy.

*** CEMFI and *Universidad de Montevideo*. Email: ngonzalez@correo.um.edu.uy.

**** Economics Department, Faculty of Social Sciences, *Universidad de la República*. Email: mito@decon.edu.uy.

doi 10.7764/LAJE.50.2.289

on health and education, improves democratic participation and is associated with low levels of corruption, without affecting economic freedom. Specifically, some empirical research has emphasized that an expansion of the middle class leads to better institutional outcomes (Barro, 1999; Easterly, 2001), as well as its effect on human capital accumulation, consumption and savings. López-Calva *et al.* (2012) find that a characteristic of middle-class values is moderation.

Therefore, another important factor for economic development is the degree of income polarization. In addition, polarization is linked to other relevant features for the development of a country such as social stability or lack of social conflict. In that sense, Gasparini *et al.* (2008) find that for LAC countries in the period 1989-2004, high levels of income polarization are positively correlated with a high level of social conflict.

Polarization measures are also strongly related to the size of the middle class. Splitting an economy into three categories (lower, middle, and upper) according to an income measure, a declining middle class could be an indicator of increasing polarization. For instance, polarization could increase in the case of bipolarization, when we observe greater mass in the lower and upper tails of the income distribution than in the middle. From an economic and social perspective, the middle class could play an important role in the development of a democratic country since it contributes a significant share of the labor force, and therefore is closely related to the country's output and usually represents the main source of tax revenue.[1] Moreover, an increase in the middle class resulting from the reduction of the lower and upper classes could enhance the positive externalities mentioned above, that is, the reduction of income inequality and antagonism between classes, which is an important source of social tension.

The aim of this paper is to define and characterize the middle class and to analyze the evolution of income polarization in a middle-income country (Uruguay) in recent years. Uruguay is an interesting case study because it is a middle-income country (with a large proportion of households around the median of the income distribution) and was the least polarized Latin American country at the end of the 1990s (see Gasparini, *et al.* 2008). Moreover, we focus on two particular periods during which the middle class is affected by economic conditions and

1. In Uruguay, considering the 2001 tax system, Grau and Lagomarsino (2002) show that the first two income quintiles contribute 22% of tax revenue, while the top two income quintiles contribute 18%. The middle-income deciles (third to sixth) contribute 60%.

political measures. In regard to the first period (1994-2004), previous research has observed a tendency toward income inequality during the 1990s in almost all Latin American and Caribbean (LAC) countries (IADB, 1998 or Bourguignon and Morrison, 2002). Uruguay is not an exception and during the 1990s the country did experience an increase in income inequality (Amarante and Vigorito, 2006). In the second period (2004-2010), inequality tended to decrease and the Uruguayan economy experienced a recovery from the downturn suffered in 2002. There is a common perception that because of the vigorous economic growth in recent years, the middle class in LAC countries is declining, but this perception is rarely confirmed by the research.

What makes this latter period interesting is that different kinds of redistributive policies, which potentially could have an impact on income distribution, have been introduced. For instance, in 2005 a conditional cash transfer program was launched[2] and the real minimum wage grew 63%. In addition, in 2007 a tax reform was implemented. Rodriguez and Perazzo (2007) conclude that the changes in sales tax (VAT) favors households in the first and last quintiles of the income distribution. Regarding the new personal income tax, Barriex and Roca (2007) find that the Gini index decreases 0.022 points. A shortcoming of these studies on the implications of the tax reform for income distribution is the lack of general equilibrium effects. However, we expect the redistributive policies to have an impact on income distribution.

Additionally, between 2005 and 2010 the Uruguayan economy grew around 30% in real terms (5% yearly). In the literature, the relationship between inequality and economic growth remains unresolved (Aghion, Caroli, and Gracia-Penalosa, 1999). For instance, growth could lead to wage inequality by spreading the gap across educational cohorts. Nevertheless, the new theoretical framework does not imply a trade-off between growth and inequality. In recent years, we have observed that income inequality fluctuates without a trend. Therefore, if growth is positively correlated with inequality then policy efforts could slow down inequality but not reduce it.

However, an opening question is, what is the appropriate definition of the middle class? Social class is a concept with a long history in sociology and economics. From the point of view of sociology it refers to the place of people in a social hierarchy, based on opportunities,

2. See Borraz and González (2009).

lifestyles and economic and social attitudes (Lora and Fajardo, 2011). Goldthorpe (1980) develops a social stratification analysis with seven categories in terms of source and level of personal income, degree of economic security and possibility of economic advancement, location within the systems of authority and control of production processes and performing their tasks and roles (Marshall, 1998 and Goldthorpe, 1980). The middle class in economics is usually defined on the basis of the distribution of a social welfare indicator such as household income (the most commonly used), household expenditure, labor status, educational attainment, etc. Most of the time economists use objective income definitions but some authors use self-perceived social status measures (Lora and Fajardo, 2011). In a new approach, López-Calva and Ortiz Suarez (2012) develop a vulnerability approach using the probability of falling into poverty, in order to determine the lower income bound of the middle class. In this paper, the definition of the middle class is related to the distribution of one variable: income. Therefore, the main problem is the arbitrary identification of the range of the income distribution that represents the middle class. The literature is not unanimous on this issue (see Foster and Wolfson, 2009 for further discussion of developed countries and Cruces *et al.*, 2010 for developing countries) and different definitions could lead to diverse and incomparable results. In order to analyze the evolution of the middle class, Foster and Wolfson (2009) develop a methodology that lacks arbitrariness in that it is based on the concept of "partial orderings" and first (and second) degree stochastic dominance. This method yields two curves (one for each population we would like to compare) that enables us to unambiguously determine which distribution concentrates more population around its median and also a bipolarization index.

Another complementary measure that enables comparison of the entire income distribution at two different points in time, in order to analyze the evolution of the entire distribution, is that developed by Handcock and Morris (1998, 1999). They provide the theoretical framework for the relative distribution approach, which enables us to compare two different distributions. Moreover, this non-parametric methodology gives us the tools to separately estimate the effects attributable to changes in the shape of the income distribution and those which come from changes in the location of the income distribution.

In the literature, the best definition of the middle class is still the subject of debate. Banerjee *et al.* (2008) characterize the middle class around the world, concluding "Nothing seems more middle class than the fact of having a steady well-paying job."

Several other issues motivate us to carry out this research: 1) the opportunity to apply different complementary tools to analyze the Uruguayan case (to show whether there are discrepancies) and contribute new evidence for discussion of this topic; 2) the opportunity to analyze whether the tendency toward income polarization and inequality observed during the 1990s has reversed in recent years; and 3) the opportunity to analyze the sensitivity of the results to certain components of household income, specifically, if we do not consider imputed income as a result of the new national healthcare system (NHS) implemented in Uruguay in 2008.

For our research we use data from the Uruguayan National Household Survey to apply different and complementary methodologies. In order to define the middle class, we follow Esteban and Ray (1994) and estimate a multinomial (and ordered) logit model to disentangle some features of the middle class. To quantify polarization and bipolarization we compute the polarization index developed by Duclos *et al.* (2004) and the bipolarization index derived by Foster and Wolfson (2009). We also use Foster and Wolfson's curves to analyze the evolution of the middle class. Finally, following Handcock and Morris (1998, 1999) we apply the relative distribution approach. We conclude that the middle class shrinks while income polarization increases between 1994 and 2004 and decreases from 2004 to 2010. However, this last result is attenuated when we do not consider the household income imputation due to the new healthcare system. Finally, the different approaches applied yield similar outcomes and therefore, we do not find discrepancies or inconsistencies across methods.

2. Measuring the middle class

One important issue regarding the concept of the middle class is the lack of consensus about the definition of the term, principally because different definitions lead to dissimilar results. Using the income distribution function, our main concern is to define which specific income range the middle class belongs to. For instance, let m be the middle of the income distribution measured by the median. We could consider that those households with income between $m - \varepsilon$ and $m + \varepsilon$ belong to the middle class and therefore, the proportion of households in the range represent a measure of middle class size. However, this definition depends on the value of ε. In this context, the methodology proposed by Foster and Wolfson (2009) is not subject to a specific income interval and hence it does not suffer from arbitrariness. This approach is derived from the idea of partial ordering and stochastic dominance.

Let F represent an income distribution function in one period. Since different distribution functions might have different medians, we consider a median-normalized distribution denoted as $\widetilde{F}$ to make a robust comparison between two different distributions functions. The middle class index M for $\widetilde{F}$ given an income range $I = [\underline{\varepsilon}, \bar{\varepsilon}]$ is defined as:

$$M_{\widetilde{F}}(I) = M_{\widetilde{F}}(\underline{\varepsilon}) + M_{\widetilde{F}}(\bar{\varepsilon}) = [\widetilde{F}(1) - \widetilde{F}(\underline{\varepsilon})] + [\widetilde{F}(\bar{\varepsilon}) - \widetilde{F}(1)] \quad \text{with } 0 \leq \underline{\varepsilon} \leq 1 \leq \bar{\varepsilon} \tag{1}$$

where $M_{\widetilde{F}}(\underline{\varepsilon})$ and $M_{\widetilde{F}}(\bar{\varepsilon})$ are the "lower middle class" and the "upper middle class," respectively, and $\widetilde{F}(1) = 0.5$. Conceptually, $M_{\widetilde{F}}(I)$ is defined as a function which gives the share of population for the income range I. In other words, the function $M_{\widetilde{F}}(I)$ enables measurement of the proportion of households around the median for a median-normalized distribution. For example, for the income range $I_1 = [0.5, 1.5]$ we obtain the following middle-class index: $\int M_{\widetilde{F}}(I_1) = M_{\widetilde{F}}(0.5) + M_{\widetilde{F}}(1.5)$. By considering different income ranges, we are able to construct a curve that is not restricted to one particular definition of the middle class: $M_{\widetilde{F}}(R_i)$ with $i = 1,...n$, where the index i denotes the income range thus giving the idea that the latter measurement supports any definition of the middle class. We can construct the $M(\cdot)$ function for two mean-normalized distributions functions in order to compare them. Hence, Foster and Wolfson (2009) define M as a *partial ordering*. Thus, considering two distribution functions F and G and using the notion of partial ordering, the following binary relation M can be stated[3]:

Proposition 1.

$$FMG \Leftrightarrow M_{\widetilde{F}}(R_i) \geq M_{\widetilde{G}}(R_i) \ \forall i = 1,\ldots,n$$
$$\text{and } M_{\widetilde{F}}(R_i) > M_{\widetilde{G}}(R_i) \text{ for some } i$$

In other words, if Proposition 1 holds, "F has an unambiguously larger middle class than G," for any definition of the middle class. That is, the distribution F accumulates more mass around its median than distribution G, which accumulates more mass in the upper and lower

3. See Foster and Wolfson (2009) for more details.

tails. In our case, we estimate three curves, one for the 1994 income distribution, another for the 2004 income distribution, and finally one for the 2009 income distribution. After that, we compare 1994 with 2004 and this latter year with 2010. If the estimated curves do not cross at any point in each period, we are able to draw an unambiguous conclusion about the evolution of the middle class during both periods. Otherwise, we only have the information about the different income ranges that support prior definitions. At this stage, it is important to point out that this analysis is based on the income distribution across years which is constructed using cross-section data. It would be interesting to consider panel data to construct a transition matrix for determining which households change categories over time, in order to analyze the dynamics of social classes and mobility.

2.1. Polarization measures

A declining middle class could be related with a more bipolar income distribution whenever the middle-class reduction occurs jointly with an increase in the lower and upper classes. The Foster and Wolfson bipolarization index and polarization curves are based on the idea that movements away from the middle via increased spread or more distant extremes in the income distribution lead to a rise in polarization. Thus, they divide the income distribution in two, forming two income groups, one above and one below the median. The approach to derive the first "degree" polarization curve is similar to the one used to measure the middle class, but here the aim is to determine the income interval that includes all the households belonging to a given population range. For example, for a given population range $Q = [\underline{w}, \bar{w}]$ the distribution F has a certain income range. The greater the income range required to quantify any defined population range, the greater the income spread (growth in polarization). Hence, we are interested in measuring income spread as the width of the income range in the distribution F given a population range. Formally,

$$\begin{aligned} S_F(w_i) &= \left| \widetilde{F}^{-1}(w_i) - \widetilde{F}^{-1}(0.5) \right| \\ \text{with } 0 \leq w_i &\leq 1 \quad \forall i = 1,\ldots,n \end{aligned} \tag{2}$$

Note that in this case i refers to population range. Again, using the notion of partial ordering, the following proposition is derived[4]:

4. See Foster and Wolfson (2009) for more details.

Proposition 2.

$$FSG \Leftrightarrow S_F(w_i) \geq S_G(w_i) \ \forall i = 1,\ldots,n$$
$$\text{and } S_F(w_i) > S_G(w_i) \text{ for some } i$$

This proposition states that for a given population range i the income distribution F reveals a greater income spread than the income distribution G, that is, F has a greater income polarization than G. This result holds for any population range. Furthermore, since a greater income spread implies a lower proportion of population around the middle, Proposition 3 implies that income distribution G has a larger middle class than income distribution F, and therefore G dominates *F (GMF).* Additionally, Foster and Wolfson construct a second curve called "second-degree" polarization which simultaneously considers both sources of polarization: "increased spread" and "increased bipolarity." It is defined as the area under the first degree polarization curve between 0:5 and a population share w_i:

$$B_F(w_i) = \left| \int_{w_i}^{0.5} S_F(p)\,dp \right| \text{ with } 0 \leq w_i \leq 1 \quad \forall i = 1,\ldots,n \tag{3}$$

The second-degree polarization curve is similar to the Lorenz curve, which accumulates the population share from the lowest to the highest incomes. This new curve accumulates income spreads from the middle to the top and the bottom, respectively, and places more weight on changes around the middle of the income distribution. The following proposition applies when income distribution F presents a greater level of polarization than income distribution G[5],

Proposition 3.

$$FBG \Leftrightarrow B_F(w_i) \geq B_G(w_i) \ \forall i = 1,\ldots,n$$
$$\text{and } B_F(w_i) > B_G(w_i) \text{ for some } i$$

Finally, Foster & Wolfson construct a polarization index consistent with the first and second polarization curves and similar to the Gini index. It is defined as twice the area under the second-degree polarization curve: $P = \int_0^1 2B_F(w)\,dw$. As mentioned before, this analysis is based on an

5. See Foster and Wolfson (2009) for more details.

income distribution that is divided into two groups: those with income below the median and those with income above the median. For this reason, this index can be defined as a bipolarization index. A greater value could be indicative of a greater income spread between these two groups and/or that the groups have become more sharply defined. The distance between these two groups as a proportion of the overall mean is defined as the *relative median deviation*: $\mathrm{T} = (\mu^U - \mu^L)/\mu$. Then, it can be proved that: 1) $T = 2G^B$, where G^B is the between-groups Gini index; 2) $G = G^B + G^W$, that is, the Gini index is equal to the sum of the between Gini index G^B and the within-groups Gini index G^W; and 3) the polarization index is equal to $P = (T - G)(\mu / m)$, where μ is the overall mean and m is the median. Based on these three results, we can define the polarization index as:

$$P = (G^B - G^W)\frac{\mu}{m} \tag{4}$$

Equation (4) reflects the fact that an increment in inequality between the two defined groups raises polarization, in other words it increases alienation. However, an increment in inequality in each group decrease polarization, that is, each group is less homogeneous. Equation (4) also tells us that polarization increases depending on the source of inequality and thus, polarization and inequality may or may not move in the same direction. For example, a rise in the spread of income distribution as a result of a regressive transfer tends to enhance both polarization and inequality. On the other hand, an increment in bipolarization as result of a progressive transfer leads to a growth in polarization but not in inequality.

The polarization measure presented above is focused on the idea of only two income groups. In order to relax this assumption and based on the concepts of alienation and identification, Esteban and Ray (1994) develop a polarization index in which the number of income groups are determined by the analyst or by using common rules. Formally,

$$P(F) = \iint T(I(y,F), r(\delta(y_i, y_j)) \, dF(x) \, dF(y) \tag{5}$$

where T is the “effective antagonism” between individual y and individual x (under F) which is compounded by the identification function I that measures the degree of association of an individual with a group in terms of income; and the alienation function, which measures the distance (usually the Euclidean metric) between the identified income groups. The main drawback of this index is that it assumes that individuals have been "regrouped" in each of the relevant groups. The problem then is how to set the optimal "partition" for a given number n of groups. Esteban *et al.* (1999) introduce some refinements to the previous polarization index in order to determine the optimal way to construct the boundaries that define the n groups. Relying on the assumption that the income distribution can be represented by a density function f in a bounded interval, the function f could have an "n-spike" representation denoted by ρ. The "n-spike" representation differs from the actual representation of f, in an error term $\varepsilon(f, \rho)$ which can be called the "grouping error." This error term needs to be introduced in order to correct the previous polarization measure. Moreover, the error term $\varepsilon(f,\rho)$ can be defined as $G\ (f) - G\ (p^*)$ which is the difference between the Gini index using the actual density function and the one that arises from optimally separating the population in defined n number of groups. Thus, this polarization measure is obtained by minimizing the within-group dispersion using an iterative procedure. The new polarization measure is:

$$P(f,\alpha,\beta) = \mathrm{ER}(\alpha,\rho) - \beta\varepsilon(f,\rho) \tag{6}$$

where ρ is the "n-spike" representation of the density function f, a is a parameter related to the importance of the identification factor and is defined by the user, and finally β is the weight placed on the grouping error term and it is also a user-defined parameter. As a result of the application of this method with $n = 3$, we can define the lower, middle and upper class because we can calculate the values of income that define each category. After that, we characterize the middle class and estimate a multinomial ordered logit to find out the main features of the middle class.

Duclos *et al.* (2004) extend the prior analysis by letting the number of groups be determined endogenously. The identification process is based on the estimation of a non-parametric kernel density for the income variable (y_i). The density for a given income range can be viewed as the proportion of population in this range. The degree of identification arises when this proportion or density is powered by the parameter α (with $\alpha \in [0,1]$), which is an ethical parameter that expresses the level of a sense of identification within a population group given by a level of income. In other words, for each density point a "window of identification" is defined. Individuals belonging to a particular window are weighted by their distance from each density point. In this context, the alienation factor is a measure of the income distance between each previously determined group. Then, the polarization index for the distribution F can be defined as,

$$P_\alpha(F) = \int_y f(y)^\alpha a(y)\, dF(y) \tag{7}$$

where y represents the income variable and F its distribution function. The identification effect, which is sensitive to the parameter α, is denoted as $f(y)^\alpha$ and finally, $a(y)$ denotes the alienation effect. One drawback of this index is that it is subject to the choice of the parameter α, which as mentioned previously is related to the identification process. A higher value of α emphasizes the role of identification in the construction of this polarization indicator. In contrast, when α is zero, there is no weight placed on the identification effect and therefore, the polarization index equals the alienation effect (the Gini index). In order to circumvent this disadvantage, we estimate Duclos *et al.*'s polarization index for a set of values of α. In addition, $f(y)^\alpha$ is estimated using a kernel procedure. We use a Gaussian kernel function and the "optimal" bandwidth is derived by minimizing the mean square error (see Duclos *et al.* for more details).

Finally, the polarization index can be decomposed as follows:

$$P_\alpha(f) = \bar{a}\bar{i}_\alpha[1 + \rho] \tag{8}$$

where $\bar{a}$ is the average alienation effect, $\bar{i}_\alpha$ is the average identification effect and ρ is the normalized covariance between $\bar{i}_\alpha$ and a. This equation provides interesting information since we can observe the contribution of each component to polarization.

3. Relative distribution approach

Although this approach is different than those previously described, it can be viewed as a complement to them. Based on the "relative distribution" method, this tool is helpful for finding changes in patterns across the entire income distribution for a given period and it is also capable of distinguishing between changes in the location and shape of the income distribution. The theoretical framework is introduced by Handcock and Morris (1998, 1999) and assumes that we have two different populations: the "reference" population and the "comparison" population. The initial step is to define a relative rank. First, we introduce some notation: let Y_t and Y_{t+1} be the income variable with cumulative distribution functions F_t and F_{t+1}, respectively. Then, a relative rank R between 0 and 1 is defined as $R = F_t(y_{t+1})$. This relative rank is considered a random variable and it quantifies the accumulated mass of population in t according to the income variable in $t + 1$. For one realization of R we have $r = F_t(y_{t+1,r})$ with $0 \leq r \leq 1$ and the associated quintile function $F_t^{-1}(r) = y_{t+1,r}$. Then, the relative distribution function is defined as $G(r) = F_{t+1}(F_t^{-1}(r))$ with $0 \leq r \leq 1$ and the relative density function of interest is defined as:

$$g(r) = \frac{f_{t+1}(F_t^{-1}(r))}{f_t(F_t^{-1}(r))} \, with \; 0 \leq r \leq 1 \tag{9}$$

where f represents the density function in $t + 1$ and t, respectively; $g(r)$ is the relative density function evaluated at the income level of the reference group t at the quintile r. This function is defined as the ratio of the density of the reference group to the density of the comparison group evaluated in the income level of the reference group at quintile r. It has the properties of a density function (for example, it integrates to 1). When the relative density function shows values near one, it means that the two density functions have a similar density at the quintile r of the reference group and thus, R has a uniform distribution in the interval [0,1]. A relative density greater than one means that the comparison density has more density than the reference density evaluated at the quintile r of the reference group. Finally, a relative density function of less than one indicates the opposite.

The density functions are estimated using a non-parametric kernel method. Once we obtain the estimated relative density functions for different realizations of R, we fit a local polynomial for each estimated point in order to have an accurate description of the relative density. One of the major advantages of this method is the possibility to decompose the relative distribution into a location effect, usually associated with changes in the mean of the income distribution, and a shape effect, which could be linked with several factors, for instance social policies or polarization. Formally,

$$g(r) = \underbrace{\frac{f_{t+1}(F_t^{-1}(r))}{f_t(F_t^{-1}(r))}}_{\text{Overall effect}} = \underbrace{\frac{f_{t,L}(y_{t+1,r})}{f_t(y_{t+1,r})}}_{\text{Location effect}} x \underbrace{\frac{f_{t+1}(y_{t+1,r})}{f_{t,L}(y_{t+1,r})}}_{\text{Shape effect}} \text{ with } 0 \leq r \leq 1 \tag{10}$$

where $f_{t+1,L}(y_{t+1,r}) = f_{t+1}(y_{t+1,r} + \rho)$ is a density function adjusted by an additive shift $\rho = \text{median}(Y_{t+1}) - \text{median}(Y_t)$. An increasing location effect means that the comparison income distribution is greater than the reference income distribution and vice versa. The second term, which is the shape effect function, is useful for identifying movements in the entire distribution function. For instance, as a consequence of the redistributive policies launched in 2005 we could expect a reduction in the upper tail in 2010, which could lead to an increase in the middle class, observing a shape effect function with some sort of U form. We could expect the opposite (an inverse U shape) if we compare the 1994 income distribution with the 2004 income distribution.

This approach also includes a "median relative polarization index" that is based on changes in the shape of the income distribution to account for polarization. This index measures the average of the absolute value from the median of the shape effect function normalized to vary between -1 and 1. Negative values indicate that income polarization decreases, while positive values indicate the opposite. When the index value is zero, there are no changes in polarization patterns. The index is formally defined for the reference population (period $t+1$) and the comparison population (period t) as follows:

$$MRP \text{ index} = 4\int_0^1 \left| r - \frac{1}{2} \right| \underbrace{\frac{f_{t+1}(y_{t+1,r})}{f_{t,L}(y_{t+1,r})}}_{g_s(y_{t+1,r})} dr - 1 \tag{11}$$

where $g_s(y_{t+1,r})$ is the shape effect function. The index can be estimated using non-parametric techniques. Finally, the MRP index can be decomposed into a lower and upper relative polarization index, which are also normalized to vary between -1 and 1. These two new indices can shed light on income bipolarization and therefore on issues concerning the declining middle class. They are formally defined as:

$$LRP \text{ index} = 8\int_0^{1/2} \left| r - \frac{1}{2} \right| g_s(y_{t+1,r})\, dr - 1 \tag{12}$$

$$URP \text{ index} = 8\int_{1/2}^1 \left| r - \frac{1}{2} \right| g_s(y_{t+1,r})\, dr - 1 \tag{13}$$

4. Data and results

We use the annual National Household Survey (ECH) conducted yearly by the National Statistical Office of Uruguay (INE).We employ cross-sectional data for 1994, 2004 and 2010 to analyze two different periods, 1994-2004 and 2004-2010. The first period is characterized by increasing inequality and it comprises the 2002 economic downturn[6], while in the second one redistributive policies were introduced and average yearly real GDP growth was 6%. The ECH is the main source of socioeconomic information on Uruguayan households and their members at the national level. Because the 1994 and 2004 surveys only include households in urban areas with more than 5,000 inhabitants, we restrict the analysis to this population.[7] We are interested in the total household income variable of the survey. This variable includes all sources of income (salaries, pensions, benefits from cash transfer programs, etc.) as well as imputed income (for example, in the case of homeowners, the imputed rental income is the hypothetical value that household members would have to pay for it). It is necessary to point out that the household income reported in the survey is net of social security and income taxes. Specifically, our outcome variable is the per-capita household income in March 1997 Uruguayan pesos since we adjust it by the consumer price index with base in March 1997.

4.1 Characterizing the middle class

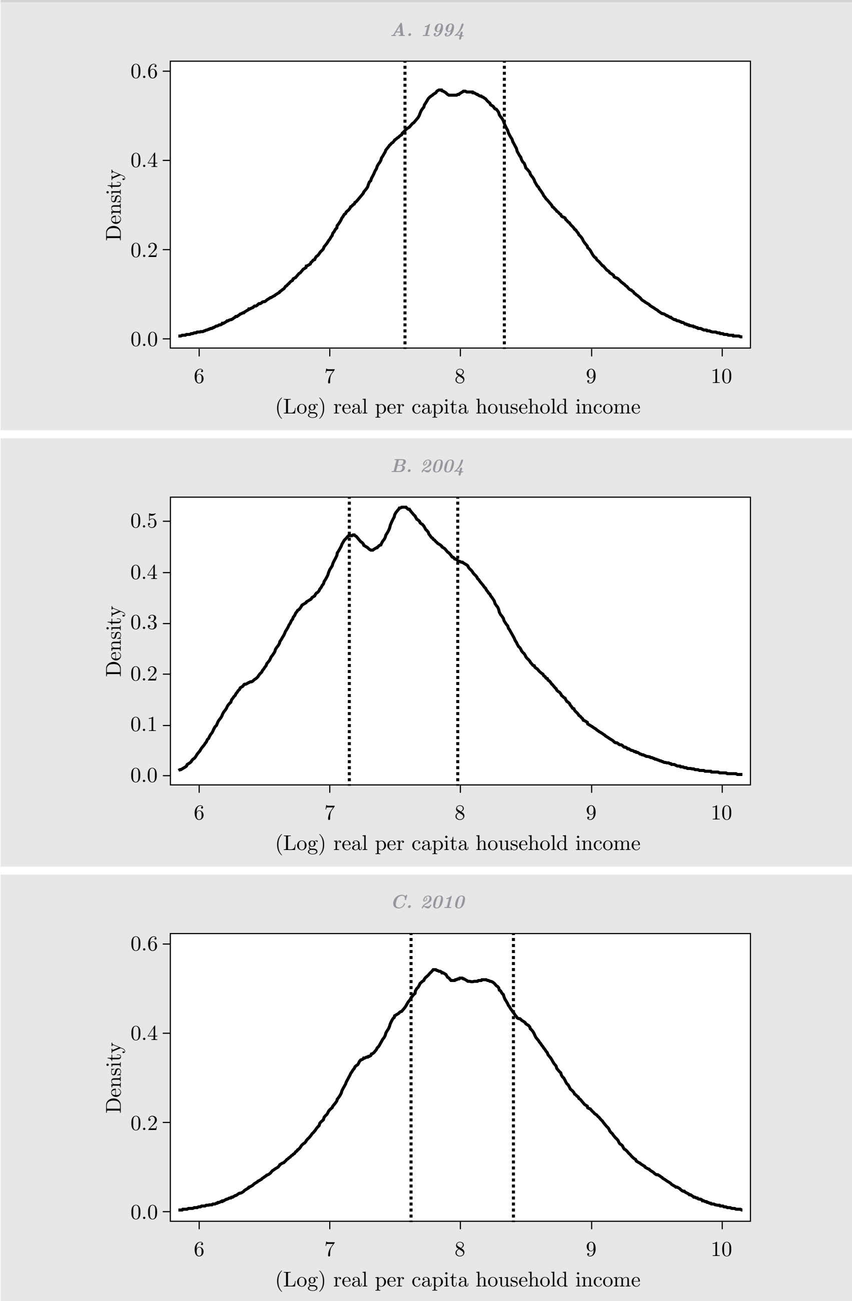
In this section, we define and characterize the middle class following Esteban and Ray (1994) and then compare it with the other social classes (lower and upper). In Figure 1, we observe the density of the (log) real household income jointly with the middle class boundaries in 1994, 2004 and 2010. In 1994 and 2010 the definition seems to be quite similar, while in 2004 the middle-class interval shifts to the left, probably due to the 2002 economic crisis.

Based on these middle-class intervals, Table 1 shows summary statistics of the middle classes. First of all, we observe that in Uruguay around 37% of the households belong to the middle class. The low-income class is the largest, with approximately 45%, and the upper class is the smallest (around 12%). Therefore, Uruguay is basically comprised

6. Real GDP decreased 11% in 2002 and unemployment reached 17% that year.

7. Note that only around the 5% of the Uruguay population is located in rural areas.

Figure 1. Middle class definition



Source: Authors' computations based on data from ECH.

Table 1. Characteristics of social classes: summary statistics

Variables	1994			2004			2010		
	Low	Middle	High	Low	Middle	High	Low	Middle	High
% of persons	54.17	32.50	13.33	58.29	30.16	11.56	56.62	31.37	12.01
% of households	44.57	37.26	18.17	46.42	36.95	16.62	45.45	37.15	17.40
Average (per capita) household income	1.977	4.578	11.427	1.346	3.396	9.252	2.103	4.975	12.891
Standard deviation	716.72	1.002	6.134	520	794	5.859	751	1.115	8.782
Household income share	25.68	37.26	37.06	23.14	35.33	41.53	26.01	37.03	36.96
% of labor income	61.67	60.28	60.1	60.06	58.17	57.21	58.88	60.96	59.15
% of hhs below the poverty line	29.69	0.00	0.00	57.33	0.00	0.00	27.72	0.00	0.00
% of hhs below the extreme poverty line	1.56	0.00	0.00	4.66	0.00	0.00	1.11	0.00	0.00
Education									
Average household education	6.47	7.85	10.8	7.56	9.24	12.50	7.59	9.61	12.80
% households in < 7	51.98	38.80	16.88	36.36	27.59	11.08	34.73	23.19	9.25
% households in [7,9]	32.85	28.58	18.26	38.02	23.08	10.04	40.36	23.75	9.88
% households in [10,12]	12.02	21.34	28.70	20.36	27.77	24.53	19.97	29.99	25.65
% households in >12	2.97	11.27	36.16	5.25	21.56	54.36	4.94	23.07	55.22
Head of hhld years of education	6.05	7.41	10.74	7.17	8.94	12.60	7.30	9.38	12.80
% head of hhld with high school degree	10.34	20.44	47.47	9.51	28.14	62.9	8.34	28.24	61.23
% head of hhld with university degree	0.59	2.79	15.59	0.65	5.23	24.19	0.28	3.01	16.04

Table 1. (continued)

Variables	1994			2004			2010		
	Low	Middle	High	Low	Middle	High	Low	Middle	High
Attendance rate by age interval:									
[6,12]	98.23	99.78	99.53	98.34	99.23	99.14	98.88	99.12	99.51
[13,17]	68.86	86.56	94.47	82.83	95.84	99.45	80.56	95.09	98.68
[18,23]	20.95	42.46	61.83	33.59	64.99	83.05	28.42	55.85	77.55
Average education gap - children in [7,15]	0.76	0.37	0.32	0.76	0.34	0.31	0.73	0.34	0.33
% of children in [7,15] with education gap	43.16	25.69	26.48	46.64	27.00	26.68	47.22	29.02	28.69
Living conditions									
Homeownership	59.91	73.27	81.05	57.88	71.24	81.17	54.38	63.65	71.65
Persons per room	2.07	1.45	1.18	2.00	1.38	1.20	1.87	1.28	1.03
% of overcrowded households	30.60	7.05	1.32	27.02	3.64	0.93	22.66	2.74	0.37
Water supply - general network	97.22	98.44	99.43	98.24	99.22	99.75	98.17	98.95	99.57
Network evacuation	44.62	69.90	91.08	52.91	78.07	92.50	48.74	71.90	87.15
Asset index	0.14	0.22	0.33	0.17	0.29	0.42	0.24	0.33	0.44
Wealth Index	-0.80	0.10	1.22	-1.14	0.33	1.60	-1.04	0.40	1.70
Household size	3.92	2.81	2.34	3.83	2.52	2.15	3.54	2.40	1.95

Table 1. (continued)

Variables	1994			2004			2010		
	Low	Middle	High	Low	Middle	High	Low	Middle	High
Population composition									
Children ages 0-5	11.38	5.53	3.65	10.91	4.49	3.46	10.43	4.75	3.40
Children ages 12-17	16.37	8.47	6.37	16.64	7.15	5.49	17.09	7.25	4.91
Adults > 60	14.63	26.62	29.22	13.16	29.36	33.73	12.83	26.19	31.77
Labor status									
Employment rate	86.27	94.37	96.63	81.73	91.69	95.53	89.50	95.85	97.67
Wage earner	63.56	69.19	63.16	54.89	68.64	65.70	63.68	73.27	68.69
Self-employed	19.65	18.51	18.77	24.06	17.78	18.17	22.77	16.72	16.77
Entrepreneur	1.37	4.77	12.86	1.08	4.16	10.70	1.73	5.07	11.61
Zero income	1.67	1.89	1.79	1.69	1.10	0.96	1.27	0.78	0.58
Informal workers	27.64	17.33	10.65	25.47	14.33	6.91	25.24	13.05	5.22
Unemployment rate	13.72	5.62	3.37	18.26	8.30	4.46	10.49	4.14	2.32
% of inactive	30.48	36.01	35.21	29.05	37.64	38.18	27.16	31.30	32.33
% of pensioner	5.47	8.10	9.67	4.15	4.59	4.52	4.07	3.79	3.41
% of retired	8.00	13.56	11.63	6.82	19.19	20.90	5.98	15.60	18.30

Source: Authors' calculation based on the Uruguayan National Household Survey (ECH).
Note: Calculation based on real per-capita household income in 1997 Uruguayan pesos, net of social security and income tax. Data weighted using sample weights.

of low- and middle-income households. Another interesting feature we observe is a great income dispersion in the upper class, while the lower class appears to be more homogeneous. Also, the income share of the middle and upper classes seems to be similar between 1994 and 2010, despite their size differences. As expected, the income share of the lower class decreases in 2004, while that of the upper class increases.

We present a second group of indicators that are related to education. Overall, we observe that educational attainment increases from the lower to the upper class. For instance, if we consider the average years of education of adult household members, the upper class has the highest average while the lower class has the lowest[8]. The attendance rate is similar across classes for the age cohort [6, 12][9]. However, when we take into account higher cohorts the attendance rate decreases, mainly in the lower class case. In addition, the lower class shows a high education gap in children between 7 and 15 years old in comparison with the middle and upper class, which have a similar education gap.

Regarding living conditions, around 70% of the middle-class households own their homes. Nevertheless, it is interesting to note that this proportion has declined in recent years and around 64% are homeowners in 2010. This trend is also evident in the other social classes. In addition, there is a considerable difference in terms of overcrowded households between the lower and the middle and the upper classes. Sanitation is another variable that increases as we move to higher-income classes. We construct an asset index as a weighted average of a series of indicator variables for the availability of the following household assets: refrigerator, dishwasher, washing machine, broadcast TV, Internet connection, computer, car and household help. The weights are the relative distance between 1 and the proportion of households having this item and therefore the index places more weight on items possessed only by few households. The index varies between 0 and 1. The asset index shows a difference between the low and middle classes of around 0.10 point and this gap remains constant for the three years. The asset gap between the middle and upper class is wider (approximately 0.14). We construct another wealth index with the same variables but considering a normal, standardized transformation of them. In this case, we observe larger gaps and this index varies among a higher set of values than before.

8. The same conclusion arises when we consider the average years of education of the head of household.

9. This is not surprising since primary school attendance is almost universal.

In regard to population composition, the lower class is comprised of more younger people than the other classes, while the middle and upper classes have more adults older than 60.

The labor status indicators show that the unemployment rate is the highest for the lower-income class. The majority of the middle-class workers are wage earners, followed by the self-employed and then entrepreneurs. This pattern is quite similar in the other class categories. The major difference is that the high-income class has a greater proportion of workers in the entrepreneur category. We use the definition of informal workers adopted by the International Labour Organization at the 15th International Conference of Labor Statisticians (1993), which considers informal workers as those who work in the housekeeping sector, unpaid household members, private wage earners working in firms with less than five employees and self-employed workers (excluding administrative workers, professionals and technicians). Using this definition, the highest proportion of informal workers is in the low-income class. The proportion of informal workers in the middle class in 1994 is just over 0.17 and decreases in both 2004 and 2010. Finally, the middle and upper classes show a similar share of inactive people.

Table 2 presents multinomial logit estimates for the three years. As a dependent variable we use the category variable, which takes the value of 1 if the household belongs to the low-income category, 2 if it belongs to the middle and 3 if belongs to the high-income category. We consider the middle class as the base category and we report the marginal effects. It is interesting to note that the signs of the coefficients do not change when we consider different years and that almost all the coefficients are statistically different from zero at the 1% level. For instance, the probability of being a low-income household (with respect to being a middle-income household) decreases if the household is in the capital city (Montevideo). The opposite occurs when we analyze the probability of being a high-income household. This could be associated with differences in the cost of living between the capital and the rest of the country. As mentioned earlier, households with young children have higher probabilities of being low-income than middle-class. The same holds for the household size variable. A more educated head of household raises the probability of being high-income and decreases the probability of being low-income (with respect to the middle). Concerning labor market variables, a head of household who is unemployed or works in the informal sector increases the probability

Table 2a. Multinomial logit estimates

Variables	1994		2004		2010	
	Low	High	Low	High	Low	High
Capital	-0.132***	0.074***	-0.044***	0.055***	-0.028***	0.037***
	(0.007)	(0.006)	(0.007)	(0.006)	(0.005)	(0.004)
Household with children ages 0-5	0.080***	-0.048***	0.051***	-0.031***	0.075***	-0.032***
	(0.009)	(0.008)	(0.008)	(0.009)	(0.006)	(0.006)
Household with children ages 12-17	0.094***	-0.042***	0.078***	-0.027***	0.099***	-0.034***
	(0.008)	(0.007)	(0.008)	(0.008)	(0.005)	(0.005)
Household with adults > 60	-0.047***	0.019***	-0.055***	0.038***	-0.041***	0.015***
	(0.008)	(0.006)	(0.008)	(0.006)	(0.006)	(0.004)
Household size	0.083***	-0.075***	0.098***	-0.073***	0.111***	-0.088***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Head of hhld average education	-0.017***	0.015***	-0.018***	0.016***	-0.022***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Head of hhld unemployed	0.190***	-0.069***	0.153***	-0.118***	0.134***	-0.081***
	(0.022)	(0.020)	(0.015)	(0.020)	(0.013)	(0.014)
Head of hhld occupation: entrepreneur	-0.111***	0.072***	-0.099***	0.079***	-0.083***	0.072***
	(0.016)	(0.008)	(0.018)	(0.009)	(0.011)	(0.006)
Households with informal workers	0.032***	-0.040***	0.025***	-0.030***	0.052***	-0.052***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.005)	(0.005)
Homeownership	-0.094***	0.059***	-0.043***	0.046***	-0.048***	0.046***
	(0.007)	(0.006)	(0.007)	(0.006)	(0.004)	(0.004)

Table 2a. (continued)

Variables	1994		2004		2010	
	Low	High	Low	High	Low	High
Overcrowded households	0.060***	-0.042***	0.047***	-0.032*	0.031***	-0.050***
	(0.010)	(0.016)	(0.012)	(0.019)	(0.009)	(0.019)
Network evacuation	-0.060***	0.051***	-0.032***	0.005	-0.044***	0.017***
	(0.007)	(0.007)	(0.008)	(0.009)	(0.005)	(0.005)
Wealth index	-0.080***	0.050***	-0.078***	0.049***	-0.071***	0.046***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Pseudo R^2	0.374	0.374	0.414	0.414	0.395	0.395
Log Likelihood	-7,678.08	-7,678.08	-7,064.40	-7,064.40	-17,023.34	-17,023.34
Observations	11,906	11,906	11,748	11,748	27,914	27,914

Marginal effects and robust standard errors reported. Base category = middle class
* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2b. Ordered logit estimates

Variables	1994 Middle	2004 Middle	2010 Middle
Capital	0.042***	0.015***	0.012***
	(0.002)	(0.002)	(0.001)
Household with children ages 0-5	-0.026***	-0.013***	-0.021***
	(0.003)	(0.002)	(0.002)
Household with children ages 12-17	-0.028***	-0.018***	-0.026***
	(0.002)	(0.002)	(0.001)
Household with adults > 60	0.013***	0.015***	0.010***
	(0.002)	(0.002)	(0.001)
Household size	-0.030***	-0.027***	-0.035***
	(0.001)	(0.001)	(0.001)
Head of hhld average education	0.007***	0.006***	0.007***
	(0.000)	(0.000)	(0.000)
Head of hhld unemployed	-0.056***	-0.043***	-0.040***
	(0.007)	(0.004)	(0.004)
Head of hhld occupation: entrepreneur	0.035***	0.028***	0.029***
	(0.004)	(0.003)	(0.002)
Households with informal workers	-0.013***	-0.008***	-0.017***
	(0.002)	(0.002)	(0.001)
Homeownership	0.031***	0.014***	0.017***
	(0.002)	(0.002)	(0.001)
Overcrowded households	-0.015***	-0.012***	-0.007***
	(0.003)	(0.003)	(0.003)
Network evacuation	0.021***	0.007***	0.012***
	(0.002)	(0.002)	(0.001)
Wealth index	0.025***	0.021***	0.021***
	(0.001)	(0.001)	(0.000)
Pseudo R^2	0.368	0.413	0.393
Log Likelihood	-7,756.56	-7,077.75	-17,140.93
Observations	11,906	11,748	27,914

Marginal effects and robust standard errors reported. Base category = middle class
* significant at 10%; ** significant at 5%; *** significant at 1%.

of being low-income, while decreasing the probability of being high-income. In addition, if the head of household is an entrepreneur, this increases the probability of being high-income. The housing variables have the expected signs.

Because in this case the dependent variable seems to have a natural order, an ordered logit model appears to be the most appropriate. However, if this assumption does not hold we will have a bias estimator. Otherwise, the ordered logit model produces more efficient estimates than the multinomial logit. However, the results of both models are quite similar, so we do not report the ordered logit estimation.[10]

4.2. Evolution of the middle class and polarization

In this section, we apply the methodology related to the evolution of the middle class and the polarization measures. Table 3 presents summary statistics that help describe the income distribution for the different years. As we can see, the mean and the median of the income distribution fall between 1994 and 2004 and both increase in 2010. The mean is greater than the median, indicating that the income distribution is skewed left. With respect to income concentration, the first quintile has approximately 5% of the total income, while the fifth quintile represents approximately 50%. Interestingly, during the first period the proportion of the first quintile declines, whereas that of the fifth rises. In the second period we observe the opposite pattern. This also can be viewed in the income share measures. The bottom five percentile has an income share of just under 1%, which decreases in the first period and subsequently increases. The top five percentile has an income share of 20% in 1994, which rises one percentage point and then declines to just over 20%. The next group of indicators measures the population share given a specific income range. For instance, we observe that 10% of households have income less than 40% of the median in 1994, and so on. Considering low and high income values as a percentage of the median, we observe that the population share grows in the first period and in the subsequent period it drops. However, if we consider income intervals near or around the median this trend reverses. This generates the perception of a decrease in the middle class during the 1994-2004 period, and an increase in the next period.

Using the M curve, this perception is confirmed. The middle class decreases by around 3% in the 1994-2004 period (the movement from the middle was both upward and downward), and then rises by 2 percentage points in the following period. When analyzing different population ranges around the middle, we also observe that a larger income spread is required to capture those ranges in 2004, reflecting greater income variation in the income distribution in that year. For example, given a population range between 20% and 80%, we require an income spread of 141% of the median income in 2004. This percentage falls by 8% in 2010.

10. The results are available from the authors upon request.

Table 3. Description of income distribution and middle class measures

	1994 (1)	2004 (2)	2010 (3)	2004[a] (4)	2010[b] (5)	Difference (2) - (1)	Difference (3) - (2)	Difference (5) - (4)
Centrality measures								
Mean	4,664.18	3,373.71	5,002.64	3,262.06	4,791.46	-1,290.47	1,628.91	1,529.40
Median	3,476.21	2,409.77	3,669.35	2,286.45	3,444.06	-1,066.44	1,259.58	1,157.61
Median/mean	74.53	71.43	73.35	70.09	71.88	-3.10	1.92	1.79
Quintiles (%)								
1st Quantile	5.61	4.22	5.17	4.14	4.88	-0.66	0.95	0.74
2sd Quantile	10.32	8.27	9.63	8.04	9.18	-0.85	1.36	1.14
3rd Quantile	14.99	13.03	14.34	12.86	14.04	-0.61	1.31	1.18
4th Quantile	21.96	20.69	21.50	20.43	21.40	-0.22	0.81	0.97
5th Quantile	47.12	53.80	49.37	54.54	50.49	2.34	-4.43	-4.05
Income share (%)								
Bottom 5%	0.78	0.71	0.81	0.71	0.77	-0.07	0.10	0.06
Bottom 10%	2.07	1.83	2.07	1.82	1.94	-0.24	0.24	0.12
Bottom 20%	5.60	4.94	5.50	4.85	5.16	-0.66	0.56	0.31
Top 20%	47.13	49.47	47.83	50.19	48.95	2.34	-1.64	-1.24
Top 10%	30.84	33.15	31.61	33.79	32.52	2.31	-1.54	-1.27
Top 5%	19.69	21.65	20.46	22.14	22.13	1.96	-1.19	-0.01

Table 3. (continued)

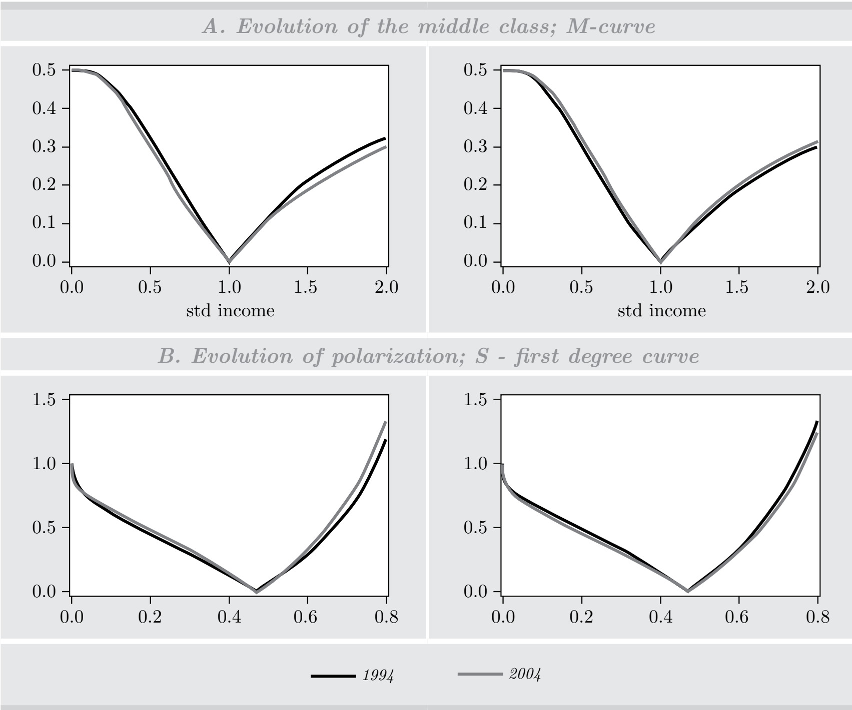
	1994 (1)	2004 (2)	2010 (3)	2004[a] (4)	2010[b] (5)	Difference (2) - (1)	Difference (3) - (2)	Difference (5) - (4)
% of Population with income:								
< 40% of median	10.38	12.68	10.71	12.81	11.90	2.30	-1.97	-0.91
< 50% of median	16.61	19.39	17.10	19.56	18.22	2.78	-2.29	-1.34
< 60% of median	23.41	26.20	24.14	26.38	25.35	2.79	-2.06	-1.03
60% to 75%	10.76	9.26	10.58	9.27	10.28	-1.50	1.32	1.01
75% to 100%	15.82	14.54	15.28	14.34	14.38	-1.28	0.74	0.04
100% to 125%	12.71	11.24	11.93	10.88	11.38	-1.47	0.69	0.50
125% to 150%	8.80	8.79	8.99	8.51	8.83	-0.01	0.20	0.32
> 200%	17.57	18.60	17.78	19.28	18.75	1.03	-0.82	-0.53
% in M-curve given income range								
75% to 150% of median	37.33	34.57	36.19	33.73	34.59	-2.76	1.62	0.86
75% to 125%	28.53	25.78	27.20	25.22	25.76	-2.75	1.42	0.54
50% to 150%	54.90	50.64	53.82	49.82	51.99	-4.26	3.18	2.17

Table 3. (continued)

	1994 (1)	2004 (2)	2010 (3)	2004[a] (4)	2010[b] (5)	Difference (2) - (1)	Difference (3) - (2)	Difference (5) - (4)
S given pop range:								
40% to 60%	35.58	39.92	37.23	40.87	39.26	4.34	-2.69	-1.61
35% to 65%	54.34	60.38	57.22	62.24	60.20	6.04	-3.16	-2.04
30% to 70%	75.30	83.80	79.08	86.08	82.57	8.50	-4.72	-3.51
25% to 75%	100.47	109.91	103.67	113.47	109.12	9.44	-6.24	-4.35
20% to 80%	131.25	141.30	133.23	144.78	140.30	10.05	-8.07	-4.48
Avg distance given pop range:								
40% to 60%	1.006	1.009	1.007	1.009	1.005	0.003	-0.002	-0.004
35% to 65%	1.011	1.019	1.014	1.020	1.013	0.008	-0.005	-0.007
30% to 70%	1.021	1.033	1.025	1.036	1.025	0.012	-0.008	-0.011
25% to 75%	1.036	1.051	1.042	1.055	1.044	0.015	-0.009	-0.011
20% to 80%	1.057	1.078	1.064	1.083	1.068	0.021	-0.014	-0.015
Observations	**18.386**	**18.392**	**40.539**	**18.392**	**40.539**			

Source: Authors' calculation based on the Uruguayan National Household Survey (ECH).
Note: Calculation based on the real per capita household income in March 1997 Uruguayan pesos, net of social security and income tax. Income data weighted using sample weights.
a household income without considering the old health system (OHS) income.
b household income without considering the new health system (NHS) income.

Figure 2. Middle class and polarization curves



Source: Authors' computations based on data from ECH.

All of these observed features are illustrated in Figure 2. In the top panels we plot the M-curve, which measures the concentration of mass around the median of the income distribution. We observe that the M-curve of the income distribution of 1994 is above the M-curve of the income distribution of 2004 (and they do not cross each other), and thus Proposition 1 holds: "The income distribution function in 1994 has an unambiguously larger middle class than the income distribution function in 2004." In other words, the 1994 income distribution has more mass around the median than the 2004 income distribution. Moreover, the first- and second-degree polarization curves (middle and lower panels) lead to the same conclusions as before. Those latter curves indicate that polarization in the income distribution in 2004 is higher than polarization in the income distribution in 1994, revealing that the latter has a greater income spread. Since a greater income spread implies a lower proportion of population around the middle, according to Proposition 3 the 1994 income distribution has a larger middle class than the 2004 income distribution and therefore, the former dominates the latter. Additionally, the second-degree polarization curve in 1994 is below the second-degree polarization curve in 2004, which implies that the income distribution in 2004 has a greater spread, as well as a greater bipolarity than the income distribution in 1994. The second period, 2004-2010, shows the opposite picture. The middle class increases while polarization tends to decline.

Table 4 contains inequality and polarization indices. The inequality indicators show a sharp increase between 1994 and 2004. For instance, the Gini index rises from 0.409 to 0.439. The generalized entropy index, the Atkinson index and the coefficient of variation index increase 0.054, 0.021 and 0.143 points, respectively. As mentioned above, this period is characterized by a tendency toward increasing inequality which is enhanced by the economic downturn that began in the late 1990s. This period of growing inequality is also accompanied by a significant rise in income polarization. The Duclos *et al.* index grows around 0.015 for different levels of identification represented by the parameter α. That is, for different values of α, the change in the Duclos *et al.* index between 1994 and 2004 is statistically different from zero at the 1% level. A greater value of α means that more emphasis is placed on the identification process. In order to analyze the contribution of each of the sources of polarization, the index can be decomposed into three (multiplicative) components: identification, alienation (which is equal to the Gini index) and correlation (between the two measures). It is interesting to note that while the alienation and correlation components evolve positively, the identification component declines. This result holds for different values of the α parameter.

In other words, polarization basically increases because the gap between the identified groups rises. For the second period, 2004-2010, the first main result is a decline in inequality. With the exception of the coefficient of variation index, the reduction is statistically different from zero. The second interesting result is that, as we have already noted, polarization falls. If we focus on the Duclos *et al.* (2004) index, the magnitude of the reduction decreases with the value of the α parameter. This can be explained by the fact that we give the greatest weight to the identification effect, which in this case goes in the opposite direction. Despite polarization declining slightly, the identification component rises but not enough to offset the reduction of the alienation effect.

The Foster and Wolfson polarization measure deserves a very similar reading. In the first period, we observe a statistically significant increase in the bipolarization index. In this case, we observe an

Table 4. Polarization and inequality measures

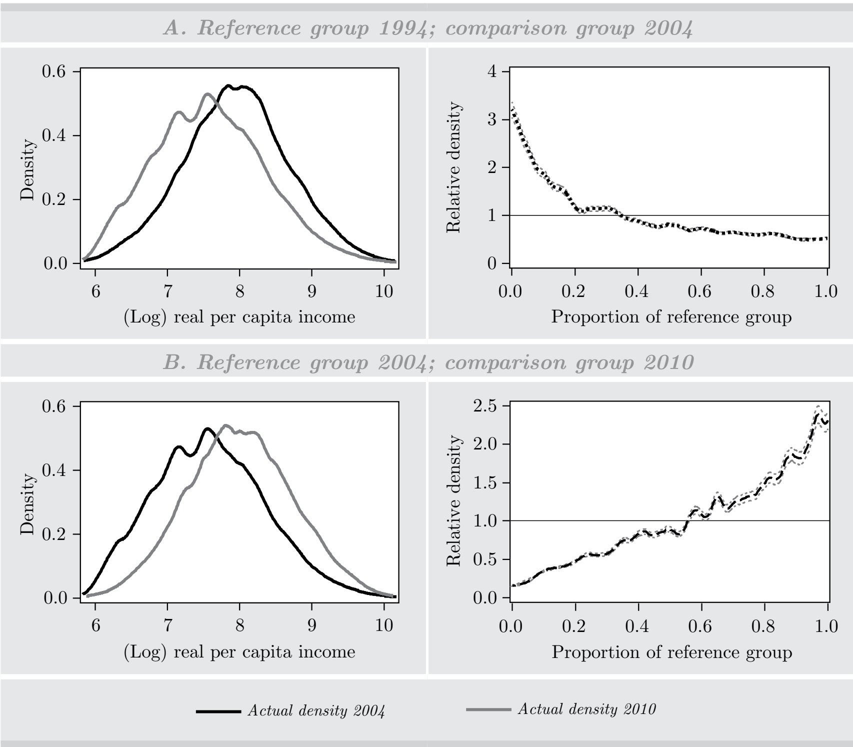
	1994 (1)	2004 (2)	2010 (3)	2004[a] (4)	2010[b] (5)	Difference (2) - (1)	Difference (3) - (2)	Difference (5) - (4)
Inequality								
Gini index	0.409 (0.003)	0.439 (0.003)	0.418 (0.002)	0.447 (0.003)	0.432 (0.002)	0.030***	-0.021***	-0.015***
Generalized entropy index	0.298 (0.005)	0.353 (0.008)	0.321 (0.011)	0.366 (0.008)	0.343 (0.007)	0.054***	-0.032***	-0.023**
Atkinson index	0.136 (0.002)	0.158 (0.002)	0.143 (0.002)	0.163 (0.003)	0.152 (0.002)	0.021***	-0.015***	-0.011***
Coefficient of variation index	0.934 (0.016)	1.077 (0.038)	1.049 (0.039)	1.106 (0.040)	1.092 (0.040)	0.143***	-0.028	-0.013
Polarization								
Duclos, Esteban and Ray (polarization) Index								
$\alpha = 0.25$	0.299 (0.001)	0.317 (0.002)	0.305 (0.001)	0.322 (0.002)	0.314 (0.001)	0.018***	-0.013***	-0.008***
Identification	0.834	0.817	0.831	0.818	0.829			
Correlation	0.877	0.884	0.878	0.881	0.877			
$\alpha = 0.50$	0.243 (0.001)	0.258 (0.001)	0.249 (0.001)	0.262 (0.001)	0.256 (0.001)	0.015***	-0.009***	-0.006***
Identification	0.725	0.702	0.723	0.704	0.720			
Correlation	0.819	0.837	0.824	0.833	0.823			
$\alpha = 0.75$	0.209 (0.001)	0.222 (0.001)	0.215 (0.001)	0.226 (0.002)	0.221 (0.001)	0.013***	-0.007***	-0.005***
Identification	0.646	0.621	0.645	0.625	0.642			
Correlation	0.791	0.814	0.797	0.809	0.797			

Table 4. (continued)

	1994 (1)	2004 (2)	2010 (3)	2004[a] (4)	2010[b] (5)	Difference (2) - (1)	Difference (3) - (2)	Difference (5) - (4)
Foster & Wolfson								
Relative median deviation	0.545	0.582	0.555	0.591	0.574	0.037***	-0.014***	-0.017***
Bipolarization Index	0.182	0.200	0.188	0.206	0.197	0.018***	-0.012***	-0.009***
Within gini index	0.137	0.148	0.140	0.151	0.145	0.011***	-0.008***	-0.006***
Between gini index	0.272	0.291	0.278	0.296	0.287	0.019***	-0.014***	-0.009***
Observations	18.386	18.392	40.539					

Source: Authors' calculation based on the Uruguayan National Household Survey (ECH).
Note: Calculation based on the real per capita household income in March 1997 Uruguayan pesos, net of social security and income tax. Income data weighted using sample weights.
a household income without considering the old health system (OHS) income.
b household income without considering the new health system (NHS) income.
* significant at 10%; ** significant at 5%; *** significant at 1%.

Figure 3. Actual and relative density



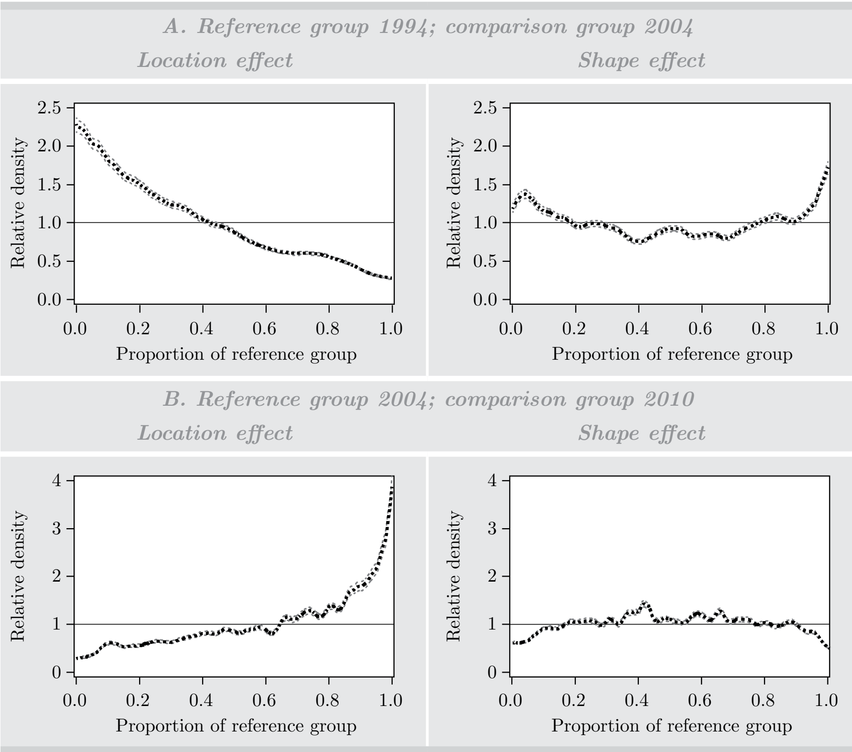
Source: Authors' computations based on data from ECH.

increase in inequality within and between the two groups.[11] Therefore, both groups spread out and the distance between them increases ("increased spread" and "increased bipolarity"). In the second period, the reduction in the within- and the between-Gini indices indicates a decreases in polarization.

We apply the relative distribution approach in order to find changes in the entire income distribution. Figure 3 shows the actual income distribution in 1994 and 2004 in the left plot and the relative distribution in the right plot of the top panel. At first glance, there is a shift from the right to the left which implies a reduction of mean income in this period. On the contrary, we observe a shift from the

11. As previously mentioned, Foster and Wolfson identify only two groups, those above and those below the median of the income distribution.

Figure 4. Location and shape effects



Source: Authors' computations based on data from ECH.

left to the right in the income distribution during the 2004-2010 period (see lower panel of Figure 3).

In Figure 4, we observe the location and the shape effect. The plots on the left confirm our prior observation since we find a decreasing and increasing location effect for the first and second periods, respectively. The right (top) plot shows how the lower and upper tail of the income distribution increase during the 1994-2004 period. This fact supports prior findings concerning a decline around the middle of the income distribution. In the other period, the shape effect shows that the lower and upper tail decline and the middle increases slightly. To formalize this result, and based on the relative density, we calculate relative polarization measures where positive values mean that polarization increases. In fact, we observe positive values that are statistically different from zero for the three measures in the first period. In the second period, the three indices are negative. This means that polarization decreases, which is in line with our previous findings. However, the change is smaller than in the first period.

To summarize, throughout the 1990s and until 2004 the income distribution becomes more unequally distributed and more polarized and the middle class shrinks considerably, while during the 2004-2010 period we observe some improvements.

4.3. Robustness analysis

In order to analyze the robustness of our results, we do not include as household income the health services derived from the new healthcare system (NHS) implemented in 2008.[12] In the new scheme, children under 18 years old of formal employees automatically acquired the right to medical services and therefore are not subject to the monthly payment.[13] The reform implies an important increase in the number of persons affiliated with private hospitals.[14]

The National Statistical Office of Uruguay (INE) accounts for this change by imputing a monthly payment for healthcare services to household income[15]. From a theoretical point of view, it is not clear whether this should include be included as income. If we do not impute this income, the results could change because the income distribution is sensitive to this imputation (mainly for low-income households). As we can see in Table 3 (fourth and fifth column) the proportion of income in the first and in the second quintiles decreases. What's more, the percentage of households around the median drops while the proportions at the extremes tend to increase. This is confirmed in the summary statistics related to the M-curve. In the previous section, we conclude that the middle class increases. However, if we do not include the imputation for health reform as household income (as well as the imputed income for healthcare services to wage earners in 2004), the change in the middle is ambiguous. This situation is illustrated by Figure 5, where the M-curve of the income distribution of 2004 is still

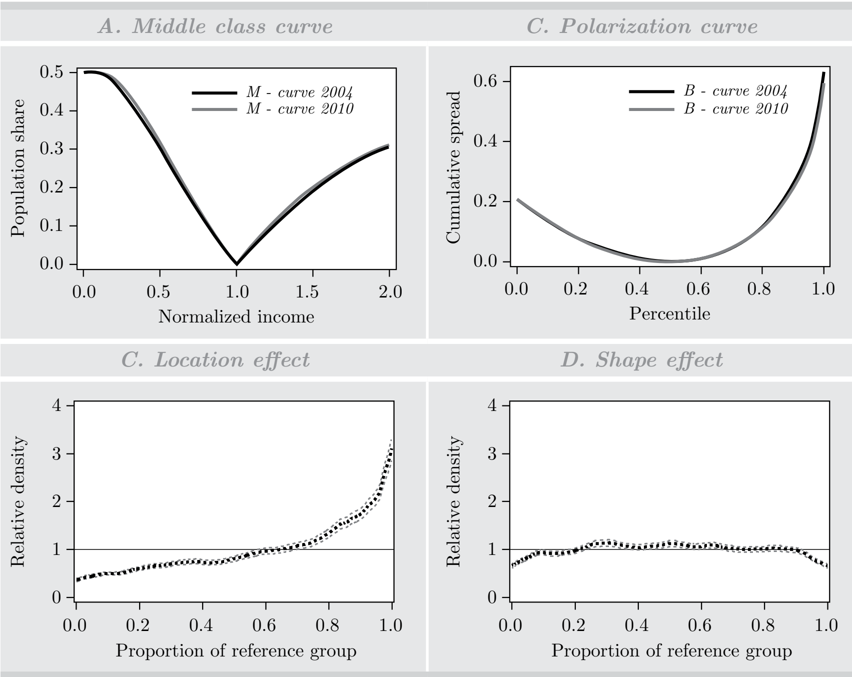
12. For a complete discussion of the 2008 health reform see Bérgolo and Cruces (2010).

13. This change was financed with an increase in worker contributions.

14. According to the Ministry of Public Health, the number of customers of collective healthcare institutions, which are the main private healthcare suppliers, increased by 314,976 between December 2007 and December 2008.

15. The INE also includes in 2004 household income the amount accounting for the health services for each household member who is a wage earner.

Figure 5. Robustness analysis



Source: Authors' computations based on data from ECH.

below the M-curve of the income distribution of 2010, but around the middle both curves are quite similar. This result implies that the middle class increases. Nevertheless, around the middle its increment is not as pronounced as observed in the previous section. The same is also evident in the shape effect panel, in which the extreme poles seem to decline while the middle increases, but to a lesser extent than in the previous section.

With respect to the polarization and inequality measures in Table 4, the various indicators decrease as before, but to a lesser extent. For instance, the Gini index decreases from 0.439 in 2004 to 0.418 in 2010 (0.021 points), while if we do not impute income for healthcare services the decline is 0.015. The same picture holds for the other indices where the changes are statistically different from zero, but with a lower change than in the original case. Furthermore, polarization grows for different values of the α parameter. In this case, the most important component of polarization is alienation since identification remains steady. Therefore, the higher the value of α, the lower the

Table 5. Relative polarization measures

	1994-2004 (1)	2004-2010 (2)	2004[a]-2010[b] (3)
Median relative polarization index	0.069*** (0.006)	-0.052*** (0.005)	-0.035*** (0.005)
Lower relative polarization index	0.076*** (0.010)	-0.058*** (0.009)	-0.041*** (0.009)
Upper relative polarization index	0.062*** (0.010)	-0.045*** (0.009)	-0.029* (0.009)

Source: Authors' calculation based on the Uruguayan National Household Survey (ECH).
Note: Calculation based on the real per capita household income in March 1997 Uruguayan pesos, net of social security and income tax. Income data weighted using sample weights. (1) Reference group 1994 and comparison group 2004; (2) reference group 2004 and comparison group 2010.
a household income without considering the old health system (OHS) income.
b household income without considering the new health system (NHS) income.
* significant at 10%; ** significant at 5%; *** significant at 1%.

effect. The conclusions are the same with respect to the Foster and Wolfson indexes and relative polarization measures (Tables 4 and 5).

5. Concluding remarks

In recent years there has been increasing concern about inequality and polarization. The expansion of the middle class is one of the key issues contributing to a reduction in inequality and less polarization. From an economic and social perspective, the middle class could play an important role in the development of a democratic country because it contributes a significant share of the labor force, and therefore is closely related to the country's output and usually represents the main source of tax revenue. Furthermore, an increase in the middle class resulting from reduction of the lower and upper classes could enhance the positive externalities mentioned above, decreasing income inequality and social tension.

We analyze the middle class and polarization in Uruguay over the last two decades. We conclude that the middle class decreases in size and income polarization increases between 1994 and 2004, while the opposite occurs between 2004 and 2010. However, when we do not include the income imputation due to the health reform implemented in 2008, the results tend to be attenuated. In other words, the expansion of the middle class between 2004 and 2010 is reduced and the magnitude of the decline is affected by the health income imputation, highlighting the importance of analyzing income imputation when using household surveys.

REFERENCES

Aghion, P., E. Caroli, and C. Garcia-Penalosa (1999), "Inequality and economic growth: The perspective of the new growth theories," *Journal of Economic Literature* **37**(4): 1615–60.

Amarante, V. and A. Vigorito (2006), "Evolución de la pobreza en el Uruguay," *Instituto Nacional de Estadística.*

Andersen, L.E. (2001), "Social mobility in Latin America: Links with adolescent schooling," Research Network Working Papers R-433, Inter-American Development Bank.

Azzimonti, M. (2011), "Barriers to investment in polarized societies," *American Economic Review* **101**(5): 2182–204.

Banerjee, A.V. and E. Duflo (2008), "What is middle-class about the middle classes around the world?" *Journal of Economic Perspectives* **22**(2): 3–28.

Barro, R. (1999), "Determinants of democracy," *Journal of Political Economy* **107**(S6): S158–S183.

Barriex, A. and J. Roca (2007) "Reforzando un pilar fiscal: El impuesto a la renta dual a la Uruguaya," *Revista de la CEPAL*: 123–42.

Bérgolo, M. and G. Cruces (2010), "Labor informality and the incentive effects of social security: Evidence from a health reform in Uruguay," CEDLAS, *Universidad Nacional de la Plata.*

Borraz, F. and N. González (2009), "The impact of the Uruguayan conditional cash transfer program," *Latin American Journal of Economics* **46**(134): 243–71.

Bourguignon, F. and C. Morrison (2002), "Inequality among world citizens: 1820-1992," *American Economic Review* **92**: 727–43.

Cruces, G., L.F. Lopez-Calva, and D. Battiston (2010), "Down and out or up and in? In search of Latin American´s elusive middle class," Research for Public Policy, Inclusive Development, ID-03-2010, RBLAC-UNDP.

Duclos, J., J. Esteban, and D. Ray (2004), "Polarization: concepts, measurements, Estimation," *Econometrica* **72**: 1737–72.

Easterly, W. (2001). "The middle class consensus and economic development," *Journal of Economic Growth* **6**(4): 317–35.

Esteban, J. and D. Ray (1994), "On the measurement of polarization," *Econometrica* **62**: 819–52.

Foster, J. and M. Wolfson (2009), "Polarization and the decline of the middle class: Canada and the U.S.," *Journal of Economic Inequality* **8**(2): 247–73.

Gasparini, L., E. Horenstein, E. Molina, and S. Olivieri (2008), "Income polarization in Latin America: Patterns and links with institutions and conflict," *Oxford Development Studies* **36**(4): 463–84.

Goldthorpe, J. (1980), *Social mobility and class structure in modern Britain.* Oxford: Clarendon Press.

Grau, C. and G. Lagomarsino (2002), "La estructura tributaria de Uruguay y su incidencia en la distribución del ingreso de los hogares," *Fundación de Cultura Universitaria.*

Handcock, M.S. and M. Morris (1998), "Relative distribution methods," *Sociological Methodology* **28**: 53–97.

Handcock, M.S. and M. Morris (1999), *Relative distribution methods in the social Sciences.* New York: Springer-Verlag.

Inter-American Development Bank (1998), "América Latina frente a la desigualdad," *Informe de Progreso Económica y Social.*

Loaiza, N., J. Rigolini, and G. Llorente (2012), "Do middle classes bring institutional reforms?" IZA Discussion Papers No. 6430.

López-Calva, L. and E. Ortiz-Juarez (2013), "A vulnerability approach to the definition of the middle class," *Journal of Economic Inequality*, January.

López-Calva, L., J. Rigolini, and F. Torche (2012), "Is there such thing as middle class values? Class differences, values and political orientations in Latin America." IZA Discussion Papers No. 6292.

Marshall, G. (1998) "Goldthorpe class scheme." *A Dictionary of Sociology Oxford Reference.*

Massari, M., M. Pittau, and R. Zelli (2009), "A dwindling middle class? Italian evidence in the 2000s," *Journal of Economic Inequality*, **7**(4): 333–50.

Lora, E. and J. Fajardo (2011), "Latin American middle classes: The distance between perception and reality," IADB Working Paper Series No. IDB-WP-275.

Rodríguez, S. and I. Perazzo (2007), "Impactos de la reforma tributaria sobre el ingreso de los hogares," Quantum, Instituto de Economía de la Facultad de Ciencias Económicas y de Administración, **II**(1): 61–78.