The Cyclicality of Enrollment in Tertiary Education

International Evidence from a Country-Pair Panel

Felix Schafmeister*

September 2016

* Student number: 434583fs. E-Mail: felix.schafmeister@student.eur.nl. Supervisor: Dr. Lorenzo Pozzi. Master's Thesis in International Economics.

Abstract

This paper aims at analyzing the cyclicality of enrollment in tertiary education by exploiting variation in student flows across countries. For this purpose, we investigate the relationship between the enrollment of foreign students and the business cycle in their country of origin in a country-pair panel dataset, spanning the years from 1999 to 2012. While we do not find a clear pattern of cyclicality for the sample as a whole, we argue that this is due to heterogeneity across countries and examine this possibility by allowing for different degrees of cyclicality in certain sub-samples. In line with previous results in the literature, we find that there is a significant difference in the cyclicality of tertiary enrollment between OECD and non-OECD countries. By splitting the sample with respect to the initial values of GDP per capita, years of schooling, and the index of institutionalized democracy, we are able to confirm the notion that more developed countries tend to have more countercyclical enrollment, even though our analysis is not suited to explain which specific factors are driving these differences. Moreover, we find that countries with a higher level of inequality tend to exhibit more procyclical enrollment, an insight that is in line with theoretical models of human capital accumulation based on credit constraints. Finally, we test for the importance of credit constraints by exploiting variation in both the real effective exchange rate and the ratio of domestic credit to the private sector over GDP. Our results confirm the relevance of credit constraints in shaping the pattern of cyclicality and are shown to be robust to the use of different data sources.

Table of Contents

1. Introduction

Until the beginning of the 21st century the macroeconomics literature was characterized by a striking asymmetry. While ample research had been conducted on the cyclical fluctuations of the stock of physical capital, the cyclical swings of the stock of human capital had not been subjected to such scrutiny (Dellas and Sakellaris 2003). However, there are numerous reasons for macroeconomic scholars and policymakers to be interested in the degree of cyclicality of human capital accumulation. Virtually all general equilibrium models aiming at an explanation of the business cycle – in particular real business cycle models in the spirit of Kydland and Prescott (1982) and Long and Plosser (1983) – are based on the assumption of intertemporal substitution in and out of labor as a main channel for the propagation of shocks. As empirical studies generally find the elasticity of intertemporal substitution between labor and leisure to be negligible (see i.a. Mankiw et al. 1985, Altonji 1986), substitution between labor and human capital accumulation activities might provide a more realistic alternative. Moreover, it is unclear whether cyclical fluctuations in the accumulation of human capital cancel out over the business cycle or have long-lasting effects. If the latter holds true, this could potentially serve as an explanation for the phenomena of hysteresis (Blanchard and Summers 1986) or 'super-hysteresis' (Ball 2014) that have recently regained attention in the literature (i.a. Stiglitz 2015).[1]

While both of the above-mentioned arguments refer to the same concept, distinct analyses are required to shed light on the implications for each of the two issues. The reason for this can be found in the composition of the stock of human capital. In his seminal contribution on the analysis of human capital, Becker (1975) postulates that the two main sources of human capital are formal schooling and on-the-job training. Since on-the-job training is unlikely to be an alternative to employment, it follows that the question of intertemporal substitution in and out of labor can be addressed by focusing exclusively on the cyclicality of schooling decisions. However, to resolve the second issue it is essential to consider cyclical movements in all components of the human capital stock. Due to the lack of reliable data for on-the-job training, this work aims at disentangling the direction of cyclical swings in schooling decisions and is

[1] The concept of hysteresis establishes a relationship between short-run fluctuations and long-run economic outcomes. Originally proposed by Blanchard and Summers (1986) as an explanation for the observed persistence of unemployment in Europe, it has since been considered in a broader sense as the relationship between business cycle fluctuations and potential output. While the original hysteresis theory regards fluctuations as persistent losses in the sense that the economy does not catch up to the level of output it would have attained in the absence of a recession, Ball (2014) proposes a more extreme version of hysteresis in which short-run fluctuations alter the economy's growth trajectory, thus leading to losses that increase over time.

therefore best suited to illuminate the importance of intertemporal substitution between labor and schooling.

For this purpose, section 2 reviews both the existing theoretical and empirical literature on the determinants and the cyclicality of human capital accumulation. Subsequently, in section 3, the empirical model is presented together with a detailed description of potential econometric problems that have to be overcome in the main analysis. In particular, we pay special attention to the issues of dynamic panel bias as well as instrumentation in a GMM setting. Section 4 gives an overview of the dataset used in the analysis and provides some first graphic evidence on the cyclical relationship. Building on the preceding sections, section 5 contains the empirical results. We start by analyzing the enrollment data in a univariate setting in order to determine which estimation approach is most readily suited for our purposes. We then proceed by applying our empirical model to the data and provide evidence on potential factors driving the observed divergence in patterns of cyclicality across countries. Furthermore, the section tests for the relevance of credit constraints and shows the robustness of the findings to the source of enrollment data. Section 6 concludes.

2. Literature overview

2.1 The determinants of human capital accumulation

As a starting point to the analysis of the cyclicality of human capital accumulation it is necessary to understand the underlying factors driving the decision of individuals to accumulate human capital in the form of schooling. There has been a broad literature in both development economics and macroeconomics attempting to uncover the determinants of schooling decisions from either a micro- or a macroeconomic standpoint. The theoretical literature on human capital accumulation was shaped by the early contributions of Mincer (1958), Becker (1962), and Ben-Porath (1967), who developed models formally introducing human capital accumulation into the household decision-making process. In essence, these models can be reduced to a trade-off between the costs and returns of education, whereby the costs consist mostly of opportunity costs due to foregone earnings and direct costs in the form of tuition fees and expenditures on textbooks, while the returns are due to increased life-time earnings (Card 1999, Mincer 1984). An important issue that was not appropriately addressed in these early theoretical contributions is the presumption that most schooling is undertaken early in life and hence, in the absence of access to perfect credit markets, often has to be financed by the parents. This fact is captured in the model by Loury (1981) who considers an overlapping generations (OLG) model in which parents are required to finance the education of their offspring. This adds another relevant dimension to the trade-off described above, namely the ability to pay for education, which will be shown to play a major role in the cyclicality considerations outlined below.

Aside from these theoretical papers, there also exists a large strand of literature that approaches the topic empirically. Kane (1994) examines the decrease and subsequent recovery in college enrollment of black students in the US during the 1980s. The author identifies changes in tuition fees, returns to education, and family background as the three main drivers of enrollment rates during this period. Specifically, increases in tuition fees were driving enrollment downwards while increases in the return to education and parental education levels applied upward pressure to enrollment rates. These findings are mirrored by Berger and Kostal (2002), who moreover show that not only demand-side but also supply-side factors, such as state appropriations to universities, help in explaining enrollment rates. Black and Sufi (2002) also focus on US enrollment rates and find differential reactions to tuition fees and returns to education depending on an individual's socio-economic status. Card and Lemieux (2001a) argue that while changes in family background across cohorts are important determinants of enrollment, they are not sufficient to explain observed patterns in US enrollment rates. Instead, they highlight the role of returns to education and cohort size as drivers of time-series variation in enrollment rates. This idea is corroborated by Abramitzky and Lavy (2014) who exploit a natural experiment to uncover a strong positive effect of the returns to schooling on enrollment.

Furthermore, some authors have analyzed a broader relationship between macroeconomic factors and human capital accumulation. Galor and Zeira (1993) develop an OLG model with liquidity constraints and intergenerational altruism. In their model, human capital is a decreasing function of wealth inequality since poor individuals might choose to enter the labor market as unskilled workers due to being unable to afford education. The same relationship is found in Chiu (1998) who extends the model by a heterogeneous distribution of talent, resulting in an even stronger adverse impact of wealth inequality on human capital accumulation as talented individuals from poor families do not receive an appropriate education. Implicit in these models is the idea that borrowing constraints are not only prevalent but also harmful with respect to the accumulation of human capital. Early evidence for this presumption is provided by Jacoby (1994) who shows that children from Peruvian families that were more likely to be subject to borrowing constraints dropped out of school earlier than their peers. De Gregorio (1996) develops an OLG model with endogenous growth and borrowing constraints and shows that human capital accumulation is a negative function of the degree of borrowing constraints faced by the households. He then proceeds to show that countries characterized by tighter borrowing constraints accumulate less human capital. Christou (2001) implements a similar model and concludes that borrowing constraints have a negative impact on human capital accumulation which is more profound when government spending on education is low.[2]

2.2 The cyclicality of human capital accumulation

2.2.1 Theory

From the examination of these underlying determinants it becomes clear that cyclical swings in the accumulation of human capital are likely to occur, since several of the determinants themselves fluctuate over the business cycle. This can best be illustrated by recalling the trade-off between expected returns and costs of the individual's schooling decision as well as the ability to pay for education. Regarding expected returns, Schady (2004) argues that the expected returns to education might decrease in a recession, if it is expected to be persistent. Bowlus and Liu (2003) formalize this idea and propose that entering the labor market during a recession results in long-lasting earnings losses because individuals gain less experience and earn lower wages in their entry-level jobs.[3] As a result, the recession might either depress expected lifetime earnings, rendering education less profitable and thus implying a drop in the optimal amount of schooling an individual will pursue, or might lead individuals to avoid entering the labor market in a recession by pursuing additional years of schooling.[4]

Considering the costs of education, there are several different factors that have to be taken into account. On the one hand, the opportunity costs of education can generally be viewed as procyclical since hours worked have been shown to be procyclical (Lilien and Hall 1986), while unemployment fluctuates countercyclically. In combination with the assumption that real wages are at least moderately procyclical (Bils 1985), this implies that individuals have a lower likelihood of finding employment in a recession and in case they do, they tend to work fewer hours and earn less. Moreover, in case of a procyclical real interest rate,[5] a simple life-cycle

---

[2] Furthermore, other researchers have uncovered relationships between human capital accumulation and labor mobility (Di Maria and Stryszowski 2009), technological change (Nelson and Phelps 1966, Foster and Rosenzweig 1996), as well as resource abundance (Gylfason 2001).

[3] Empirical support for this notion is obtained frequently in the literature (see i.a. Oreopoulos et al. (2012) for a study of Canada; Genda et al. (2010) for a study of Japan and the US; Bowlus and Liu (2003), Kondo (2007) and Hershbein (2012) for studies of the US).

[4] Bowlus and Liu (2003) find evidence for the latter mechanism, which they label "forced opportunity", as it forces individuals to deviate from their otherwise optimal decision not to pursue further education.

[5] So far there has been mixed evidence on the cyclicality of the real interest rate. Mishkin (1981) finds the real interest rate to be acyclical, a result which the author attributes to the low power of the econometric approach. Stock and Watson (1990) note that real interest rates are less cyclical than their strongly procyclical nominal counterparts and overall suggest a slightly countercyclical pattern. By contrast, Wilcox (1983) reports a drop in the real interest rate during the oil price shocks of the 1970s and therefore shows a procyclical pattern of the real interest rate.

model predicts that in booms there is a larger incentive to work and save for the future (Dellas and Koubi 2003). When combined, these opportunity cost considerations imply a countercyclical pattern for the accumulation of human capital. On the other hand, the direct costs of schooling are likely to follow a different pattern. Johnson (2013) argues that tuition fees might rise more strongly in recessions as governments reduce their expenditure on education.[6] This argument gains additional relevance when viewed in light of the cyclical swings in the ability to pay for education. Most importantly, Card and Lemieux (2001a) assert that liquidity constraints faced by individuals and their parents are tighter in recessions since parental income is reduced. Furthermore, financing programs in the form of student loans, aid, grants, and scholarships are often awarded in a procyclical manner (Dellas and Sakellaris 2003). Viewed in isolation, these notions therefore imply a procyclical pattern for the accumulation of human capital.

From the discussion above it becomes obvious that there are counteracting cyclical forces active in the determination of human capital accumulation. As there is no theoretical basis for arguing unequivocally which factor predominates, the direction of cyclicality remains to be determined by means of an empirical study.

2.2.2 Empirical evidence

There exists ample anecdotal evidence on the issue. A well-known stylized fact can be found in Goldin (1999), who shows in her analysis that there has been an unprecedented increase in secondary school enrollment in the US during the Great Depression. Card and Lemieux (2001b) provide evidence of intertemporal substitution between labor and schooling in the context of the Vietnam War. Moreover, Hotchkiss et al. (2012) argue that during the Great Recession there has been a marked increase in the share of individuals choosing not to participate in the labor force for reasons related to schooling. Finally, Riddell and Sweetman (1999) note that the steady increase in average years of schooling in Canada throughout the 1970s and 1980s is frequently attributed to worsening labor market conditions.

Several authors devote their attention to the analysis of cyclical swings in children's schooling in developing countries, obtaining mixed results. While Binder (1999) and McKenzie (2003) focus on the evolution of school enrollment rates in Mexico from 1977 to 1994 and from 1992 to 1996, respectively, and find a strongly countercyclical reaction, Schady (2004) analyzes

[6] Heyneman (1990) shows that developing countries frequently reduced their spending on education during recessions. However, Van Damme and Karkkainen (2011) provide evidence that most OECD countries held their government expenditures on education fairly constant throughout the Great Recession.

the impact of the Peruvian crisis of 1988 to 1992 and discovers only a marginally countercyclical enrollment pattern. On the other side of the spectrum, Jensen (2000) examines the impact of negative shocks to agricultural output on the investment in children in Côte d'Ivoire and reports a sharp drop in enrollment rates. In the same vein, Thomas et al. (2004) identify a reduction in school enrollments during the 1998 financial crisis in Indonesia.[7] The opposing results of these studies suggest a relationship between the pattern of cyclicality and a country's development status. If ability to pay considerations are assumed to become more important the less developed a country is, these findings appear fully in line with the theoretical arguments outlined above.

This presumption is also supported by the wide array of studies investigating the cyclicality of enrollment in tertiary education in the US. Early work on the determinants of enrollment rates in the US sometimes includes implicit tests for the cyclicality of enrollment by introducing the local unemployment rate into the regression framework. While Kane (1994) and Card and Lemieux (2001a) find a strong positive impact of the unemployment rate on enrollment (i.e. countercyclical enrollment), Black and Sufi (2002) report that the unemployment rate enters with a significant positive coefficient only for Whites, while it has no impact for black students. Moreover, in the empirical model of Berger and Kostal (2002), the unemployment rate is unrelated to changes in enrollment, which might be due to the fact that the authors base their study on data from the 1990s, a time that was characterized by comparative economic tranquility in the US.

Since then, several authors have performed more rigorous investigations of the cyclicality of US enrollment rates, focusing on micro-level data of different subgroups and time periods. Betts and McFarland (1995) consider enrollment in two-year programs at community colleges between the 1960s and 1980s and detect a strong countercyclical pattern of enrollment. Their setting is particularly remarkable as individuals enrolling in two-year programs at community colleges are already in the labor force but have low savings and are largely excluded from parental assets and could therefore be expected to have limited access to credit. Closely related are the studies by Dellas and Sakellaris (2003) and Dellas and Koubi (2003) which are based on broader samples of US undergraduates. Both find a strongly countercyclical enrollment pattern and additionally identify the real interest rate to have a negative impact on enrollment, providing suggestive evidence that ability to pay considerations also play a role in the US

[7] See Ferreira and Schady (2009) for a broad survey of studies on the cyclicality of children's schooling in developing countries.

setting. Christian (2007) attempts to directly test for the importance of liquidity constraints by splitting his sample of recent high-school graduates with respect to the likelihood of being credit constrained, as indicated by parental wealth, and finds that individuals more likely to be subject to liquidity constraints entail a more procyclical enrollment pattern. Other studies have investigated the cyclicality of graduate education in the US and arrived at rather mixed results. While Johnson (2013) studies graduate students in all fields and finds significant countercyclical enrollment only for females, Bedard and Herman (2008) limit their analysis to MINT graduates and report a countercyclical pattern for male graduates but not for female graduates.[8]

Finally, some authors have taken a broader perspective to the investigation of the cyclicality of human capital accumulation. Skidmore and Toya (2002) provide evidence that climatic disasters have a positive impact on human capital accumulation in a cross-country panel regression, where human capital accumulation is measured by changes in secondary school enrollment or average years of schooling. In a related study, Heylen and Pozzi (2007) also consider a large country panel data set and find a positive effect of inflation crises on human capital accumulation as measured by changes in average years of schooling. The study most closely related to the analysis in the work at hand is Sakellaris and Spilimbergo (2000). They adopt a unique strategy to examine the cyclicality of human capital accumulation by estimating the impact of the business cycle in the country of origin on the enrollment of foreign students in the US. Using this approach, the authors are able to paint a more nuanced picture of the cyclicality of human capital accumulation across the world. By splitting the sample along several dimensions, most notably related to OECD membership status, the authors are able to show that while student flows from OECD countries to the US reflect the commonly observed countercyclical pattern, flows from non-OECD countries are generally procyclical.

In what follows, we build on this literature by considering a dynamic panel model in which we relate the enrollment of foreign students in selected OECD countries to the business cycle in student's home countries. In doing so, we make use of a country-pair panel dataset with a large number of individuals and a moderate time-series dimension and employ state-of-the-art econometric methods to overcome several common methodological problems outlined in the following chapter.

8 Johnson (2013) argues that cyclical swings in graduate enrollment are more muted because most universities tend to enroll students at capacity, irrespective of economic conditions. This serves as an explanation for the tendency of graduate enrollment to display a rather acyclical pattern.

3. Methodology

In principle we would be interested in estimating a simple regression equation relating enrollment rates of foreign students to contemporaneous as well as lagged real GDP per capita growth (henceforth GDP growth) in the country of origin. However, this analysis is complicated by the nature of the enrollment data which are aggregated over several cohorts of students and are hence likely to exhibit a considerable degree of persistence. Combined with unobserved heterogeneity across country-pairs in the sample, this warrants the use of a dynamic panel approach.

In its most general form, the model of interest is

$$E_{ijt} = \alpha E_{ij,t-1} + \sum_{k=0}^{p} \gamma_k g_{j,t-k} + \boldsymbol{X_{jt}\beta} + \mu_{ij} + \lambda_t + \varepsilon_{ijt}, \qquad (1)$$

where E and g denote enrollment and GDP growth, respectively.[9] The matrix $\boldsymbol{X_{jt}}$ contains additional exogenous control variables for the sending country, while μ_{ij} is a country-pair fixed effect, λ_t is a time fixed-effect and ε_{ijt} is an idiosyncratic error term, assumed to be independent across country pairs.[10]

3.1 Dynamic panel bias

It is well known in the literature that estimation of equation (1) by ordinary least squares (OLS) or within-groups (WG) estimation yields inconsistent parameter estimates because the lagged dependent variable is correlated with the error term.[11] As shown in Appendix A.1, due to the presence of the fixed effect μ_{ij} there is a positive correlation between the lagged dependent variable and the error term, which implies an upward bias in the OLS estimate. The same line of reasoning also implies a negative correlation between the transformed lagged dependent variable and the transformed error term in the WG estimator, thus suggesting a downward bias. The issue has been studied more thoroughly by Nerlove (1971) and Nickell (1981), who provide Monte Carlo simulations and analytical evidence on the nature of the dynamic panel bias. In particular, Nickell (1981) shows that the bias in the WG estimator is of order in probability O(T^{-1}) and hence approaches zero only when the time dimension grows

9 The subscripts i, j, and t refer to the receiving country, the sending country, and the time dimension, respectively, and take the values $i = 1, 2, \ldots, N_1$, $j = 1, 2, \ldots, N_2$, $t = 1, 2, \ldots, T$.

10 Independence of the idiosyncratic error term across individuals is a crucial assumption for the consistency of the estimation techniques outlined below. Allowing for time fixed-effects helps in alleviating potential concerns about the validity of this assumption.

11 For simplicity, the exposition in this chapter focuses on solving endogeneity of the lagged dependent variable, even though the coefficient on this variable is not the focus of our analysis. As, among others, Blundell and Bond (2000) show, the insights in this section are easily generalized to the analysis of an endogenous regressor that depends on a time-invariant fixed effect.

large. Using Monte Carlo simulations, Judson and Owen (1999) show that the dynamic panel bias can still be considerably large – up to 20% – in panels with a time dimension of 30.

3.2 Overcoming dynamic panel bias

Several approaches have been advanced to solve the issue of dynamic panel bias. In early work, Anderson and Hsiao (1982) (hereinafter AH) have proposed applying first differences to the model of interest to cancel out the unobserved heterogeneity in form of the fixed effect and subsequently estimating the first differenced model by two-stage least squares (2SLS), instrumenting the lagged, first-differenced dependent variable by twice lagged levels or first-differences of itself. While this estimator is consistent as long as there is no serial correlation in the untransformed error terms, it has been shown to exhibit undesirable characteristics such as a singularity point (Arellano 1989) and generally large variances (Arellano and Bond 1991). In response to these weaknesses, Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998) have built on the AH methodology and proposed more encompassing solutions to the dynamic panel bias based on the Generalized Method of Moments (GMM).[12]

The so-called difference-GMM estimator developed by Arellano and Bond (1991) also considers the first-differenced model but instead of using only the twice lagged level as an instrument, the estimator exploits all $\frac{(T-2)(T-1)}{2}$ available moment restrictions implied by the sequential exogeneity of the dependent variable and absence of serial correlation in the error terms.[13] This becomes possible by making use of an instrument matrix of the form first

[12] Due to the focus of this study being on the empirical analysis, we refrain from an in-depth description of the mechanism underlying GMM estimation. In general, the standard GMM estimator minimizes the quadratic form of the moment conditions and is given by $\hat{\beta}_{GMM} = (X'ZAZ'X)^{-1}X'ZAZ'Y$, where Z is the instrument matrix and A is a weighting matrix. While consistency of the GMM estimator does not depend on the specific choice of A, efficiency hinges crucially on the choice of an optimal weighting matrix that weighs moments inversely to their respective variances. A GMM estimator can usually be obtained by either a one-step or a two-step procedure – while the one-step procedure simply assumes a weighting matrix with favorable properties under certain assumptions, the two-step estimator uses the estimated residuals from a first-stage regression to construct the weighting matrix and then repeats the estimation (see e.g. Cameron and Trivedi 2005). The interested reader is referred to the seminal work of Hansen (1982) or the overview article of Wooldridge (2001), both of which provide an excellent treatment of GMM estimation.

[13] The moment conditions take the form $E(y_{i,t-j}\Delta\varepsilon_{it}) = 0, \forall\, t \geq 3, j \geq 2$ and are valid as long as the error term ε_{it} is not serially correlated and y_{it} is sequentially exogenous (Wooldridge 2010). It should be noted that the first-differenced error term is by definition serially correlated and hence testing for second order serial correlation in the error term is required. If the untransformed error terms are serially correlated, fewer valid moment restrictions are available as this induces a correlation between the twice-lagged dependent variable and the error term. In general, we have to distinguish between two forms of serial correlation, namely autoregressive and moving average processes. While autoregressive error terms would render all lags invalid instruments, the presence of a moving average process only results in some lags being invalid as the correlation between the error terms rapidly converges to zero. As a result, the standard procedures to dealing with both problems differ greatly. Whereas autoregressive error terms require the inclusion of additional lags of the dependent variable, a moving average process can be overcome by using deeper lags as instruments.

considered by Holtz-Eakin, Newey, and Rosen (1988) (henceforth HENR) that allows to overcome the trade-off between lag length and sample depth usually found in standard 2SLS regressions (Roodman 2009).[14] By using the additional information, the difference GMM estimator is more efficient than the AH estimator and has been shown to be more precise in Monte Carlo simulations (see i.a. Arellano and Bond 1991, Kiviet 1995, and Judson and Owen 1999). However, both the AH and the difference GMM estimators can be expected to suffer from weak instrument bias when the dependent variable is highly persistent because the lagged levels are less informative for subsequent changes (Arellano and Bover 1995).[15] In the extreme case of a random walk process, past levels contain no information about the change in the variable of interest, thus rendering the difference GMM estimator unidentified.

These shortcomings have led Arellano and Bover (1995) and Blundell and Bond (1998) to propose the use of additional moment restrictions, resulting in the so-called system-GMM estimator. This estimator builds a stacked system of equations in which the moment conditions of the first-differenced model are combined with an additional set of moment conditions for the equation in levels which are obtained by using multiply lagged first differences as instruments for the dependent variable in levels. While the use of this estimator requires non-trivial additional assumptions on the data generating process (DGP), it has the advantage that lagged first-differences retain their informational content about subsequent levels even when the process is highly persistent (Bond 2002).[16] Blundell et al. (2001) show that the system-GMM estimator is a weighted average of difference- and levels-GMM, whereby the weight attached to the levels estimator increases in the weakness of the first-differenced instruments.

[14] While a standard 2SLS instrument matrix allows for the use of different instruments for each variable, the instrument matrix proposed by HENR enables the use of separate instruments for each variable and time-period by constructing a block-diagonal instrument matrix in which the instruments related to each time-period are on the (block-)diagonal (Bazzi and Clemens 2013). In standard 2SLS the use of additional lags as instruments would therefore diminish the sample size as early time-periods have to be omitted due to the lack of lagged values. A brief comparison of the two approaches is outlined in Appendix A.2.

[15] Blundell et al. (2001) moreover show that the instruments are likely to be weak when the ratio of the respective variances of the individual effects and the error terms is large.

[16] The additional moment conditions take the form $E\left(\Delta y_{i,t-1}\varepsilon_{it}\right) = 0 \; \forall \; t \geq 3$ (Roodman 2009). Blundell and Bond (1998) show that validity of the additional moment conditions requires that changes in the variable of interest are uncorrelated with the fixed effects. This condition is always fulfilled when the series is stationary and has converged to its long-run equilibrium. Even if the series exhibits a time-varying mean, the initial conditions can be fulfilled when enough time has passed so that the influence of the initial conditions is small (Blundell et al. 2001).

3.3 Threats to identification

3.3.1 Invalid instruments

In practice, there are several potential reasons for GMM instrumentation to break down. Most obviously, this is the case when the assumed moment restrictions do not actually hold. As outlined in footnotes 11 and 14, respectively, validity of the difference-GMM and system-GMM instruments requires specific assumptions of the DGP which are in principle testable by means of a test originally proposed by Sargan (1958) and further refined by Hansen (1982) (henceforth Hansen test). The test statistic directly follows from the minimized moment criterion in the GMM estimation and thus tests whether the minimization brings the moment conditions reasonably close to zero. Under the alternative hypothesis, the moment restrictions do not hold, indicating that the chosen instrument set is invalid. The additional instruments included in the system GMM estimator can be tested by calculating a so-called 'difference in Hansen' statistic because the instrument set used in difference-GMM is nested in the full system-GMM instrument set (Bond 2002). Moreover, since instrument validity depends on characteristics of the DGP, the underlying assumptions can be substantiated by analyzing certain features of the data. One of the most well-known tests of this sort has been proposed by Arellano and Bond (1991). Specifically, it tests for serial correlation in the first-differenced error terms which would invalidate the use of certain lags as instruments.[17] Furthermore, for the assumption of sequential exogeneity to be fulfilled, it is necessary that the dynamic process is correctly specified. Hence the appropriate number of lags as well as the stationarity properties of all underlying series should be tested.[18] The examination of the individual data series also helps in determining the degree of persistence of the variables in the model, which, as mentioned above, is an important indicator of instrument strength.

3.3.2 Weak instruments

Another problem that has received considerable attention in the literature on GMM estimation is that of weak instruments. Arellano and Bover (1995) and Alonso-Borrego and

---

[17] Cf. footnote 11.

[18] Testing for unit roots in dynamic panel models of the sort we are investigating in this paper is complicated by the fact that the vast majority of panel unit root tests is only valid under large T asymptotics (Binder et al. 2005). Harris and Tzavalis (1999) propose a variation of the Levin et al. (2002) unit root test requiring only N going to infinity for the validity of its asymptotic distribution. A drawback of this test is that its assumptions of independent and identically distributed error terms and homogeneity are quite restrictive. A simple alternative is brought forth by Blundell and Bond (2000) who exploit the fact that the OLS estimator of the lagged dependent variable is upward biased and hence argue that rejection of the unit root hypothesis on the basis of the OLS estimate is sufficient in many cases.

Arellano (1999) note that the GMM estimator entails a finite-sample bias that increases in the weakness of the instruments.[19] Unfortunately, there is no direct test available for instrument strength in dynamic panel data models, which led several authors to consider heuristic techniques. A first criterion follows from the known biases of both OLS and WG estimation which, due to the opposite directions of their respective biases, act as bounds on the true parameter value. Therefore, a reasonable parameter estimate for the lagged dependent variable should always lie in between the two estimates obtained by OLS and WG (Bond 2002). Moreover, recalling that the bias due to weak instruments is in the direction of the WG estimate, a GMM estimate very close to the one obtained by WG estimation might be indicative of instrument weakness. Going beyond these rough indicators of instrument strength, Bazzi and Clemens (2013) propose estimating a 2SLS first-stage, using the GMM instrument matrix, and applying the standard 2SLS weak instrument diagnostics developed by Stock and Yogo (2005). In case the instruments are found to be weak, there are two main remedies available. First, the instrument set can be altered by exploiting different moment conditions that are more directly related to the variable of interest, e.g. by switching from difference GMM to system GMM. Second, Bun and Windmeijer (2010) suggest the use of the Hansen et al. (1996) continuous updated GMM estimator (CUE) supplemented with a weak-instrument variance correction due to Newey and Windmeijer (2009).[20]

3.3.3 Too many instruments

A closely related issue is due to the amount of instruments included in the regression. As already noted above, in the full difference GMM estimator, the number of instruments is a quadratic function of the time dimension. Accordingly, in particular in samples with a moderate time dimension and more than one endogenous variable, the number of instruments can become considerably large. Roodman (2009) argues that using a large number of instruments results in overfitting the endogenous variables which causes instrumentation to break down because the endogeneity is not removed from the model. Moreover, the author shows that estimation of the

[19] Blundell and Bond (2000) argue that this finite sample bias is likely to be in the direction of the within estimator, i.e. downwards. This notion is supported in the Monte Carlo simulation results of Blundell et al. (2001) and Hauk and Wacziarg (2009).

[20] The CUE is a GMM variation that views the weighting matrix as a function of the parameter vector and thus does not take the weighting matrix as given when minimizing the quadratic moment criterion, but instead allows it to vary continuously when the estimated parameter vector is altered in the minimization (Hansen et al. 1996). This estimator has the advantage of being invariant to the scaling of the moment conditions and has been shown to entail a smaller bias than other GMM estimators when instruments are weak (Newey and Windmeijer 2009). However, similar to the standard two-step GMM estimator, the usual standard errors are found to be too small, leading standard tests of coefficient significance to overreject the null hypothesis. This issue can be alleviated by making use of the variance correction suggested in Newey and Windmeijer (2009).

optimal weighting matrix in two-step GMM becomes exceedingly imprecise when the number of instruments increases. Many authors have shown in Monte Carlo simulations that the number of instruments induces a trade-off between bias and efficiency of the GMM estimators, with the bias increasing and the standard deviation decreasing in the number of instruments (see i.a. Judson and Owen 1996, Windmeijer 2005, Everaert and Pozzi 2007, and Roodman 2009). An even more adverse consequence of using a large number of instruments is the loss in power of the Hansen test of instrument validity. Bowsher (2002) shows that the test becomes excessively undersized in panels with a time dimension of 11 or larger, with the size rapidly approaching zero. As a result, the unadjusted power of the test goes to zero, rendering it ineffective. This deterioration in power is especially detrimental since it is likely to result in a strong false negative result, namely that the test is unable to reject the hypothesis of instrument validity at a 'perfect' p-value of 1, even if it is false (Roodman 2009). The most obvious solution to this problem is to reduce the number of instruments used in the regression, either by limiting the number of lags used as instruments or by 'collapsing' the HENR instrument matrix which also results in a reduction of the column count (Roodman 2006). While this might cause a further weakening of the instruments and thus might not be appropriate for making inference about the regressors in the model, at least the exclusion restriction should be tested under such a reduced instrument set to overcome the weakness of the testing procedure (Roodman 2009).

4. Data description

Our main variable of interest is the total number of students enrolled in tertiary education in a foreign country, disaggregated at the country of origin. Data of this kind is published by three agencies, namely the UNESCO, the OECD, and the US Institute for International Education (IIE), whereby the data vary in terms of availability over time and the definition of a student enrolled in a foreign country.[21] Due to the sketchy availability of the UNESCO data and the sole focus on the US by the IIE, we consider the OECD data series as our main variable of interest. This implies that we limit our attention to students enrolled in OECD countries which has the advantage of alleviating concerns due to data quality and the comparability of the provided education. We make several adjustments to the data (described in Appendix B.1), resulting in a balanced panel of 948 country pairs, unevenly distributed across 163 sender and 19 recipient countries and observed on a yearly basis from 1999 to 2012.

[21] Detailed data sources and definitions are given in Appendix A.1.

Table 1 presents summary statistics for the distribution of country pairs across recipient and sending countries. On average, every recipient country is part of almost 50 country pairs with a standard deviation of 38. The recipient countries making up the largest share of the total country pairs are the US with 146 pairs, the UK with 112 pairs, and Germany with 107 pairs. On the other side of the spectrum, there are only five country pairs with South Korea as the recipient country, which is due to the fact that enrollment of foreign students in South Korea was quite low in the beginning of the sample period and only recently reached higher levels. As a result, most series based on South Korea as the country of destination do not fulfill our selection criteria. Of the 163 sending countries, the average country is paired with six recipient countries with a standard deviation of 4. China and Russia both are part of 18 country pairs, while some small African countries such as Comoros and Djibouti are only paired with a single recipient country.

In order to obtain a general overview of the development of international student flows over time, figure 1 depicts the number of outbound students disaggregated by region.[22] It can be seen that the total number of internationally mobile students approximately doubled from 1998 to 2013, reaching a value of about 3.5 million. Almost two thirds of this increase was driven by the surge in the number of outbound students from Asia, which grew by over 150% from around 750,000 in 1998 to almost two million in 2013. This increase, in turn, reflects the economic rise of the world's two most populous nations, China and India, which account for the majority of outbound students from Asia. Notably, the lowest growth rates of the number of outbound students are found for the predominantly highly developed regions of Oceania, Europe, and North America that exhibit growth rates of around 60%. While figure 1 describes student flows from the perspective of the sending country, figure 2 features the development of the number of inbound students in the five most important destination countries in our sample, namely the US, the United Kingdom, Germany, France, and Japan.[23] It becomes obvious that the US were by far the most important destination country throughout the sample period, although recently the gap to the United Kingdom has narrowed. Taken together, the US and the United Kingdom account for slightly over 50% of the students captured in the sample.

[22] This data is based on the UNESCO definition of internationally mobile students and therefore does not necessarily reflect the same flows of students as the data used in the main analysis.

[23] It should be noted that figure 2 no longer includes all countries of the world but only those that fulfill our sample selection criteria. The trends and proportions depicted might therefore not be perfectly representative as they are in part driven by the unavailability of complete data series for some countries and the exclusion of non-OECD destination countries.

In the following, we take a first look at the cyclical enrollment patterns in the data. For this purpose, figure 3 plots the growth rates of the total number of outbound students over all destination countries in our sample and the growth rate of GDP per capita over time for four selected countries, in particular Argentina, Italy, Slovenia, and Trinidad and Tobago. Even though the time dimension is too short to draw robust conclusions, for Argentina and Italy there appears to be a strikingly close negative connection between the growth rate of GDP and the growth rate of the number of outbound students, especially when lags of the GDP growth series are considered. For both countries, the correlation coefficient between the growth rate of the number of outbound students and the once-lagged growth rate of GDP is strongly negative at -0.67 and -0.7, respectively. Most notably, both the Argentinian 'great depression' (Kehoe 2003) of the early 2000s, as well as the drop in Italian GDP per capita during the Great Recession were followed by remarkably large increases in the number of outbound students. The remaining panels of figure 3 paint a much more nuanced picture for the cyclicality in Slovenia and Trinidad and Tobago. For both countries, there is a slight positive relationship between the two series, with a correlation of 0.18 and 0.52, respectively. Considering Slovenia, the correlation between the two series is even higher when twice-lagged GDP growth is considered, in which case the correlation coefficient takes a value of 0.53. From figure 3, we can therefore conclude that while cyclical movements appear to be a relevant feature of the data, the direction of cyclicality might be heterogeneous and remains to investigated in a more thorough analysis.

5. Empirical analysis

5.1 Univariate regressions

As a first step, we investigate the properties of the enrollment series in order to determine which estimation approach is most appropriate for the cyclicality analysis. For this purpose, table 2 presents OLS and WG estimates of a univariate regression of the natural logarithm of enrollment on up to three of its own lags. It becomes obvious that the series is highly persistent with the OLS and WG estimates of the AR(1) model indicating a value of α of 0.988 and 0.832, respectively. However, it should be noted that the null hypothesis of a unit root can be rejected even when the upward biased OLS estimate is considered. The high persistence of the series and concomitant significance of the lagged dependent variable warrants the use of GMM techniques as outlined in chapter 3. In table 3, we re-estimate the univariate model using both difference- and system GMM and including up to two lags of the dependent variable. While the estimates obtained by difference GMM lie within the bounds given by the OLS and WG estimates of table 2, we find that the system GMM estimates are too large and even exceed unity. In all regressions reported in table 3, the Hansen J-test overwhelmingly rejects the null hypothesis of instrument validity with a p-value of 0.0000, while the Arellano-Bond test of serially autocorrelated error terms rejects even the null hypothesis of no autocorrelation of order 6 at the 5% level. The standard approach to alleviating problems due to autocorrelated error terms of including additional lags of the dependent variable into the regression framework therefore does not appear to be a valid solution as this would rapidly deplete the sample. Instead, we will proceed by considering the growth rate of enrollment as our dependent variable, which can be expected to entail a much lower degree of persistence.

Therefore, in table 4, we repeat the above analysis and estimate the persistence of the growth rate of enrollment in a univariate framework using OLS and WG estimation. The estimated coefficients are strikingly different, as all OLS estimates are essentially equal to zero while the WG estimates even imply a small negative effect. Considering these estimates to be bounds on the true coefficient, it becomes obvious that the lagged dependent variable has at most a modest impact on the dependent variable and might not even need to be part of the regression framework. If this was the case, this would drastically simplify the main analysis since the WG estimator is consistent in the absence of a lagged dependent variable.

In order to confirm whether a lagged dependent variable should be part of the empirical framework, in table 5, we estimate the univariate model by GMM to obtain consistent estimates of the coefficients on the lagged dependent variable. Columns 1 and 2 present difference- and system-GMM estimates, respectively, using the full instrument set, which both yield estimated coefficients that are essentially equal to 0 and insignificant. Although we find the instruments to be strong, with a first stage Cragg-Donald (1993) F-statistic of 74.80, the Hansen J-test again rejects the validity of the instrument set, even though in this case the Arellano-Bond test does not indicate any serial autocorrelation beyond order 1 in the error terms.[24] A potential explanation for the rejection of the Hansen J-test can be found in the failure of the assumption of independence of the error terms across individuals. This is likely to occur in our sample as the same source and destination countries appear multiple times in different country pairs. This has two quite distinct effects on the GMM estimation that both might cause the Hansen test to reject. First, the weighting matrix is estimated without allowing for correlation across individuals, leading to an inappropriate weighting of the moment conditions and thus causing

[24] This rejection also appears not to be caused by serial correlation of the moving average form in the error term, as the Hansen test still rejects when only deeper lags are used as instruments (results not reported).

the Hansen test statistic to be excessively large. Second, as mentioned in chapter 3, correlation of the error terms across individuals likely also renders the moment conditions used in both the difference- and system-GMM estimators invalid as these are derived under the assumption of independence across individuals.

We attempt to test whether the moment restrictions used in columns 1 and 2 are actually invalid by allowing for clustering of the error terms at the level of the recipient country. Columns 3 and 4 provide difference- and system-GMM estimates, respectively, allowing for such clustering. For both regressions, the Hansen J-test yields a p-value of 1, indicating that the test might be weakened by the presence of an excessive amount of instruments in the regression and should therefore not be relied upon. In columns 5 and 6, we attempt to overcome this weakness by making use of a collapsed instrument set in order to reduce the number of instruments used in the regression while retaining most of the information of the original instrument set. The resulting p-values of the Hansen J-test are 0.06 and 0.25, respectively, which, notwithstanding the lack of rejection at conventional levels in the system-GMM estimation, call the appropriateness of the instrumentation strategy into question. Although this exercise is far from conclusive, we take the fact that none of the estimates in table 5 are significantly different from zero as indication that a lagged dependent variable need not necessarily be part of the empirical model.[25]

5.2 Cyclicality analysis

Building on the insights derived in the preceding section, we estimate a simplified version of equation (1), in which we replace the dependent variable by the growth rate of enrollment and omit the lagged dependent variable to obtain

$$e_{ijt} = \sum_{k=0}^{p} \gamma_k g_{j,t-k} + \boldsymbol{X_{jt}\beta} + \mu_{ij} + \lambda_t + \varepsilon_{ijt}, \qquad (2)$$

where e denotes the growth rate of enrollment. For the main analysis, we set $\boldsymbol{X_{jt}}$ to zero, i.e. we refrain from including additional regressors into the model as these might absorb some of the cyclicality and render the interpretation of the coefficients less straightforward.

Table 6 presents the first set of results in which the growth rate of enrollment is regressed on contemporaneous and up to 5 lags of GDP growth as well as year dummies, using the WG estimator. Column 6 shows our preferred specification in which all 5 lags are included into the regression. The resulting pattern appears acyclical as the majority of estimated coefficients are

[25] The appropriateness of this assumption will further be addressed in section 5.3.

insignificant and very close to zero. Moreover, the F-test testing for the sum of the coefficients on all included lags of GDP growth being equal to zero is generally insignificant.

5.2.1 Heterogeneity

An issue that is not appropriately addressed in the regression above is that of heterogeneity. As outlined in the literature review, there is plenty of reason to expect patterns of cyclicality to vary with characteristics specific to the sending country. Since the estimates displayed in table 6 aggregate over all countries, it might not come as a surprise that no clear direction of cyclicality emerges. In order to examine whether patterns of cyclicality differ with certain characteristics, in the following, we assign the sending countries to groups according to whether they fulfill specific criteria and re-estimate our model allowing for different slope coefficients for each group. For this purpose, the matrix $\boldsymbol{X_{jt}}$ now includes interaction terms between a dummy variable, indicating whether country j is part of the group, and contemporaneous as well as lagged GDP growth.[26]

Following one of the core findings of Sakellaris and Spilimbergo (2000), column 1 of table 7 investigates whether there is a significant difference in the pattern of cyclicality between OECD members and non-OECD members. Out of the six interaction terms, three are found to be statistically significant, two of which enter with a negative sign. While our findings are therefore much less clear-cut than those of Sakellaris and Spilimbergo (2000), as indicated by the F-test, we can nevertheless argue that there is a significant difference in cyclical behavior between members and non-members of the OECD, even though the direction of the divergence cannot unequivocally be established.

The remaining columns of table 7 attempt to shed some light on the specific factors that might cause the divergence in patterns of cyclicality across countries.[27] Building on the presumption of a relationship between a country's development status and the cyclicality of enrollment, column 2 allows for different slope coefficients for sending countries with above-

[26] The groups are defined as follows: The OECD dummy (column 1) takes a value of 1 if a country is currently a member of the OECD. The Schooling dummy takes a value of 1 if the country's population in 1995 had average years of schooling above the median of our sample (column 3). All other dummy variables are defined as being equal to 1 if the average of a certain characteristic from 1995 to 1999 exceeds the sample median, whereby the characteristics are GDP per capita (column 2), the growth rate of the GDP deflator (column 4), the Gini coefficient (column 5), the index of institutionalized democracy (column 6), and the ratio of credit to the private sector over GDP (column 7). As all dummy variables are defined on the basis of observations prior to the start of our sample period, we do not allow for the group composition to vary over time.

[27] As we are unable to exploit any exogenous variation to assign countries to groups, the results in table 7 should not be interpreted as causal relationships but rather as correlations upon which a more thorough analysis of the macroeconomic determinants of cyclical patterns can be based.

and below-median GDP per capita. In line with the evidence presented in the literature review, we find that the group of rich countries exhibits significantly more countercyclicality than the group of poor countries, with two interaction terms being significant and strongly negative and the F-test rejecting the null hypothesis of the sum of all coefficients being equal to zero. A similar result is found in columns 3 and 6, which split the sample into groups with respect to average years of schooling and the index of institutionalized democracy, respectively. Average years of schooling can be seen as both an indicator of a country's development status and a measure of the valuation of schooling in a country. The index of institutionalized democracy is a typical measure of institutions in the sense of Acemoglu et al. (2001) and as such proxies both economic development and the valuation of economic and personal rights. Since both of the resulting groups display a correlation with the groups split according to GDP per capita of 0.51 and 0.42, respectively, columns 3 and 6 should be interpreted as corroborating our finding that more developed countries exhibit a more countercyclical enrollment pattern than less developed countries.

In column 4, we allow for different degrees of cyclicality for countries with above- and below-median inflation rates. Even though there is no general agreement in the literature about whether moderate rates of inflation have any effect on macroeconomic measures such as the growth of total factor productivity and thus GDP growth, or if an effect only occurs at very high levels of inflation (see i.a. the discussion in Bruno and Easterly 1998), we deem it an interesting exercise to test whether patterns of cyclicality differ with respect to prevalent inflation rates. The majority of interaction terms enters with a positive sign, with the second lag being strongly positive and significant. Accordingly, it appears that countries with higher inflation rates exhibit more procyclical enrollment. Column 5 defines the groups based on the level of inequality as measured by the Gini coefficient. In line with the theoretical models of Galor and Zeira (1993) and Chiu (1998), as discussed in the literature review, we would expect countries with higher inequality to display a more procyclical enrollment pattern, since a larger share of the population can be expected to have low savings and to be excluded from sources of credit. We do not find strong support for this notion as only two interaction terms are statistically significant and of opposite signs and the F-test does not reject. However, the fact that the positive coefficient is about twice as large as the negative coefficient is indicative of a more procyclical pattern among countries with high inequality.

5.2.2 The role of credit constraints

The final column of table 7 attempts to provide a tentative test for the importance of credit constraints in shaping the pattern of cyclicality. For this purpose, we follow De Gregorio (1996) and proxy credit constraints by the ratio of domestic credit to the private sector over GDP and define a dummy variable to take the value of 1 if the five-year average of the credit ratio between 1995 and 1999 exceeds the sample median. This measure has the advantage of broad availability, but, as noted by De Gregorio and Guidotti (1995), might be only weakly related to financial development in highly developed countries. Nevertheless, we find evidence for a more countercyclical pattern in countries with a high credit ratio as two coefficients are negative and significant and the F-test is close to significance with a p-value of 0.14. This is in line with the idea that individuals with less access to borrowing should exhibit a more procyclical response in their enrollment decision as they might find it more difficult to finance their studies during a recession.

In order to corroborate these findings, in table 8, we perform another test of the role of credit constraints by exploiting variation in the real effective exchange rate (henceforth REER). As noted by, among others, Sakellaris and Spilimbergo (2000), a depreciation of the REER implies an increase in the price of education abroad, as long as individuals do not hold internationally diversified portfolios. Coupled with the assumption of imperfect capital markets, a depreciation of the REER should therefore be expected to have a negative impact on enrollment growth. As can be seen in column 1 of table 8, when the REER is included as an additional explanatory variable, while controlling for the cyclical swings due to changes in GDP growth, all included lags enter with a positive sign, whereby the three-year lag is strongly significant. This finding is supported in column 2 where, instead of including the REER directly, we include two dummy variables taking the value of 1 if the REER appreciated or depreciated, respectively, by more than 15% per year. Since we found only the third lag of the REER to be significant in column 1, we lag these dummy variables three times. As a result, the impact of a REER depreciation is found to be negative and highly significant, while strong appreciations do not appear to have any impact on enrollment. This finding parallels that obtained by Sakellaris and Spilimbergo (2000) who report a similar asymmetry between depreciations and appreciations of the REER.

Finally, in columns 3 and 4, we combine our two measures used thus far and re-estimate the regression in column 1 separately for the groups of low and high credit ratio countries. Unfortunately, the coefficient estimates in column 4 are exceedingly imprecise, thus rendering a direct comparison between the two groups to be misleading. However, it should be noted that for the group of countries with a low credit ratio, all lags of the REER enter with a positive coefficient, with the first and third lag being statistically significant, while none of the lags are significant for the countries with a high credit ratio.

5.3 Robustness

In testing the robustness of our results, we employ an alternative set of enrollment data obtained from the UNESCO database. Instead of considering the student flows disaggregated by country pairs, this dataset contains the total number of outbound internationally mobile students for each sending country. This alternative dataset has both advantages as well as disadvantages over the data used in the main analysis. Most importantly, it allows us to test whether the results obtained in previous sections are robust to the different definitions of foreign or internationally mobile students used by the OECD and the UNESCO, respectively. Moreover, by getting rid of the country-pair structure, the correlation across units in the panel should be reduced, thus alleviating one of the factors previously rendering GMM estimation infeasible. However, this comes at the cost of a reduction in the reliability of the data and a loss of information.[28] Finally, since the data aggregates over all destination countries and not just a selected number of OECD countries, we are no longer able to control for the quality of the education and the specific features of the underlying education systems in the recipient countries that are likely to be heterogeneous.

As a first test of robustness, we employ the new dataset to re-estimate equation (1) using GMM. It was noted in section 5.1 that a likely reason for the rejection of the Hansen test in tables 3 and 5 was the failure of the assumption of the error terms being independent across individuals induced by the fact that the same countries appeared in multiple country pairs. Since this assumption appears more likely to hold in the alternative sample, this might allow us to test the robustness of the results with respect to the estimation methodology used. Table 9 shows OLS, WG, and GMM coefficient estimates of the growth rate of the number of outbound students on two of its lags and contemporaneous as well as five lags of GDP growth. It can be seen that, again, the coefficient on the lagged dependent variable is insignificant except for a single negative and significant coefficient using the downward biased WG estimate. Furthermore, the estimated coefficients on lagged GDP growth are almost identical to those obtained using the WG estimator on a model estimated without a lagged dependent variable (column 1). Nevertheless, the difference- and system-GMM estimates presented in columns 4

[28] This deterioration in data quality is highlighted by the fact that for most countries and years the total number of outbound students is merely an estimate by the UNESCO Institute for Statistics.

to 7 again cast doubt on the validity of the instruments since the Hansen test rejects at the 10% level when a reduced instrument set is considered. Therefore, it appears that also in this dataset, the necessary assumptions for the use of GMM techniques are not fulfilled. Nevertheless, we construe the fact that the lagged dependent variable is again found to be insignificant as supportive evidence for our estimation strategy used to obtain the main results.

Finally, table 10 reproduces the results from table 7 for this new dataset. While most of the results appear robust to the use of different enrollment data, there are three notable changes. First, the results for the groups of OECD and non-OECD countries are far more pronounced and more closely match those obtained by Sakellaris and Spilimbergo (2000). Except for one (insignificant) interaction term, all coefficients are strongly negative and predominantly significant, clearly indicating more countercyclical enrollment in OECD countries. Second, the finding of a more procyclical enrollment pattern for countries with higher inflation rates breaks down. Even though the majority of interaction terms is positive, the only significant coefficient has a negative sign, thus making it impossible to draw any unambiguous conclusions. Third, the finding of a larger degree of procyclicality among countries with high inequality is considerably more marked, as the vast majority of interaction terms enter with a positive and significant coefficient and the F-test rejects on all conventional levels. Overall, it can be argued that, with the exception of column 4, all results in table 7 are reasonably robust to the use of a different data source.

6. Summary and conclusion

This paper aimed at analyzing the cyclicality of enrollment in tertiary education by exploiting variation in student flows across countries. For this purpose, we investigated the relationship between the enrollment of foreign students and the business cycle in their country of origin in an innovative country-pair panel dataset, spanning the years from 1999 to 2012.

While we do not find a clear pattern of cyclicality for the sample as a whole, we argue that this is due to heterogeneity across countries and examine this possibility by allowing for different degrees of cyclicality in certain sub-samples. In line with previous results in the literature, we find that there is a significant difference in the cyclicality of tertiary enrollment between OECD and non-OECD countries. By splitting the sample with respect to the initial values of GDP per capita, years of schooling, and the index of institutionalized democracy, we are able to confirm the presumption that more developed countries tend to have more countercyclical enrollment, even though our analysis is not suited to explain which specific factors are driving this divergence. Moreover, we find that countries with a higher level of inequality tend to exhibit more procyclical enrollment, an insight that is in line with theoretical models of human capital accumulation based on credit constraints. Finally, we test directly for the importance of credit constraints by exploiting variation in both the real effective exchange rate and the ratio of domestic credit to the private sector over GDP. Our results confirm the relevance of credit constraints in shaping the pattern of cyclicality and are shown to be robust to the use of different data sources.

Overall, we are able to confirm the existence of a systematic relationship between cyclical swings and enrollment in tertiary education, thus underscoring the importance of intertemporal substitution between labor and schooling. However, there are several issues that cannot appropriately be addressed in this study and therefore leave room for further research. First, it remains unclear how the cyclical swings in the number of foreign students enrolled in a country benefit the destination country and the country of origin. In order to answer this question, it would be necessary to track in which country the individuals apply their human capital – a piece of information missing in our sample. Second, as already mentioned in the introduction, our analysis is not suited to answer the question of the long-run effects of the cyclical swings in human capital accumulation. This is due to the fact that we focus exclusively on schooling and largely ignore on-the-job training, which constitutes a considerable part of human capital accumulation and whose cyclical properties remain rather elusive to this day. Méndez and Sepúlveda (2012) make use of a detailed panel of individuals in the US to show that while training appears acyclical in the aggregate, some components, most notably firm-financed training display a procyclical pattern. Furthermore, the authors note that the cyclicality of training activities tends to depend on the individual’s education level. While this paper therefore serves as a useful starting point, additional work on the topic is needed.

References

Abramitzky, Ran, and Victor Lavy. 2014. How Responsive Is Investment in Schooling to Changes in Redistributive Policies and in Returns? *Econometrica*. Vol. 82(4), pp. 1241-1272.

Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*. Vol. 91(5), pp. 1369-1401.

Altonji, Joseph. 1986. Intertemporal Substitution in Labor Supply: Evidence from Micro Data. *Journal of Political Economy*. Vol. 94(3 pt.2), pp. 176-215.

Anderson, Theodore, and Cheng Hsiao. 1982. Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics*. Vol. 18, pp. 47-82.

Angrist, Joshua, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricists Companion*. First Edition. Princeton University Press.

Arellano, Manuel. 1989. A Note on the Anderson-Hsiao Estimator for Panel Data. *Economics Letters*. Vol. 31, pp. 337-341.

Arellano, Manuel, and Stephen Bond. 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*. Vol. 58(2), pp. 277-297.

Arellano, Manuel, and Olympia Bover. 1995. Another Look at the Instrumental Variable Estimation of Error-Components Models. *Journal of Econometrics*. Vol. 68, pp. 29-51.

Ball, Laurence. 2014. Long-Term Damage from the Great Recession in OECD Countries. NBER Working Paper No. 20185.

Barro, Robert, and Jong Wha Lee. 2013. A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*. Vol. 104, pp. 184-198.

Bazzi, Samuel, and Michael Clemens. 2013. Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth. *American Economic Journal: Macroeconomics*. Vol. 5(2), pp. 152-186.

Becker, Gary. 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*. Vol. 70(5 pt.2), pp. 9-49.

Becker, Gary. 1975. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago: The University of Chicago Press. Second Edition.

Bedard, Kelly, and Douglas Herman. 2008. Who Goes to Graduate/Professional School? The Importance of Economic Fluctuations, Undergraduate Field, and Ability. *Economics of Education Review*. Vol. 27, pp. 197-210.

Ben-Porath, Yoram. 1967. The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*. Vol. 75(4 pt.1), pp. 352-365.

Berger, Mark, and Thomas Kostal. 2002. Financial Resources, Regulation, and Enrollment in US Public Higher Education. *Economics of Education Review*. Vol. 21, pp. 101-110.

Betts, Julian, and Laurel McFarland. 1995. Safe Port in a Storm: The Impact of Labor Market Conditions on Community College Enrollments. *Journal of Human Resources*. Vol. 30(4), pp. 741-765.

Bils, Mark. 1985. Real Wages over the Business Cycle: Evidence from Panel Data. *Journal of Political Economy*. Vol. 93(4), pp. 666-689.

Binder, Melissa. 1999. School Indicators during Mexico's "Lost Decade". *Economics of Education Review*. Vol. 18, pp. 183-199.

Binder, Michael, Cheng Hsiao, and Hashem Pesaran. 2005. Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration. *Econometric Theory*. Vol. 21, pp. 795-837.

Black, Sandra, and Amir Sufi. 2002. Who Goes to College? Differential Enrollment by Race and Family Background. NBER Working Paper No. 9310.

Blanchard, Olivier, and Lawrence Summers. 1986. Hysteresis and the European Unemployment Problem, in Fischer, Stanley (ed.*). NBER Macroeconomics Annual 1986, Volume 1*. Cambridge, Massachusetts: The MIT Press.

Blundell, Richard, and Stephen Bond. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*. Vol. 87, pp. 115-143.

Blundell, Richard, and Stephen Bond. 2000. GMM Estimation with Persistent Panel Data: An Application to Production Functions. *Econometric Reviews*. Vol. 19(3), pp. 321-340.

Blundell, Richard, Stephen Bond, and Frank Windmeijer. 2001. Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator, in: Baltagi, Badi, Thomas Fomby, and Carter Hill (eds.). *Advances in Econometrics, Volume 15: Nonstationary Panels, Panel Cointegration and Dynamic Panels*. Emerald Group Publishing Limited.

Bond, Stephen. 2002. Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice. *Portuguese Economic Journal*. Vol. 1, pp. 141-162.

Bowlus, Audra, and Haoming Liu. 2003. The Long-Term Effects of Graduating from High School During a Recession: Bad Luck or Forced Opportunity?. CIBC Working Paper No. 2003- 7.

Bowsher, Clive. 2002. On Testing Overidentifying Restrictions in Dynamic Panel Data Models. *Economics Letters*. Vol. 77, pp. 211-220.

Bruno, Michael, and William Easterly. 1998. Inflation Crises and Long-Run Growth. *Journal of Monetary Economics*. Vol. 41(1), pp. 3-26.

Bun, Maurice, and Frank Windmeijer. 2010. The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models. *Econometrics Journal*. Vol. 13, pp. 95-126.

Cameron, Colin, and Pravin Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press. First Edition.

Card, David. 1999. The Causal Effect of Education on Earnings, in: Card, David, and Orley Ashenfelter (eds.). *Handbook of Labor Economics, Volume 3*. Elsevier Science Publishers BV.

Card, David, and Thomas Lemieux. 2001a. Dropout and Enrollment Trends in the Postwar Period: What Went Wrong in the 1970s?, in: Gruber, Jonathan (ed.). *Risky Behavior among Youths: An Economic Analysis*. University of Chicago Press.

Card, David, and Thomas Lemieux. 2001b. Going to College to Avoid the Draft: The Unintended Legacy of the Vietnam War. *American Economic Review*. Vol. 91(2), pp. 97-102.

Chiu, Henry. 1998. Income Inequality, Human Capital Accumulation and Economic Performance. *Economic Journal*. Vol. 108(1), pp. 44-59.

Christian, Michael. 2007. Liquidity Constraints and the Cyclicality of College Enrollment in the United States. *Oxford Economic Papers*. Vol. 59(1), pp. 141-169.

Christou, Costas. 2001. Differential Borrowing Constraints and Investment in Human Capital. *Journal of Macroeconomics*. Vol. 23(1), pp. 277-295.

Cragg, John, and Stephen Donald. 1993. Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory*. Vol. 9(2), pp. 222-240.

De Gregorio, José. 1996. Borrowing Constraints, Human Capital Accumulation, and Growth. *Journal of Monetary Economics*. Vol. 37, pp. 49-71.

De Gregorio, José, and Pablo Guidotti. 1995. Financial Development and Economic Growth. *World Development*. Vol. 23(3), pp. 433-448.

Dellas, Harris, and Vally Koubi. 2003. Business Cycles and Schooling. *European Journal of Political Economy,* Vol. 19(4), pp. 843-859.

Dellas, Harris, and Plutarchos Sakellaris. 2003. On the Cyclicality of Schooling: Theory and Evidence. *Oxford Economic Papers,* Vol. 55(1), pp. 148-172.

Di Maria, Corrado, and Piotr Stryszowski. 2009. Migration, Human Capital Accumulation and Economic Development. *Journal of Development Economics*. Vol. 90(2), pp. 306-313.

Everaert, Gerdie, and Lorenzo Pozzi. Bootstrap-based Bias Correction for Dynamic Panels. *Journal of Economic Dynamics and Control*. Vol. 31(4), pp. 1160-1184.

Ferreira, Francisco, and Norbert Schady. 2009. Aggregate Economic Shocks, Child Schooling, and Child Health. *World Bank Research Observer*. Vol. 24(1), pp. 1-35.

Foster, Andrew, and Mark Rosenzweig. 1996. Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution. *American Economic Review*. Vol. 86(4), pp. 931-953.

Galor, Oded, and Joseph Zeira. 1993. Income Distribution and Macroeconomics. *Review of Economic Studies*. Vol 60(1), pp. 35-52.

Genda, Yuji, Ayako Kondo, and Souichi Ohta. 2010. Long-Term Effects of a Recession at Labor Market Entry in Japan and the United States. *The Journal of Human Resources*. Vol. 45(1), pp. 157-196.

Goldin, Claudia. 1999. Egalitarianism and the Returns to Education during the Great Transformation of American Education. *The Journal of Political Economy*. Vol. 107(6), pp. 65-94.

Gylfason, Thorvaldur. 2001. Natural Resources, Education, and Economic Development. *European Economic Review*. Vol. 45(4), pp. 847-859.

Hansen, Lars. 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*. Vol. 50(4), pp. 1029-1054.

Hansen, Lars, John Heaton, and Amir Yaron. 1996. Finite-Sample Properies of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics*. Vol. 14(3), pp. 262-280.

Harris, Richard, and Elias Tzavalis. Inference for Unit Roots in Dynamic Panels Where the Time Dimension is Fixed. *Journal of Econometrics*. Vol. 91, pp. 201-226.

Hauk, William, and Romain Wacziarg. 2009. A Monte Carlo Study of Growth Regressions. *Journal of Economic Growth*. Vol. 14(2), pp. 103-147.

Hershbein, Brad. 2012. Graduating High School in a Recession: Work, Education, and Home Production. *The B.E. Journal of Economic Analysis & Policy*. Vol. 12(1), Art. 3.

Heylen, Freddy, and Lorenzo Pozzi. 2007. Crises and Human Capital Accumulation. *Canadian Journal of Economics*. Vol. 40(4), pp. 1261-1285.

Heyneman, Stephen. 1990. Economic Crisis and the Quality of Education. International *Journal of Educational Development*. Vol. 10(2), pp. 115-129.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey Rosen. 1988. Estimating Vector Autoregressions with Panel Data. *Econometrica*. Vol. 56(6), pp. 1371-1395.

Hotchkiss, Julie, Melinda Pitts, and Fernando Rios-Avila. 2012. A Closer Look at Nonparticipants during and After the Great Recession. Federal Reserve Bank of Atlanta Working Paper No. 2012-10.

Jacoby, Hanan. 1994. Borrowing Constraints and Progress through School: Evidence from Peru. *Review of Economics and Statistics*. Vol. 76(1), pp. 151-160.

Jensen, Robert. 2000. Agricultural Volatility and Investments in Children. *American Economic Review*. Vol. 90(2), pp. 399-404.

Johnson, Matthew. 2013. The Impact of Business Cycle Fluctuations on Graduate School Enrollment. *Economics of Education Review*. Vol. 34, pp. 122-134.

Judson, Ruth, and Ann Owen. 1999. Estimating Dynamic Panel Data Models: A Guide for Macroeconomists. *Economics Letters*. Vol. 65(1), pp. 9-15.

Kane, Thomas. 1994. College Entry by Blacks since 1970: The Role of College Costs, Family Background, and the Returns to Education. *Journal of Political Economy*. Vol. 102(5), pp. 878-911.

Kehoe, Timothy. 2003. What Can We Learn from the Current Crisis in Argentina? *Scottish Journal of Political Economy*. Vol. 50(5), pp. 609-633.

Kondo, Ayako. 2007. Differential Effects of Graduating During a Recession Across Race and Gender. ISERP Working Paper No. 07-05.

Kiviet, Jan. 1995. On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models. *Journal of Econometrics*. Vol. 68, pp. 53-78.

Kydland, Finn, and Edward Prescott. 1982. Time to Build and Aggregate Fluctuations. *Econometrica*. Vol. 50(6), pp. 1345-1370.

Levin, Andrew, Chien-Fu Lin and James Chu. 2002. Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. *Journal of Econometrics*. Vol. 108(1), pp. 1-24.

Lilien, David, and Robert Hall. 1986. Cyclical Fluctuations in the Labor Market, in: Ashenfelter, Orley, and Richard Layard (eds.). *Handbook of Labor Economics, Vol. 2*. Elsevier Science Publishers BV.

Long, John, and Charles Plosser. 1983. Real Business Cycles. *Journal of Political Economy*. Vol. 91(1), pp. 39-69.

Loury, Glenn. 1981. Intergenerational Transfers and the Distribution of Earnings. *Econometrica*. Vol. 49(4), pp. 843-867.

Mankiw, Gregory, Julio Rotemberg, and Lawrence Summers. 1985. Intertemporal Substitution in Macroeconomics. *Quarterly Journal of Economics*. Vol. 100(1), pp. 225-251.

Marshall, Monty, Ted Gurr and Keith Jaggers. 2015. Polity IV Project: Political Regime Characteristics and Transitions, 1800-2015. Center for Systemic Peace.

McKenzie, David. 2003. How do Households Cope with Aggregate Shocks? Evidence from the Mexican Peso Crisis. *World Development*. Vol. 31(7), pp. 1179-1199.

Méndez, Fabio, and Facundo Sepúlveda. 2012. The Cyclicality of Skill Acquisition: Evidence from Panel Data. *American Economic Journal: Macroeconomics*. Vol. 4(3), pp. 128-152.

Mincer, Jacob. 1958. Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*. Vol. 66(4), pp. 281-302.

Mincer, Jacob. 1984. Human Capital and Economic Growth. *Economics of Education Review*. Vol. 3(3), pp. 195-205.

Mishkin, Frederic. 1981. The Real Interest Rate: An Empirical Investigation. *Carnegie-Rochester Conference Series on Public Policy*. Vol. 15, pp. 151-200.

Nelson, Richard, and Edmund Phelps. 1966. Investment in Humans, Technological Diffusion, and Economic Growth. *American Economic Review*. Vol. 56(1), pp. 69-75.

Nerlove, Marc. 1971. Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections. *Econometrica*. Vol. 39(2), pp. 359-382.

Newey, Whitney, and Frank Windmeijer. 2009. Generalized Method of Moments with Many Weak Moment Conditions. *Econometrica*. Vol. 77(3), pp. 687-719.

Nickell, Stephen. 1981. Biases in Dynamic Models with Fixed Effects. *Econometrica*. Vol. 49(6), pp. 1417-1426.

Oreopoulos, Philip, Till van Wachter, and Andrew Heisz. 2012. The Short- and Long-Term Career Effects of Graduating in a Recession. *American Economic Journal: Applied Economics*, Vol. 4(1), pp. 1-29.

Riddell, Craig, and Arthur Sweetman. 2000. Human Capital Formation in a Period of Rapid Change. *Adapting Public Policy to a Labour Market in Transition*, pp. 88-141.

Roodman, David. 2006. How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata. Center for Global Development Working Paper No. 103.

Roodman, David. 2009. A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics*, Volume 71(1), pp. 135-158.

Sakellaris, Plutarchos, and Antonio Spilimbergo. 2000. Business Cycles and Investment in Human Capital: International Evidence on Higher Education. *Carnegie-Rochester Conference Series on Public Policy*. Vol. 52, pp. 221-256.

Sargan, Denis. 1958. The Estimation of Economic Relationships Using Instrumental Variables. *Econometrica*. Vol. 26(3), pp. 393-415.

Schady, Norbert. 2004. Do Macroeconomic Crises Always Slow Human Capital Accumulation? *The World Bank Economic Review*, Vol. 18(2), pp. 131-154.

Skidmore, Mark, and Hideki Toya. 2002. Do Natural Disasters Promote Long-Run Growth? *Economic Inquiry*. Vol. 40(4), pp. 664-687.

Stiglitz, Joseph. 2015. The Measurement of Wealth: Recessions, Sustainability and Inequality. NBER Working Paper No. 21327.

Stock, James, and Mark Watson. 1990. Business Cycle Fluctuations in US Macroeconomic Time Series, in: Taylor, John, and Michael Woodford (eds.). *Handbook of Macroeconomics, Volume 1*. Elsevier Science Publishers BV.

Stock, James, and Motohiro Yogo. 2005. Testing for Weak Instruments in Linear IV Regression, in: Andrews, Donald, James Stock, and Thomas Rothenberg. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg.*

Thomas, Duncan, Kathleen Beegle, Elizabeth Frankenberg, Bondan Sikoki, John Strauss, and Graciela Teruel. 2004. Education in a Crisis. *Journal of Development Economics*. Vol. 74(1), pp. 53-85.

UNESCO Institute for Statistics. 2015. Frequently Asked Questions: Education Statistics. September 2015. http://www.uis.unesco.org/Education/Pages/FAQ.aspx (accessed 25th June 2016).

Van Damme, Dirk, and Kiira Karkkainen. 2011. OECD Education Today Crisis Survey 2010: The Impact of the Economic Recession and Fiscal Crisis on Education in OECD Countries. OECD Education Working Paper No. 56.

Wilcox, James. 1983. Why Real Interest Rates Were So Low in the 1970s. *American Economic Review*. Vol. 73(1), pp. 44-53.

Windmeijer, Frank. 2005. A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators. *Journal of Econometrics*. Vol. 126, pp. 25-51.

Wooldridge, Jeffrey. 2001. Applications of Generalized Method of Moments Estimation. *Journal of Economic Perspectives*. Vol. 15(4), pp. 87-100.

Wooldridge, Jeffrey. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press. Second Edition.

Appendix A: Methodology

A.1 Dynamic panel bias

For simplicity, we assume a simple autoregressive process of order 1 for the enrollment series:

$$E_{ijt} = \alpha E_{ijt-1} + \mu_{ij} + v_{ijt}, \tag{3}$$

where we assume $1 > \alpha > 0$. When equation (3) is estimated by pooled OLS, the individual fixed effect μ_{ij} is treated as part of the error term, so that we can rewrite the equation to obtain

$$E_{ijt} = \alpha E_{ijt-1} + \varepsilon_{ijt},\ \ \varepsilon_{ijt} = \mu_{ij} + v_{ijt}. \tag{4}$$

In order to show the correlation between E_{ijt-1} and ε_{ijt}, we can rewrite equation (4) for period t-1 and insert the resulting expression for E_{ijt-1} in equation (4). This yields

$$E_{ijt} = \alpha(\alpha E_{ijt-2} + \varepsilon_{ijt-1}) + \varepsilon_{ijt}, \tag{5}$$

which, taking equation (4) into consideration, clearly shows a positive correlation between ε_{ijt-1} and ε_{ijt} (and thus between E_{ijt-1} and ε_{ijt}) as both terms depend positively on the fixed effect μ_{ij}. We can analytically treat ε_{ijt-1} as an omitted variable and apply the standard omitted variable bias formula (see e.g. Angrist and Pischke 2009).[29] Since both E_{ijt-2} and ε_{ijt-1} depend positively on the individual fixed effect, there is a positive correlation between the two variables. Moreover, the impact of ε_{ijt-1} on E_{ijt} is positive, as a large positive error term in period t-1 increases E_{ijt-1} and hence, due to the persistence of the process, also E_{ijt}. With both bias components being positive, it becomes clear that the dynamic panel bias for OLS should be positive.

In the same manner, we can show that there is a negative correlation in the within groups estimator. First, recall that the within groups estimator transforms the variables by subtracting the sample mean from each observation. Therefore, we rewrite equation (3) as

$$E^*_{ijt} = \alpha E^*_{ijt-1} + v^*_{ijt}, \tag{6}$$

where

[29] In a standard OLS model with only one explanatory variable and one omitted variable, the omitted variable bias is given by $Bias = \gamma\delta_{IO}$, where γ is the impact of the omitted variable on the dependent variable and δ_{IO} is the coefficient vector of an OLS regression of the omitted variable on the included explanatory variable for which we are determining the bias (Angrist and Pischke 2009).

$$E^*_{ijt} = E_{ijt} - \frac{1}{T-1}\sum_{k=2}^{T} E_{ijk} \quad (7)$$

$$v^*_{ijt} = v_{ijt} - \frac{1}{T-1}\sum_{k=2}^{T} v_{ijk}\,.[30] \quad (8)$$

It can easily be seen that there is a correlation between E^*_{ijt-1} and v^*_{ijt} since E_{ijt-1} is by definition correlated with $-\frac{1}{T-1}v_{ijt-1}$, which is part of the transformed error term (Roodman 2006).

Rewriting equation (6) as

$$E^*_{ijt} = \alpha\left(\alpha E^*_{ijt-2} + v^*_{ijt-1}\right) + v^*_{ijt}, \quad (9)$$

we can again treat v^*_{ijt-1} as an omitted variable and apply the omitted variable bias formula. The correlation between E^*_{ijt-2} and v^*_{ijt-1} is negative due to the negative correlation between E_{ijt-2} and $-\frac{1}{T-1}v_{ijt-2}$ outlined above. Moreover, the impact of v^*_{ijt-1} on E^*_{ijt} is positive due to the persistence of the series. Since the two bias components therefore have opposing signs, the overall bias is negative.

A.2 Instrument matrices

In this section we again consider equation (3), only this time in its first-differenced form,

$$\Delta E_{ijt} = \alpha \Delta E_{ijt-1} + \Delta v_{ijt}.[31] \quad (9)$$

When the first-difference transformation is applied, the first time period is lost, so that the sample now starts at t=2. In applying the simple Anderson-Hsiao estimator by instrumenting ΔE_{ijt-1} with E_{ijt-2}, we use the following instrument matrix,

$$Z_{ij} = \begin{bmatrix} \cdot \\ E_{ij1} \\ \vdots \\ E_{ijT-2} \end{bmatrix}, \quad (10)$$

where the first row refers to period t=2 which has to be omitted due to a lack of an instrument. The trade-off between the lag distance and the depth of the sample can easily be seen since adding another lag as an instrument would require the sample to start at period t=3. As outlined in detail in chapter 3, the standard difference-GMM instrument matrix includes separate instruments for each lag and time period. The resulting instrument matrix is of the 'exploded' or HENR-structure,

[30] Since the individual effect is time-invariant, subtracting the sample mean from each observation causes it to drop out of the equation.

[31] The only exception is the system GMM estimator which considers both the model in levels as well as the model in first differences.

$$Z_{ij}=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ E_{ij1} & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & E_{ij2} & E_{ij1} & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & E_{ij3} & E_{ij2} & E_{ij1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{11}$$

Alternatively, the instrument matrix might be collapsed to obtain

$$Z_{ij}=\begin{bmatrix} 0 & 0 & 0 & \cdots \\ E_{ij1} & 0 & 0 & \cdots \\ E_{ij2} & E_{ij1} & 0 & \cdots \\ E_{ij3} & E_{ij2} & E_{ij1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{12}$$

which has the advantage of reducing the number of instrument while retaining most of the informational content of the full instrument matrix.

Finally, the system-GMM estimator builds a stacked system of equations that extends the instrument matrix (11) by the instruments from the levels equation, namely

$$Z_{ij}=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ \Delta E_{ij2} & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & \Delta E_{ij3} & \Delta E_{ij2} & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \Delta E_{ij4} & \Delta E_{ij3} & \Delta E_{ij2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{13}$$

where the first row now corresponds to period t=1. As Roodman (2006) notes, many of the moment conditions implied by a stacked system of (11) and (13) are mathematically redundant. As a result, the Stata ‘xtabond2’ package only includes a single lag of the first-differenced dependent variable as an instrument for each time period (Roodman 2006).

Appendix B: Data appendix

B.1 Data definitions and sources

Enrollment (Main Dataset)	The data on enrollment disaggregated by country pairs is taken from the OECD data series on the number of non-citizen students of the reporting country. Since the OECD lacks data for the US after 2004, we have supplemented this dataset by the number of non-citizen students in the US reported in the yearly IIE publication 'Open Doors' which are identical to those published by the OECD prior to 2004.
Enrollment (Alternative Dataset)	The data on the total number of outbound internationally mobile students is published by the UNESCO. The UNESCO defines an internationally mobile students as 'an individual who has physically crossed an international border between two countries with the objective to participate in educational activities in a destination country, where the destination country is different from his or her country of origin' (UIS 2016), whereby the country of origin is defined as the country in which prior education was obtained.
GDP Per Capita	Data on GDP per capita in PPP dollars is taken from the IMF's World Economic Outlook.
Real Effective Exchange Rate	Data on real effective exchange rates is provided in the World Bank World Development Indicators database.
GDP Deflator	The data on inflation measured as the growth rate of the GDP deflator is taken from the World Bank.
Years of Schooling	Data on average years of schooling of the population aged 15 and over is provided by Barro and Lee (2013).
Gini Coefficient	Data on Gini coefficients is taken from the United Nations University's World Income Inequality Database.
Domestic Credit to the Private Sector over GDP	Data on financial resources provided to the private sector by financial corporations is published by the World Bank.
Index of Institutionalized Democracy	The index of institutionalized democracy is a composite measure of democratic institutions, constraint on the executive, and civil liberties, and is published as part of the Polity IV Project (Marshall et al. 2015).

B.2 Sample adjustments

For the main sample, we restricted the dataset to country pairs fulfilling two criteria. First, there has to be data available on both GDP per capita and enrollment for all years (1999 to 2013) of the panel. Most importantly, there should not be any missing observations in the middle of the sample since this would inhibit the use of complex lag structures we are interested in. Second, we omit all country pairs for which the number of foreign students is below 40 at any point in time. This is advantageous because it excludes all country pairs with minuscule numbers of foreign students that are unlikely to be subject to cyclical fluctuations and might introduce considerable noise to the regression. The following table shows all sending countries included in the dataset and their characteristics, as defined in table 7.

Table B1: Country Characteristics

Country	OECD	GDP	Schooling	Inflation	Inequality	Democracy	Credit
Albania	0	0	1	1	0	0	0
Algeria	0	1	0	1	0	0	0
Angola	0	0	.	1	.	0	0
Argentina	0	1	1	0	1	1	0
Armenia	0	0	1	1	1	0	0
Australia	1	1	1	0	0	1	1
Austria	1	1	1	0	0	1	1
Azerbaijan	0	0	.	1	1	0	0
Bahamas	0	1	.	1	.	.	1
Bahrain	0	1	0	0	.	0	1
Bangladesh	0	0	0	1	0	.	0
Barbados	0	1	1	0	.	.	1
Belarus	0	0	.	1	0	0	0
Belgium	1	1	1	0	0	1	1
Belize	0	0	1	0	1	.	1
Benin	0	0	0	1	.	1	0
Bolivia	0	0	1	1	1	1	1
Bosnia and Herzegowina	0	0	.	0	.	.	1
Botswana	0	1	1	1	.	1	0
Brazil	0	1	0	1	1	1	1
Bulgaria	0	1	1	1	0	1	1
Burkina Faso	0	0	.	0	1	0	0
Burundi	0	0	0	1	1	0	0
Cambodia	0	0	0	0	0	0	0
Cameroon	0	0	0	0	1	0	0
Canada	1	1	1	0	0	1	1
Cape Verde	0	0	.	0	.	1	1
Central African Republic	0	0	0	0	.	0	0
Chad	0	0	.	0	.	0	0
Chile	0	1	1	0	1	1	1
China	0	0	0	0	0	0	1
Colombia	0	1	0	1	1	1	1
Comoros	0	0	.	0	.	0	0
Congo	0	0	0	1	.	0	0
Congo Dem. Rep.	0	0	0	1	.	.	0
Costa Rica	0	1	1	1	1	1	0
Croatia	0	1	1	0	0	0	1
Cyprus	0	1	1	0	0	1	1
Czech Republic	1	1	1	1	0	1	1
Côte d'Ivoire	0	0	0	0	0	.	0
Denmark	1	1	1	0	0	1	1
Djibouti	0	0	.	0	1	0	1
Dominica	0	1	.	0	.	.	1
Dominican Republic	0	1	0	0	1	1	0
Ecuador	0	1	0	0	1	1	0
Egypt	0	1	0	1	0	0	1
El Salvador	0	0	0	0	1	1	1
Equatorial Guinea	0	0	.	0	.	0	0
Eritrea	0	0	.	1	.	0	1
Estonia	1	1	1	1	0	1	1
Ethiopia	0	0	.	0	1	0	0
Fiji	0	0	1	0	.	1	1
Finland	1	1	1	0	0	1	1
France	1	1	1	0	0	1	1

Country	OECD	GDP	Schooling	Inflation	Inequality	Democracy	Credit
Gambia	0	0	0	0	1	0	0
Georgia	0	0	.	1	1	1	0
Germany	1	1	1	0	0	1	1
Ghana	0	0	0	1	0	0	0
Greece	1	1	1	0	0	1	1
Guatemala	0	0	0	1	1	1	0
Guinea	0	0	.	0	.	0	0
Guyana	0	0	1	0	1	1	1
Haiti	0	0	0	1	.	1	0
Honduras	0	0	0	1	1	1	1
Hong Kong	0	1	1	0	1	.	1
Hungary	1	1	1	1	0	1	0
Iceland	1	1	1	0	.	.	1
India	0	0	0	1	0	1	0
Indonesia	0	0	0	1	0	0	1
Iran	0	1	0	1	1	0	0
Ireland	1	1	1	0	0	1	1
Israel	1	1	1	1	1	1	1
Italy	1	1	1	0	0	1	1
Jamaica	0	1	1	1	1	1	0
Japan	1	1	1	0	0	1	1
Jordan	0	1	1	0	0	0	1
Kazakhstan	0	1	1	1	1	0	0
Kenya	0	0	0	1	1	0	0
Kuwait	0	1	0	0	.	0	1
Kyrgyzstan	0	0	1	1	1	0	0
Laos	0	0	0	1	0	0	0
Latvia	0	1	1	0	0	1	0
Lebanon	0	1	.	1	1	.	1
Libya	0	1	0	.	.	0	1
Lithuania	0	1	1	1	0	1	0
Macedonia	0	1	.	0	0	1	0
Madagascar	0	0	.	1	0	1	0
Malawi	0	0	0	1	1	1	0
Malaysia	0	1	1	0	1	0	1
Maldives	0	0	0	.	1	.	0
Mali	0	0	0	0	.	1	0
Malta	0	1	1	0	.	.	1
Mauritania	0	0	0	1	0	0	.
Mauritius	0	1	0	0	0	1	1
Mexico	1	1	0	1	1	0	0
Moldova	0	0	.	1	1	1	0
Mongolia	0	0	1	1	0	1	0
Morocco	0	0	0	0	0	0	1
Mozambique	0	0	0	1	1	0	0
Myanmar	0	.	0	1	.	0	0
Namibia	0	0	0	1	.	1	1
Nepal	0	0	0	1	1	0	1
Netherlands	1	1	1	0	0	.	1
New Zealand	1	1	1	0	0	1	1
Nicaragua	0	0	0	1	1	1	0
Niger	0	0	0	0	1	0	0
Nigeria	0	0	.	1	1	0	0
Norway	1	1	1	0	0	1	1
Oman	0	1	.	0	.	0	1
Pakistan	0	0	0	1	0	1	1
Panama	0	1	1	0	1	1	1
Papua New Guinea	0	0	0	1	1	0	0
Paraguay	0	0	0	1	1	1	1
Peru	0	0	1	1	1	0	0
Philippines	0	0	1	1	1	1	1

Country	OECD	GDP	Schooling	Inflation	Inequality	Democracy	Credit
Poland	1	1	1	1	0	1	0
Portugal	1	1	0	0	0	1	1
Qatar	0	1	0	.	.	0	1
Republic of Korea	1	1	1	0	0	1	1
Romania	0	1	1	1	0	1	0
Russia	0	1	1	1	1	0	0
Rwanda	0	0	0	1	.	0	0
Saint Kitts and Nevis	0	1	.	0	.	.	1
Saint Lucia	0	1	.	0	1	.	1
Saint Vincent	0	1	.	0	.	.	1
Samoa	0	.	.	0	.	.	0
Saudi Arabia	0	1	0	0	.	0	0
Senegal	0	0	0	0	.	0	0
Seychelles	0	1	.	0	.	.	0
Sierra Leone	0	0	0	1	.	0	0
Singapore	0	1	1	0	1	0	1
Slovakia	1	1	1	0	0	1	1
Slovenia	1	1	1	1	0	1	1
Solomon Islands	0	0	.	1	.	1	0
South Africa	0	1	1	1	1	1	1
Spain	1	1	1	0	0	1	1
Sri Lanka	0	0	1	1	0	1	1
Sudan	0	0	0	1	.	0	0
Suriname	0	1	.	1	1	1	0
Swaziland	0	0	0	1	1	0	0
Sweden	1	1	1	0	0	1	1
Switzerland	1	1	1	0	0	1	1
Tajikistan	0	0	1	1	0	0	0
Tanzania	0	0	0	1	.	0	0
Thailand	0	1	0	0	1	1	1
Togo	0	0	0	0	.	0	0
Tonga	0	0	1	0	.	.	1
Trinidad and Tobago	0	1	1	0	.	1	1
Tunisia	0	0	0	0	1	0	1
Turkey	1	1	0	1	.	1	0
Turkmenistan	0	0	.	1	0	0	.
Uganda	0	0	0	0	0	0	0
Ukraine	0	0	1	1	0	1	0
United Arab Emirates	0	1	1	0	.	0	1
United Kingdom	1	1	1	0	0	1	1
United States of America	1	1	1	0	1	1	1
Uruguay	0	1	1	1	1	1	1
Uzbekistan	0	0	.	1	1	0	.
Venezuela	0	1	0	1	1	1	0
Vietnam	0	0	0	1	0	0	0
Yemen	0	0	0	1	0	0	0
Zambia	0	0	0	1	1	0	0
Zimbabwe	0	.	0	0	1	0	1

Appendix C: Tables and Figures

Figure 1: Total Number of Outbound Students by Region; UNESCO Data

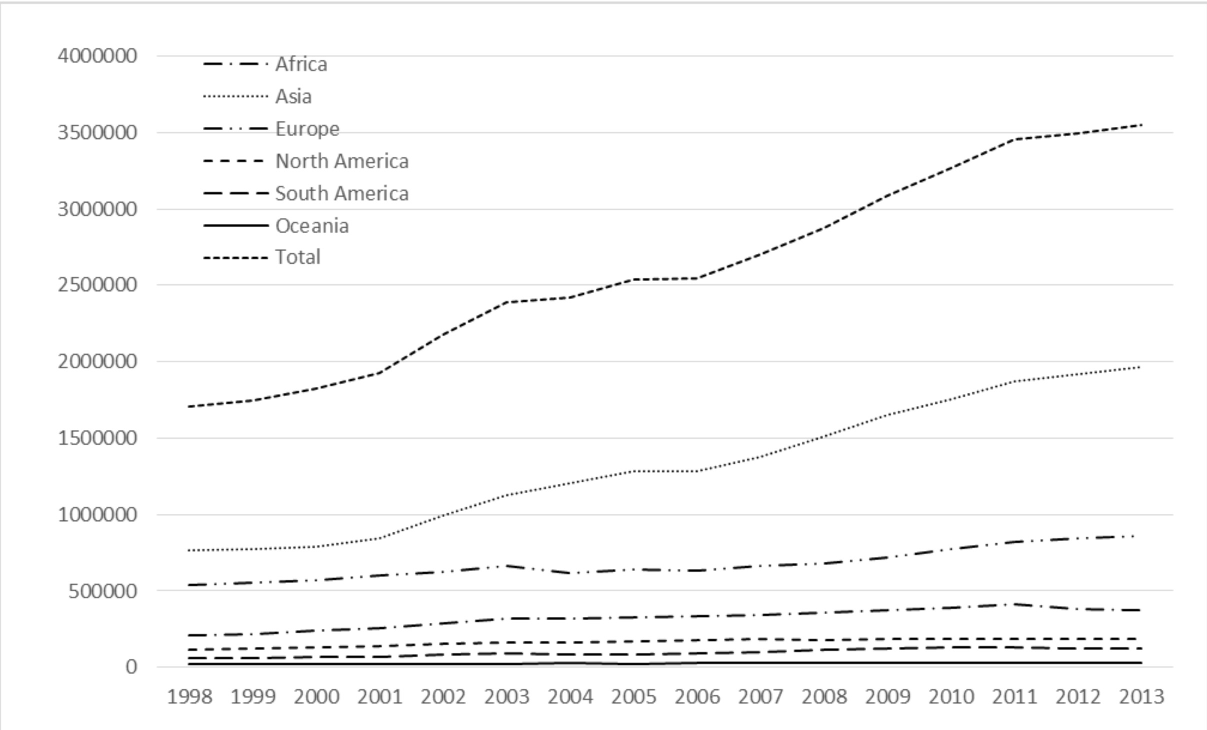

Figure 2: Total Number of Inbound Students for Selected Countries; OECD Data

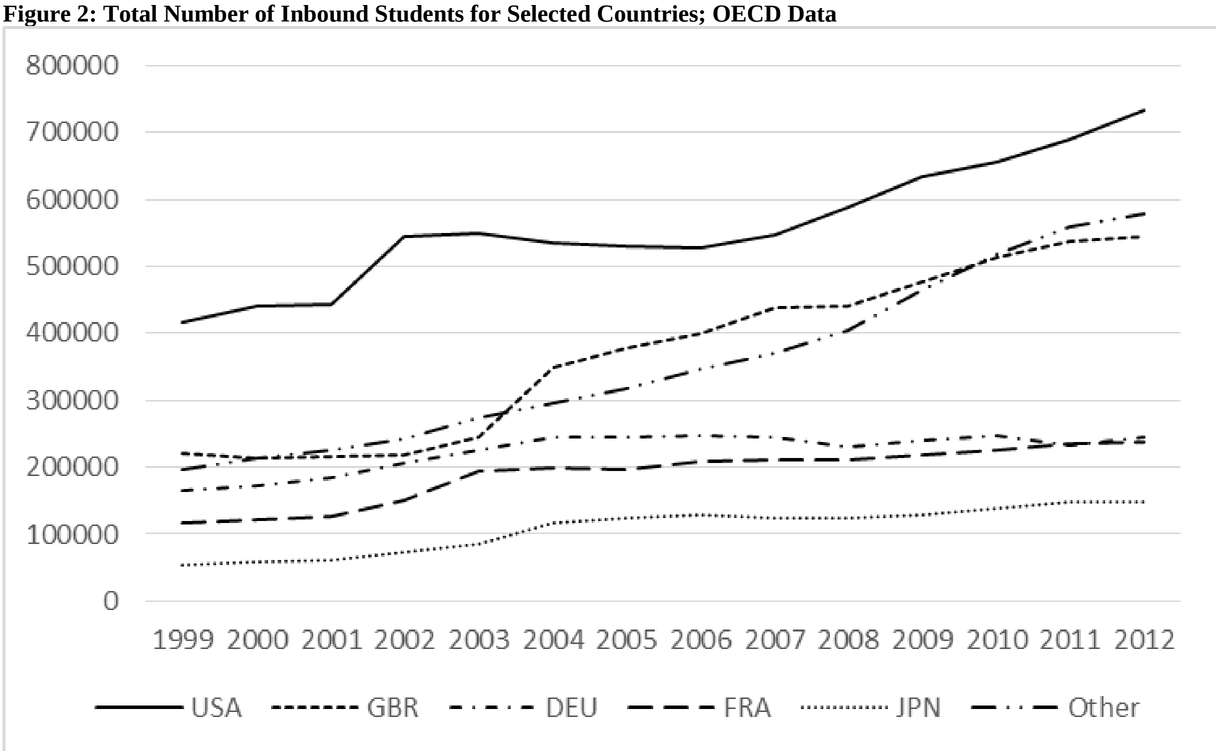

Figure 3: Time Series of GDP Growth and Enrollment Growth for Selected Countries

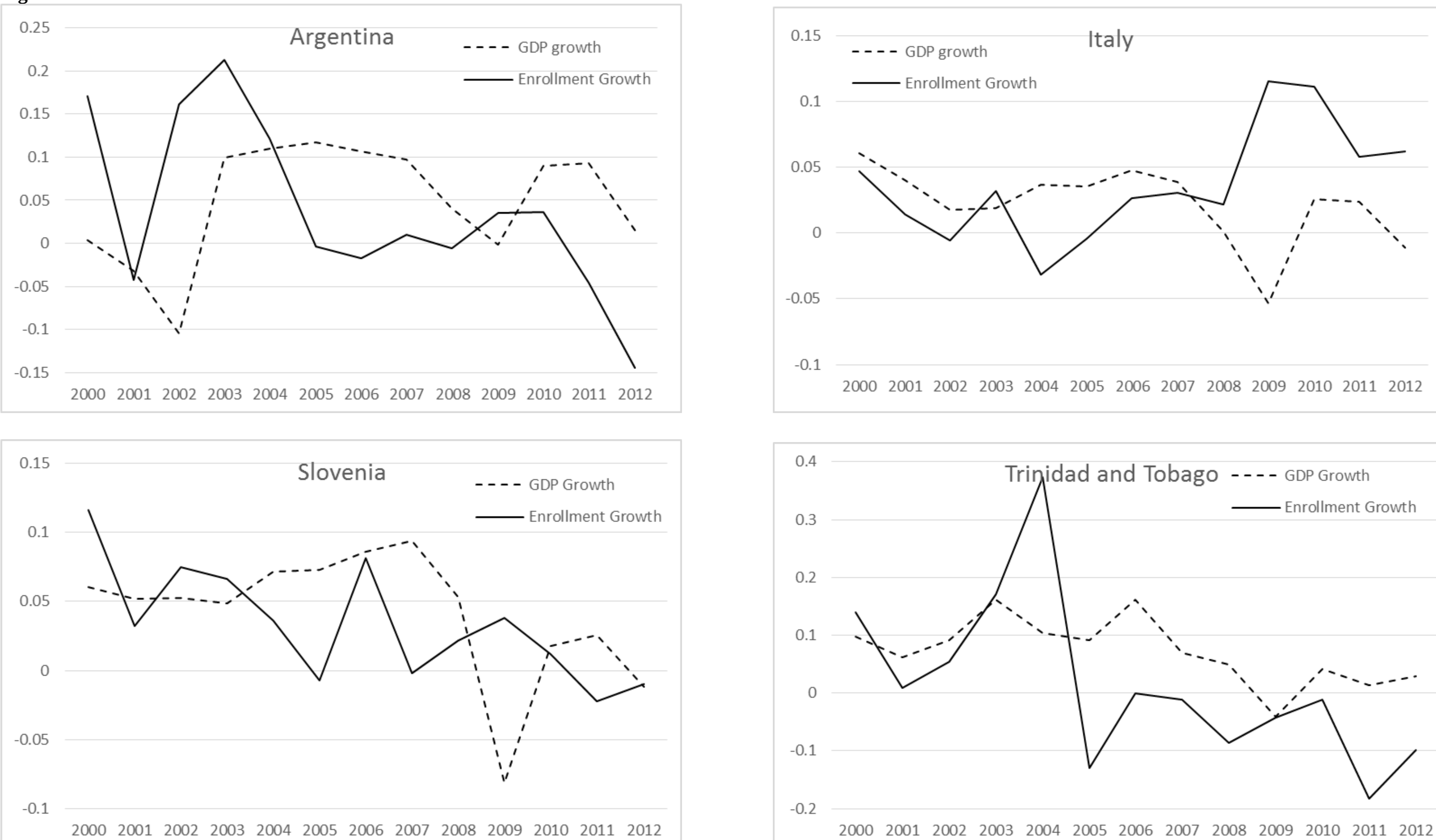

Table 1: Summary Statistics

Variable	Mean	Standard Deviation	Min.	Max.	Total Number
Country Pairs Per Recipient Country	49.895	37.982	5	146	163
Country Pairs Per Sending Country	5.816	4.410	1	18	19

Table 2: Univariate Regressions – Log Enrollment: OLS and WG Estimates

	(1) OLS	(2) OLS	(3) OLS	(4) WG	(5) WG	(6) WG
log Enrollment (-1)	0.988***	1.030***	1.036***	0.832***	0.791***	0.758***
	(0.003)	(0.062)	(0.053)	(0.024)	(0.066)	(0.064)
log Enrollment (-2)		-0.042	0.082		0.015	0.085**
		(0.061)	(0.054)		(0.046)	(0.040)
log Enrollment (-3)			-0.130***			-0.087***
			(0.024)			(0.018)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	12324	11376	10428	12324	11376	10428
Number of Groups				948	948	948
R-squared	0.98	0.98	0.98	0.81	0.79	0.74
F-statistic	22.92	24.53	30.82			
P-value	0.0001	0.0001	0.0000			

Clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1% significance level, respectively. The dependent variable is the natural logarithm of enrollment. The F-statistic refers to a test on the null hypothesis that the sum of the coefficients on the lagged dependent variables is equal to 1.

Table 3: Univariate Regressions – Log Enrollment: GMM Estimates

	(1) Diff-GMM	(2) Diff-GMM	(3) System-GMM	(4) System-GMM
log Enrollment (-1)	0.857***	0.814***	1.010***	1.048***
	(0.020)	(0.041)	(0.010)	(0.022)
log Enrollment (-2)		0.014		-0.043**
		(0.022)		(0.021)
Year dummies	Yes	Yes	Yes	Yes
Number of Observations	11376	10428	12324	11376
Number of Groups	948	948	948	948
Number of Instruments	90	88	103	101
Hansen J-test	0.0000	0.0000	0.0000	0.0000
Arellano-Bond test: AR(1)	0.0000	0.0000	0.0000	0.0000
Arellano-Bond test: AR(2)	0.0002	0.0290	0.0003	0.0008
Arellano-Bond test: AR(3)	0.2777	0.0998	0.2572	0.0691
Arellano-Bond test: AR(4)	0.1348	0.0123	0.2211	0.0136
Arellano-Bond test: AR(5)	0.0062	0.0009	0.0058	0.0013
Arellano-Bond test: AR(6)	0.0181	0.0014	0.0204	0.0012
Cragg-Donald F-Statistic	10.60	10.56		

All columns are estimated using a two-step estimator. Windmeijer (2005) standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1% significance level, respectively. The dependent variable is the natural logarithm of enrollment. Hansen denotes the p-value of the Hansen j-test of instrument validity. The rows labeled Arellano-Bond test: AR(i) contain p-values on the Arellano and Bond test serial correlation of order i in the error terms.

Table 4: Univariate Regressions – Growth Rate of Enrollment: OLS and WG Estimates

	(1) OLS	(2) OLS	(3) OLS	(4) WG	(5) WG	(6) WG
Growth Rate of Enrollment (-1)	0.001	0.002	0.002	-0.088***	-0.109***	-0.006
	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.008)
Growth Rate of Enrollment (-2)		0.001	0.002		-0.107***	-0.006
		(0.004)	(0.003)		(0.005)	(0.006)
Growth Rate of Enrollment (-3)			0.002			-0.006
			(0.002)			(0.006)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	11376	10428	9480	11376	10428	9480
Number of Groups				948	948	948
R-squared	0.00	0.00	0.02	0.01	0.02	0.02

Clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1% significance level, respectively. The dependent variable is the growth rate of enrollment.

Table 5: Univariate Regressions – Growth Rate of Enrollment: GMM Estimates

	(1) Diff-GMM	(2) System-GMM	(3) Diff-GMM	(4) System-GMM	(5) Diff-GMM	(6) System-GMM
Growth Rate of Enrollment (-1)	0.000	0.000	-0.034	0.000	0.001	0.001
	(0.001)	(0.001)	(0.209)	(0.001)	(0.004)	(0.003)
Growth Rate of Enrollment (-2)	0.000	0.000	-0.033	-0.001	0.003	0.003
	(0.003)	(0.002)	(0.210)	(0.002)	(0.005)	(0.005)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	9480	10428	9480	10428	9480	10428
Number of Groups	948	948	948	948	948	948
Number of Instruments	75	87	75	87	21	24
Hansen J-test	0.0000	0.0000	1.0000	1.0000	0.0661	0.2540
Arellano-Bond test: AR(1)	0.0329	0.0323	0.1144	0.0268	0.0373	0.0305
Arellano-Bond test: AR(2)	0.6669	0.5330	0.8632	0.2181	0.5911	0.8317
Cragg-Donald F-Statistic	74.80		74.80		479.74	
Clustered	No	No	Yes	Yes	Yes	Yes
Collapsed Instrument Set	No	No	No	No	Yes	Yes

All columns are estimated using a two-step estimator. Windmeijer (2005)-corrected standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1% significance level, respectively. The dependent variable is the growth rate of enrollment. Hansen denotes the p-value of the Hansen J-test of instrument validity. The rows labeled Arellano-Bond test: AR(i) contain p-values on the Arellano and Bond test serial correlation of order i in the error terms. Columns (3) to (6) allow for the error terms to be clustered on the level of the recipient country. Columns (5) and (6) additionally use an instrument set that has been collapsed in the way described in Appendix A.2.

Table 6: Cyclicality Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	WG	WG	WG	WG	WG	WG
GDP Growth	-0.230	-0.248	-0.312	-0.310	0.014	-0.010
	(0.257)	(0.276)	(0.316)	(0.303)	(0.129)	(0.127)
GDP Growth (-1)		0.128	0.161	0.261	-0.128*	-0.052
		(0.209)	(0.289)	(0.325)	(0.068)	(0.062)
GDP Growth (-2)			-0.388	-0.588	-0.004	-0.036
			(0.290)	(0.434)	(0.117)	(0.151)
GDP Growth (-3)				0.433	-0.124	-0.030
				(0.478)	(0.131)	(0.151)
GDP Growth (-4)					0.119	0.236
					(0.179)	(0.222)
GDP Growth(-5)						-0.401*
						(0.230)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	12324	12324	11376	10428	9480	8532
Number of Groups	948	948	948	948	948	948
R-squared	0.00	0.00	0.00	0.00	0.02	0.02
F-statistic	0.80	0.57	3.41	0.69	0.12	1.37
p-value	0.3823	0.4583	0.0812	0.4163	0.7379	0.2579

Clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1% significance level, respectively. The dependent variable is the growth rate of enrollment. The row labeled F-statistic refers to a test of the sum of the coefficients on all included lags of GDP Growth being equal to zero.

Table 7: Heterogeneity

	(1) OECD	(2) GDP	(3) Schooling	(4) Inflation	(5) Inequality	(6) Democracy	(7) Credit
Interaction	-0.222	-0.054	-0.160	0.031	0.015	-0.143	-0.012
	(0.291)	(0.110)	(0.277)	(0.213)	(0.139)	(0.098)	(0.115)
Interaction (-1)	0.048	0.249	0.164	0.017	-0.021	0.259	0.182
	(0.177)	(0.270)	(0.201)	(0.114)	(0.185)	(0.208)	(0.191)
Interaction (-2)	-0.382*	-0.462***	-0.487*	0.383**	-0.073	-0.444**	-0.250**
	(0.199)	(0.120)	(0.256)	(0.168)	(0.156)	(0.163)	(0.105)
Interaction (-3)	0.338	-0.050	-0.215	-0.006	0.091	-0.044	0.048
	(0.198)	(0.173)	(0.371)	(0.139)	(0.106)	(0.218)	(0.168)
Interaction (-4)	-2.054**	-1.207*	-1.661*	1.012	0.817**	-0.760	-1.180**
	(0.878)	(0.673)	(0.866)	(0.743)	(0.334)	(0.519)	(0.485)
Interaction (-5)	0.940*	0.653	0.920*	-0.430	-0.432*	0.448	0.658
	(0.514)	(0.639)	(0.522)	(0.550)	(0.234)	(0.488)	(0.439)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	8532	8469	7596	8478	7398	7893	8451
Number of Groups	948	941	844	942	822	877	939
R-squared	0.02	0.02	0.02	0.02	0.02	0.04	0.02
F-statistic	4.49	4.61	1.21	3.09	0.91	1.24	2.35
P-value	0.0480	0.0457	0.2854	0.0956	0.3525	0.2800	0.1426

Clustered standard errors in parentheses. All columns are estimated using the WG estimator and control for contemporaneous as well as 5 lags of GDP growth. *, **, and *** denote significance at the 10, 5, and 1% significance level, respectively. The dependent variable is the growth rate of enrollment. The rows labeled 'Interaction' contain the estimated coefficients of an interaction term between a dummy variable taking the value of 1 if the sending country is part of a certain group and GDP growth. Specifically, the dummy variables are defined as follows: The OECD dummy takes a value of 1 if a country is currently a member of the OECD. The Schooling dummy takes a value of 1 if the country's population in 1995 had average years of schooling above the median of our sample. All other dummy variables are defined as being equal to 1 when the average of a certain characteristic from 1995 to 1999 exceeds the sample median, whereby the characteristics are GDP per capita (column 3), the growth rate of the GDP deflator (column 4), the Gini coefficient (column 5), the index of institutionalized democracy (column 6), and the ratio of credit to the private sector over GDP (column 7). The F-statistic refers to a test of the sum of all interaction terms being equal to zero.

Table 8: The Role of Credit Constraints

	(1) Continuous	(2) Dummy	(3) Low Credit Ratio	(4) High Credit Ratio
Real Effective Exchange Rate Change	-0.008		0.004	-0.249
	(0.014)		(0.010)	(0.190)
Real Effective Exchange Rate Change (-1)	0.027		0.027*	0.138
	(0.017)		(0.015)	(0.157)
Real Effective Exchange Rate Change (-2)	0.017		0.009	0.106
	(0.012)		(0.100)	(0.064)
Real Effective Exchange Rate Change (-3)	0.046***		0.026*	0.303
	(0.014)		(0.013)	(0.193)
Appreciation (-3)		0.001		
		(0.016)		
Depreciation (-3)		-0.047***		
		(0.014)		
Year dummies	Yes	Yes	Yes	Yes
Number of Observations	5024	5024	1624	3400
Number of Groups	628	628	203	425
R-squared	0.00	0.00	0.01	0.00

Clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1% significance level, respectively. The dependent variable is the growth rate of enrollment. Both columns are estimated with the within groups estimator and control for contemporaneous as well as 5 lags of GDP growth. Appreciation and depreciation are dummy variables taking the value of one if the real effective exchange rate appreciated or depreciated by more than 15% in one year.

Table 9: Alternative Dataset: Different Estimators

	(1) WG	(2) OLS	(3) WG	(4) Diff-GMM	(5) System-GMM	(6) Diff-GMM	(7) System-GMM
GDP Growth	-0.061	0.057	-0.091	0.102	0.087	0.107	0.106***
	(0.059)	(0.056)	(0.059)	(0.072)	(0.058)	(0.066)	(0.040)
GDP Growth (-1)	-0.002	0.117**	-0.029	0.158**	0.118**	0.137**	0.117***
	(0.052)	(0.053)	(0.051)	(0.072)	(0.049)	(0.067)	(0.040)
GDP Growth (-2)	0.116	0.196**	0.108	0.194**	0.159**	0.140*	0.115*
	(0.084)	(0.079)	(0.084)	(0.089)	(0.075)	(0.075)	(0.062)
GDP Growth (-3)	-0.141	-0.109	-0.151	0.069	-0.044	0.079	-0.043
	(0.113)	(0.089)	(0.112)	(0.130)	(0.112)	(0.101)	(0.095)
GDP Growth (-4)	0.225***	0.242***	0.213**	0.283***	0.198***	0.278***	0.167***
	(0.086)	(0.086)	(0.082)	(0.077)	(0.067)	(0.062)	(0.062)
GDP Growth (-5)	-0.044	0.015	-0.029	0.133	0.065	0.168**	0.080
	(0.063)	(0.104)	(0.068)	(0.083)	(0.065)	(0.071)	(0.058)
Student Growth (-1)		0.003	-0.127**	-0.054	-0.008	-0.093	-0.042
		(0.058)	(0.058)	(0.079)	(0.070)	(0.087)	(0.077)
Student Growth (-2)		0.064	-0.049	0.010	0.038	-0.018	0.007
		(0.055)	(0.058)	(0.076)	(0.071)	(0.077)	(0.074)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reduced Instrument Set				No	No	Yes	Yes
No. of Observations	1716	1716	1716	1560	1716	1560	1716
No. of Groups	156		156	156	156	156	156
No. of Instruments				101	114	56	69
Hansen J-test				0.1277	0.1346	0.0771	0.1474
Arellano-Bond test: AR(1)				0.0000	0.0000	0.0000	0.0000
Arellano-Bond test: AR(2)				0.3226	0.4089	0.2211	0.2653
Cragg-Donald F-Stat				15.87		30.67	

Robust standard errors in parentheses. All GMM models were estimated using a two-step procedure and the Windmeijer (2005) correction. *, **, and *** denote significance at the 10, 5, and 1% significance level, respectively. The dependent variable is the contemporaneous growth rate of enrollment. Hansen denotes the p-value of the Hansen j-test of instrument validity. The rows labeled Arellano-Bond test: AR(i) contain p-values on the Arellano and Bond test serial correlation of order i in the error terms. In columns 6 and 7, only 3 lags are used as instruments.

Table 10: Heterogeneity – Alternative Dataset

	(1) OECD	(2) GDP	(3) Schooling	(4) Inflation	(5) Inequality	(6) Democracy	(7) Credit
Interaction	-0.414**	-0.342**	-0.084	0.177	0.308	0.014	-0.206
	(0.173)	(0.142)	(0.148)	(0.176)	(0.232)	(0.156)	(0.129)
Interaction (-1)	-0.179	-0.156	-0.265**	0.059	0.301**	-0.272*	-0.119
	(0.153)	(0.101)	(0.126)	(0.139)	(0.142)	(0.141)	(0.099)
Interaction (-2)	-0.284*	-0.246	-0.443**	0.155	0.696***	-0.166	0.030
	(0.169)	(0.154)	(0.174)	(0.164)	(0.234)	(0.155)	(0.159)
Interaction (-3)	-0.376**	-0.011	0.066	0.231	0.153	-0.101	-0.334
	(0.170)	(0.211)	(0.278)	(0.154)	(0.330)	(0.183)	(0.228)
Interaction (-4)	-0.259	-0.209	-0.381**	0.057	-0.095	-0.401***	-0.307*
	(0.226)	(0.158)	(0.171)	(0.157)	(0.244)	(0.142)	(0.167)
Interaction (-5)	0.226	-0.007	-0.037	-0.248*	0.470*	-0.071	-0.047
	(0.487)	(0.143)	(0.174)	(0.144)	(0.250)	(0.173)	(0.155)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	1716	1716	1408	1683	1298	1529	1683
Number of Groups	156	156	128	153	118	139	153
R-squared	0.07	0.07	0.07	0.07	0.08	0.07	0.07
F-statistic	4.99	9.75	7.50	1.09	9.63	8.26	8.35
P-value	0.0269	0.0021	0.0071	0.2982	0.0024	0.0047	0.0044

Robust standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1% significance level, respectively. The dependent variable is the growth rate of the number of outbound students. All columns are estimated using the WG estimator and control for contemporaneous as well as 5 lags of GDP growth. The groups are defined the same way as in table 6. The F-statistic refers to a test of the sum of all interaction terms being equal to zero.