

IZA DP No. 1612

**Student Flows and Migration:
An Empirical Analysis**

Axel Dreher
Panu Poutvaara

May 2005

Student Flows and Migration: An Empirical Analysis

Axel Dreher

*Thurgau Institute of Economics
and University of Konstanz*

Panu Poutvaara

*CEBR, Copenhagen Business School,
CESifo and IZA Bonn*

Discussion Paper No. 1612
May 2005

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Student Flows and Migration: An Empirical Analysis*

Using panel data for 78 countries of origin we examine the impact of student flows to the United States on subsequent migration there over the period 1971-2001. What we find is that the stock of foreign students is an important predictor of subsequent migration. This holds true whether or not the lagged endogenous variable is included. The relationship is robust to the inclusion of time and country dummies, and remains when we account for outliers. The basic results also hold for a cross section of 36 countries of origin and 9 host countries. Our results have important policy implications which we discuss in the last section.

JEL Classification: F22, I2, J61, O15

Keywords: migration, education, student flows, brain drain

Corresponding author:

Panu Poutvaara
Centre for Economic and Business Research (CEBR)
Copenhagen Business School
Porcelaenshaven, Bldg 65
DK-2000 Frederiksberg
Denmark
Email: panu.poutvaara@cebr.dk

* We are grateful to Julia Häring and Katarina Keller for useful comments. We also thank the EPRU Network for financial support, without implicating the sponsor for the views expressed.

I. Introduction

The determinants of migration have received a lot of attention in the economics literature, as evidenced in the review by Borjas (1994), among others. A consensus view is that both economic and cultural forces play an important role. Increased standard of living and the employment rate in the destination country tend to stimulate migration, while a higher standard of living and growth in the source country reduce push factors for emigration. Migration is easier between more similar countries, and especially a shared language plays an important role. There is also extensive evidence on the importance of existing networks of previous migrants (see Hatton 2003, Mitchell and Pain 2003, Clark et al. 2004 and Pedersen et al. 2004). On the positive side, the presence of compatriots eases the arrival of newcomers: the expatriate community helps finding jobs and apartments, and provides an existing social network. Here, there are decreasing returns to scale: once a critical mass of expatriate community is reached, further increases in its size no longer similarly ease further migration. On the negative side, large-scale immigration may generate a backlash, thus making integration of new migrants more difficult. Due to these counteracting effects, it is usually posited that the effects of existing immigrants stocks on the flow of new immigrants are initially positive, but that the effects tend to reduce in the size of the immigrant community, and may finally turn negative.

Migration studies have tended to focus on permanent migration, but a lot of migration flows are temporary – for example, a large number of students do part of their studies abroad. According to Tremblay (2002), the world's increasing integration leads to rising internationalization of educational courses. As a consequence, countries encourage their students to study abroad – benefiting from their cost-free education on their return. On the other hand, host countries are interested in educating foreign students, partly to attract human capital, benefiting the domestic economy.

In 2002, the number of foreign students in the OECD countries was 1.9 million, being on average 5.3 percent of all students in the OECD countries (OECD 2003).¹ These migration flows are highly concentrated: 71 percent of foreign students go to the United States, The United Kingdom, Germany, France or Australia (Tremblay 2002). The share of foreign students in total student population is highest in Switzerland (17 percent), Australia (14 percent), Austria (12 percent) and the United Kingdom (11 percent). In the United States, the

¹ Different countries use different definitions as to who are foreign students. Most European countries use as criterion foreign nationality or citizenship, thus including permanent residents and their children. The United Kingdom excludes permanent residents. The United States excludes permanent residents and refugees (Tremblay 2002). For the purpose of our study, the U.S. definition is most relevant.

share of foreign students is 3.5 percent. However, 16.6 percent of foreign students in the United States are at the advanced research level, making an important contribution to American Ph.D. programs (Tremblay 2002). In the OECD countries, the proportion of foreign students rose by 30.4 percent between 1995 and 2001.

An important question is who – in addition, of course, to the student – gains from students being educated abroad. Do foreign students significantly increase the amount of permanent human capital in the host countries, resulting in brain drain from the countries of origin or does the majority of those students return home? The question has important policy implications both for the host country and the country of origin. That some of the foreign students stay in their country of study is, in fact, well known (Tremblay 2002). According to the U.S. Immigration and Naturalization Service (2000), out of a sample of 4200 migrants holding temporary permission to work in the United States (H1B visa) 23 percent previously held a student visa. Clearly, there are good reasons to expect that the experience of having studied in a country considerably eases subsequent migration. In 2000, several OECD countries changed their legislation allowing a larger number of foreign students to stay, in order to alleviate shortages of skilled labor. In addition to their direct economic contributions, the presence of foreign students may also significantly invigorate the academic life of the destination country. The destination country might even benefit from economies of scale in the education system (Tremblay 2002).

Although there are some attempts to evaluate how many foreign students stay in the country they study in, we are not aware of any empirical research incorporating student flows into a more general model explaining permanent migration. As Tremblay (2002: 5) puts it, “the scale of this phenomenon is extremely difficult to ascertain” – but this is what our paper aims to do. Specifically, we try to fill the gap in the empirical literature on the effects of foreign student flows on subsequent migration flows. We take as our starting point previous studies estimating migration, most notably, Hatton (2003), Mitchell and Pain (2003), Clark et al. (2004) and Pedersen et al. (2004). These studies do not take into account the role of student flows. As in these studies, we account for push and pull factors of migration, adding the effects of studying in the destination country into the list of pull factors. Our hypothesis is that having studied previously in the prospective destination country significantly eases migration.

Due to restrictions concerning data availability, we can only estimate cross-section regressions when analyzing worldwide migration. However, there are time series data on migration to the United States. Thus, our main focus is on understanding the role of student

flows when explaining migration to the United States. In addition to being of great interest in its own right, the United States is the largest recipient of international students. This also renders the results with the U.S. data more interesting to a wider audience. Understanding the reasons behind the success of the United States in attracting skilled migrants is valuable for European countries aiming to do the same. To enable comparison, we also present a cross-section of migration to OECD countries in addition to our panel data analysis on migration to the United States.

Our paper is organized as follows. The next section presents our model. We then introduce the data and estimate the relationship between migration and student flows empirically. The final section concludes and draws policy implications.

II. The Model

When choosing where to live and work potential migrants compare both material consumption and other aspects of well-being, like social networks. Following Zavodny (1997) and Pedersen et al. (2004), we write individual k 's expected utility in country j in year t as

$$U_{ijkt} = U(S_{ijkt}, D_{ijk}, X_{ikt}, X_{jkt}) \quad (1)$$

where S_{ijkt} is a vector of characteristics that affect individual k 's utility of living in country j in year t , given that the individual lived in country i in year $t-1$. This vector includes information on individual k 's home country: it is reasonable to assume that most people have a home country preference, for example due to the existence of social networks. This vector also includes the effects of previous migrants from k 's home country in other countries. In addition to the stock of previous migrants, taken into account in the bulk of previous literature, we also include in this vector previous student flows from k 's home country to country j .

Vector D captures migration costs. For example, migration costs tend to be higher if the source country is land-locked, as such countries are often less well connected, and do not allow sea route connections elsewhere. In our analysis of migration to the United States, we also include in D the distance from the source country to the United States, and the percentage of source country population that speak English. Note that the stock of previous migrants could enter also into migration cost, especially considering psychological costs. We have chosen to include it into S as Pedersen et al. (2004), but the results would be the same if it were included, instead, in D as we later adopt a linear specification of individual utility.

The economic push and pull factors, most notably the GDP per capita ratio between the source and the destination country and the ratio of the average schooling between the two, are included in the X vectors. In some specifications, we also include here a variable

measuring democracy in the source country, as well as GDP growth and unemployment in the destination country.

We assume individual utility to be linear in the relevant characteristics. Including an idiosyncratic error ε , the utility function reads as:

$$U_{ijkt} = \alpha_1 S_{ijkt} + \alpha_2 D_{ijk} + \alpha_3 X_{ikt} + \alpha_4 X_{jkt} + \varepsilon_{ijkt}, \quad (2)$$

where α_1 , α_2 , α_3 , and α_4 are vectors of the parameters of interest that we estimate. When choosing where to migrate, an individual picks country j which results in the highest expected utility.

We estimate below migration in two different frameworks. We first study migration flows to nine OECD countries, and then migration flows to the United States. In the first specification, the probability of choosing country j among the OECD countries is

$$\Pr(j_{kt} / i_{kt-1}) = \Pr[U_{ijkt} = \max(U_{ki1t}, U_{kt2t}, \dots, U_{knt})], \quad (3)$$

where n is the total number of countries. Due to data constraints, we do not have observations for all countries, limiting us to a subset of the nine OECD destination countries below.

When studying migration flows to the United States, we only compare utility available in the United States to that available in the home country. An interpretation here, as well as in the analysis of migration to the nine OECD countries, would be that we exclude from the analysis those who have decided to migrate elsewhere, and focus only on those choosing whether to stay in the home country or migrate to the destination country. For them, migration to the United States takes place if

$$U_{kiUSA_t} > U_{kii_t}.$$

In the empirical specification, we normalize migration flows by dividing gross flows by the population of the source country. This gives us a model to evaluate migration to the nine OECD countries, parallel to Pedersen et al. (2004),

$$m_{ijkt} = \beta_1 S_{ijkt} + \beta_2 D_{ijk} + \beta_3 X_{ikt} + \beta_4 X_{jkt} + \mu_{ijkt}, \quad (4)$$

where m is the log of immigrants admitted, relative to source country population, and coefficients β are related to but not identical with coefficients α , the same holding for the error term μ . The empirical specification for migration to the United States is

$$m_{iUSAkt} = \gamma_1 S_{iUSAkt} + \gamma_2 D_{iUSAk} + \gamma_3 X_{ikt} + \gamma_4 X_{USAkt} + \nu_{iUSAkt},$$

with corresponding relationship between the parameters of interest γ and underlying parameters in the utility function α . We also estimate the equation

$$m_{iUSAkt} = \gamma_1 S_{iUSAkt} + \gamma_2 D_{iUSAk} + \gamma_3 X_{ikt} + \gamma_4 X_{USAkt} + m_{iUSAkt-1} + \mu_{iUSAkt}, \quad (5)$$

with the lagged endogenous variable included – thus taking into account that migration follows a rather stable pattern over time.

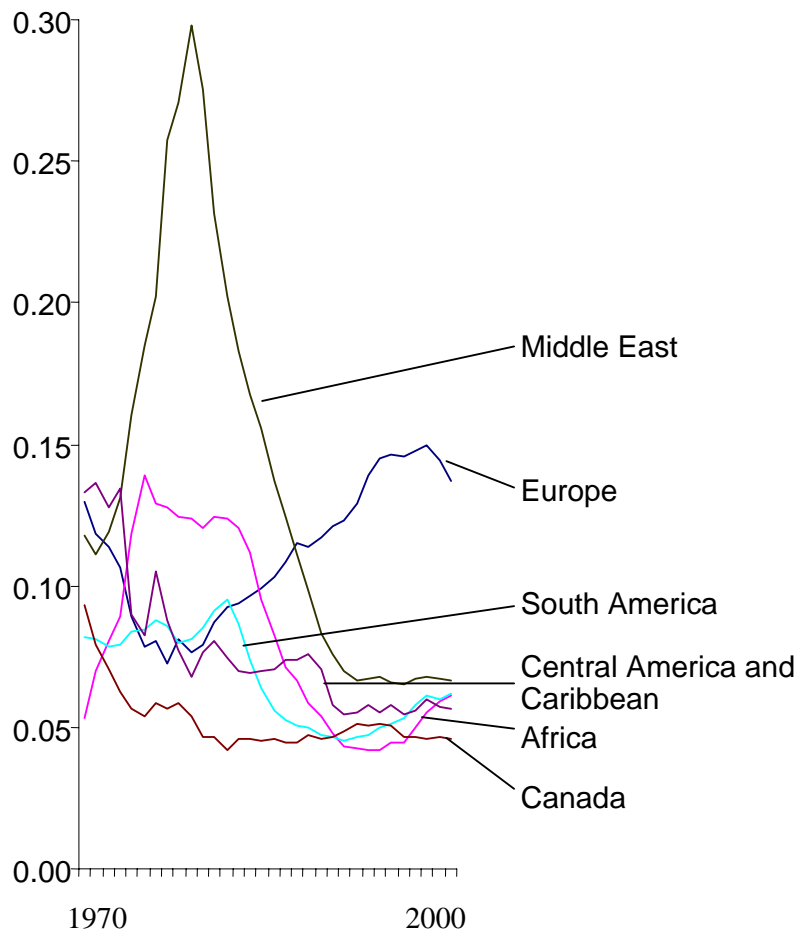
III. Data

The migration data for our cross-section analysis is taken from the OECD's (2001) International Migration Statistics. Data are provided for 94 origin countries. The destination countries covered are Australia, Belgium, Canada, Denmark, France, Germany, Japan, Luxembourg, the Netherlands, Norway, Sweden, Switzerland, the United Kingdom and the United States. The number of foreign students enrolled by country of citizenship is from the OECD's (2003) Education Database. In most cases foreign students are identified on the basis of citizenship, in some, however, other criteria – like place of birth or former domicile – apply.²

The data for the number of foreign students in the United States is taken from the Institute of International Education Network. As a nonimmigrant class of admission, the data refer to foreigners coming temporarily to the United States to pursue studies in an institution of higher education. Graph 1 shows the regional distribution of where the foreign students in the United States came from over the period 1970-2001. As can be seen, there is some variation in the composition of source countries over time. Overall, the share of students coming from the Middle East decreased from 12 to 7 percent over the last 30 years, with the hike during the seventies mainly being due to a surge in the number of Iranian and Saudi Arabian students. The share of students coming from Central America and the Caribbean fell by almost half from its 13 percent at the beginning of the seventies. The same is true for the share of Canadian students, falling from 9 to 5 percent.

² See Tremblay (2002) for an in-depth discussion and the potential bias arising from this.

Graph 1: Foreign Students in the U.S. (1970-2001), percent



The data for the number of immigrants admitted to the United States is provided by the U.S. Immigration and Naturalization Service (INS) Statistical Yearbooks.³ The data refers to all legal immigration, including refugees but does not include foreign students. The data employed here classify immigrants according to their birth country and not the country of last residence. They refer to both those who applied from abroad and those who changed to permanent status while already being in the United States.⁴

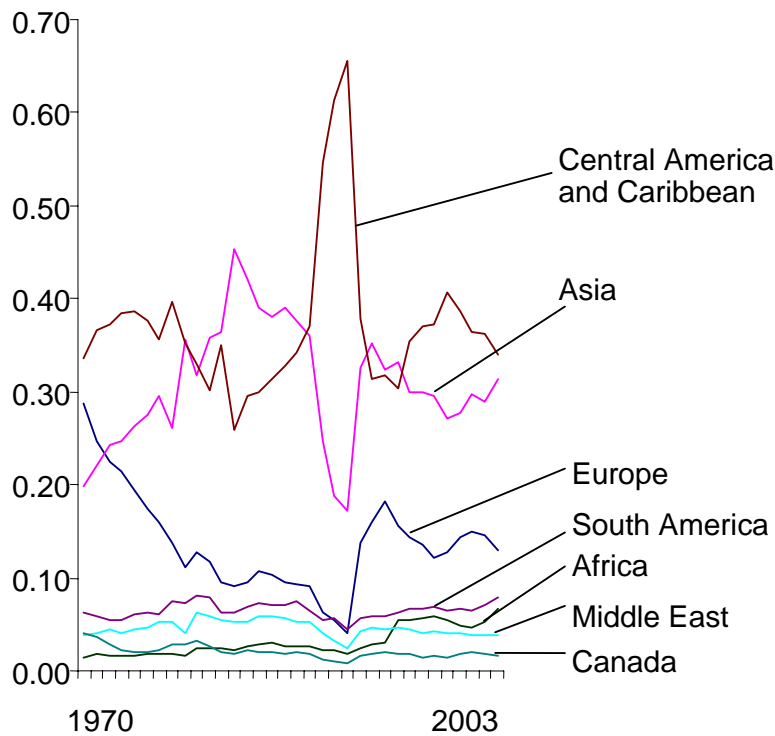
Graph 2 shows the regional composition of immigrants to the United States over the period 2000-03. As can be seen, the United States has experienced rising immigration since the 1970s. The graph shows that the composition of immigrants changed to some extent over the last 30 years. Whereas the proportion of immigrants from Europe fell from almost 30

³ We thank Timothy J. Hatton for providing these data.

⁴ For a more detailed description of the immigration data see Clark, Hatton and Williamson (2002).

percent to well below 15 percent in 2003, the share of those from Asia rose from 20 percent to over 30 percent. The share of those from Africa rose from 1 to 7 percent; while the share of those from the Middle East, South America, Central America and the Caribbean and Canada remained more or less the same. The large increase in the share of immigrants from Central America and the Caribbean at the beginning of the nineties is mainly due to a surge in Mexican immigration.

Graph 2: Migration to the U.S. (1970-2003), percent



Most important for our study, of course, is the relationship between the number of foreign students and immigration. In fact, correlation between those two flows is quite high, with 0.36 for the OECD sample and 0.67 for the United States (with both variables expressed as a share of source country population and in logarithms). We elaborate on this relationship in the next sections.

IV. Empirical Specification and Results for a cross section of OECD countries

Our selection of variables follows the analyses of Fertig (2001), Clark, Hatton and Williamson (2002) and Hatton (2003). The dependent variable is the log (immigrants admitted/source country population) ratio.⁵ The source country's GDP per capita relative to per capita GDP in the country of destination is included to capture expected relative improvements in the income of immigrants. The lower is per capita GDP in the country of origin relative to the destination country, the more attractive is immigration. We also include the ratio of average years of schooling in the source country relative to the destination country as a proxy for relative human capital stocks. By controlling for the stock of human capital, the income variable reflects the relative return on human capital instead of amount and average return (Clark, Hatton and Williamson 2002). We expect relative schooling years to have a positive impact on migration. The costs of migration depend on whether the source country is landlocked. The costs of migration are also assumed to depend on the stock of migrants already living in the destination country. The variables 'stock of migrants' and 'stock of migrants squared' capture the combined impact of friends and relatives already living in the country, and possible negative responses by the native population.⁶ On the positive side, for example, a partner having studied abroad might facilitate the other one's immigration also. The host country might give them preference for their linguistic mastery, their familiarity with the host country's social context and the lifting of entry barriers by possessing a degree "known" to employers (Tremblay 2002). The negative reactions might include increased competition in the labor market, more restrictive immigration policies, and negative sentiment towards immigrants and foreigners in general. All variables, their precise definitions and data sources are listed in the appendix.

Table 1 reports the results for a cross section of 36 source countries and 9 host countries.⁷ Due to missing data, our analysis, however, has only 93 observations. Unfortunately, the number of students studying in the OECD countries is only available for the period 1998-2000;⁸ migration flows, however, are not available after 1995. With this in mind, and the number of observations being only 93, the results of the analysis can merely be suggestive.

⁵ Note that population is measured in thousands to obtain smaller coefficients.

⁶ Ideally, one would like to control for the stock of previous migrants in specific cities rather than countries and relate them to migrant flows to those cities. We do not have the data for this exercise.

⁷ Our selection of countries is driven by data availability. The countries included in our sample are listed in Appendix C. As we focus on the United States in the next section, we do omit the United States as destination country here.

⁸ For a detailed description of these data see Tremblay (2002).

Table 1: Migration to OECD, (cross section, OLS)

Explanatory Variables	(1)	(2)
log (foreign students/ origin population)	0.44 (4.90***)	0.45 (5.05***)
GDP per capita ratio (origin/destination)	-0.85 (2.14**)	-0.80 (1.70*)
Schooling years ratio (origin/destination)	1.11 (1.82*)	1.02 (1.70*)
Landlocked, dummy	0.18 (0.24)	
Immigrant stock/ origin population	103.11 (1.82*)	30.54 (1.55)
Immigrants/ population squared	2800.37 (1.38)	
Number of host countries	9	9
Number of source countries	36	36
Number of observations	93	93
R ²	0.29	0.28

Notes:

dependent variable: log (immigrants admitted/source country population)

't'-statistics in parentheses:

***, **, * significant at the 1, 5 and 10 percent levels respectively

As can be seen in column 1, with a coefficient significant at the five percent level, migration flows are lower with a higher per capita GDP in the source country relative to the country of immigration. At the ten percent level of significance, migration flows are higher with a higher ratio of years in school and a higher stock of immigrants already living in the destination country. The results also show that student flows and migration are significantly related, with a highly significant coefficient. Squared immigration and the dummy for landlocked countries do not significantly influence migration. Column 2 thus omits those insignificant variables. The result for student flows remains. As the corresponding coefficient shows, a one percent increase in student flows is associated with about half a percent increase in immigration.

In the next section we further investigate the issue using data for one country of destination only – the United States – for which more extensive data is available.

V. Empirical Specification and Results for the United States

This section analyzes migration with the United States as the destination country. As data availability is much better for the United States, we can test our hypotheses more rigorously using cross-section time-series (panel) data of immigrants from a maximum of 78 source countries.⁹ The annual data cover the years 1971-2001. Since some of the data are not available for all countries or years, the panel data are unbalanced and our number of observations depends on the choice of explanatory variables. Since there was significant first-order autocorrelation in all models, the disturbance term is modeled as an AR(1) process.

Again, the dependent variable is the log (immigrants admitted/source country population) ratio. In addition to the variables of Table 1, we include the distance from the source country to the United States and the percentage of source country population that speak English. The costs of migration rise with distance from the United States. They are lower for English speaking people.¹⁰

Column 1 of Table 2 shows the results (of our random effects estimator). As can be seen, all coefficients are significant at the five percent level at least. In line with our hypotheses, the number of migrants to the United States rises with lower relative GDP and relatively longer education. More immigrants come from countries that are geographically closer to the United States, have a higher share of native English speakers, and have access to the sea.¹¹ The stock of immigrants already living in the United States initially attracts more people from the same country. However, as indicated by the negative coefficient of the squared stock variable, this impact reverses eventually. The results also show that immigration from a specific country is more likely, the more students from that country studied in the United States in the previous year. The coefficient of this variable is significant at the one percent level and is quantitatively important. According to the estimate of column 1, a ten percent increase in student flows leads to approximately one percent more immigration. For comparison, an increase in relative income by 10 percent reduces immigration by slightly more than 5 percent, and an increase in relative years of schooling by the same percentage increases immigration on average by almost 10 percent. A hundred mile

⁹ Our selection of countries is again driven by data availability. The countries included in our sample are listed in Appendix C.

¹⁰ We did not include these variables in the cross country regressions of the previous section as we do not have the data for all countries in the sample.

¹¹ As distance from the United States and the dummy for landlocked countries both measure travel costs, we also included them in the regression one at a time. The results are unchanged.

increase in the distance between the origin country and the United States reduces immigration by 4 percent.¹²

Table 2: Migration to the United States, (1971-2001, OLS AR1)

Explanatory Variables	(1)	(2)	(3)	(4)
log (foreign students/ origin population), t-1	0.09 (4.79***)	0.04 (6.47***)	0.03 (1.87*)	0.04 (3.46***)
GDP per capita ratio (origin/U.S.)	-0.53 (3.23***)	-0.07 (3.18***)	-0.17 (0.64)	-0.21 (2.07**)
Schooling years ratio (origin/U.S.)	0.92 (2.88***)	0.002 (0.04)	-0.67 (1.09)	-0.02 (0.12)
Distance from U.S.	-0.0004 (8.18***)	-9.25e-06 (2.64***)		
Landlocked, Dummy	-1.05 (3.52***)	-0.004 (0.17)		
English speaking (percent of population)	0.01 (2.61***)	0.0004 (1.29)		
Immigrant stock (t-1)/ origin population (t-1)	12.12 (4.82***)	1.51 (2.67***)	0.30 (0.09)	1.30 (1.65*)
Immigrants/ population (t-1) Squared	-29.25 (3.92***)	-5.20 (2.17**)	-3.36 (0.41)	-4.46 (1.79*)
Lagged Endogenous Variable		0.94 (140.99***)		0.83 (64.30***)
Number of countries	78	78	78	78
Number of observations	1903	1881	1825	1803
Country/ Year dummies	No	No	Yes	Yes
R ² (overall)	0.57	0.98	0.04	0.98

Notes:

dependent variable: log (immigrants admitted/source country population)

't'-statistics in parentheses:

***, **, * significant at the 1, 5 and 10 percent levels respectively

In column 2, we add the lagged endogenous variable. As a consequence, the estimated impact of students on migration becomes quantitatively smaller, with a ten percent increase in student flows leading to about half a percent more immigration. The coefficient stays,

¹² If we drop the share of English speaking population and distance from the U.S. (and so exactly replicate the specification of Table 1), the results are unchanged. We also included the squared distance to test for non-linearities. Indeed, the squared term is significant at the one percent level, with a positive coefficient. The coefficient of the share of English speaking population is no longer significant, while all other results remain.

however, significant at the one percent level. The lagged dependent variable is highly significant and the share of the dependent variables' variation explained by the regression rises substantially. With the inclusion of the lagged endogenous variable, relative schooling years no longer significantly affect immigration. The dummy for landlocked countries also loses its significance as does the percent of the population speaking English.

As a hausman test rejects the random effects model in favor of fixed effects and we found significant period effects in all specifications, columns 3 and 4 replicate the analysis including those dummies.¹³ As a consequence, we cannot employ variables that do not vary over time or across countries. Language, distance, and the share of population speaking English have to be omitted. The results show that most coefficients are insignificant when time and country dummies are included. Most important for our analysis, however, the coefficient of student flows stays significant at least at the ten percent level when the lagged endogenous variable is excluded and at the one percent level otherwise.

In order to further investigate the relationship between immigration and foreign students, Table 3 includes up to three lags of the foreign students variable in the regressions. Arguably, students are more likely to immigrate permanently to the United States at the end of their studies. Unfortunately, we do not have the data to test this hypothesis. The results show that the number of students studying in the United States in the previous year always increases immigration, while the coefficient of the second lag is only significant with the lagged dependent variable included. The third lag is insignificant regardless of the inclusion of the lagged endogenous variable. We also included the lags one at a time instead of in addition (not reported in the table). When only the third lag is included, its coefficient is insignificant when the lagged endogenous variable is excluded, and significant at the five percent level otherwise. The same is true for the second lag. Clearly, we have problems with collinearity here, so the results have to be interpreted with caution.

In the next section, we further test the robustness of our results.

¹³ The coefficients of the country and time dummies are not reported in the tables.

Table 3: Migration to the United States (1971-2001, OLS AR1)

Explanatory Variables	(1)	(2)	(3)	(4)
log (foreign students/ origin population), t-1	0.04 (2.37**)	0.06 (3.23***)	0.04 (2.30**)	0.06 (3.32***)
log (foreign students/ origin population), t-2	-0.001 (0.07)	0.05 (2.55**)	0.01 (0.31)	0.06 (3.22***)
log (foreign students/ origin population), t-3			-0.01 (0.55)	0.02 (0.85)
GDP per capita ratio (origin/ U.S.)	-0.19 (1.82*)	-0.14 (0.51)	-0.18 (1.56)	-0.17 (0.59)
Schooling years ratio (origin/ U.S.)	-0.04 (0.28)	0.75 (1.13)	-0.02 (0.09)	-0.76 (1.12)
Immigrant stock (t-1)/ origin population (t-1)	1.18 (1.55)	-0.07 (0.02)	1.32 (1.53)	0.07 (0.32)
Immigrants/ population (t-1) Squared	-4.34 (1.69*)	-2.14 (0.90)	-4.67 (1.76*)	-2.60 (0.32)
Lagged Endogenous Variable	0.82 (60.95***)		0.81 (56.96***)	
Number of countries	78	78	78	78
Number of observations	1753	1754	1682	1683
Country/ Year dummies	Yes	Yes	Yes	Yes
R ² (overall)	0.98	0.13	0.98	0.18

Notes:

dependent variable: log (immigrants admitted/source country population)

't'-statistics in parentheses:

***, **, * significant at the 1, 5 and 10 percent levels respectively

VI. Further Discussion

This section tests the robustness of our results with respect to inclusion of additional explanatory variables and in terms of the observations included in the regressions. First, regarding additional variables, we follow the previous literature to identify candidate variables. Mitchell and Pain (2003) suggest including the unemployment rate and economic growth in the United States. Arguably, people are more likely trying to escape countries where political participation is weak. We therefore also include a variable measuring whether or not the source country is a democracy.

When it comes to how helpful previous immigrants are in finding a job or apartment, it might be the absolute number of immigrants that matters or migrants relative to the

population in the destination rather than the origin country. The same is true for backlash against immigrants. We therefore replace the number of immigrants relative to source country population by their number relative to U.S. population.

Second, and equally important, we focus on the role of outliers. Some authors suggest using so-called robust estimation techniques to deal with them. Robust estimators can be thought of as trying to identify that part of the data best approximated by the model being estimated (see de Haan and Sturm 2003 for an illustrative discussion). We therefore employ the Least Median of Squares (LMS) estimator introduced by Rousseeuw (1984) to test the robustness of our results, minimizing the median of the squared residuals. This estimator will not be influenced by observations lying outside the typical relationship between the dependent and explanatory variables revealed by the remaining data.

Table 4 contains the results. In column 1, we include the POLITY IV measure of democracy. Column 2 includes the rate of unemployment in the United States, column 3 GDP growth. In column 4, the stock of migrants in the United States is divided by U.S. population instead of origin country population. The same is true for the stock of migrants squared. As the table shows, only the unemployment rate is significantly associated with immigration to the United States. As expected, immigration is lower with higher unemployment. Dividing the stock of migrants by U.S. population rather than origin country population does not change the results. The inclusion of the additional variables has no impact on the relationship between foreign student enrollment and immigration. In all four regressions immigration is increasing in the number of foreign students in the previous year, with a coefficient significant at the one percent level.

In column 5 we report the results of the Least Median of Squares regression. As can be seen, the previous results have not been driven by outliers. All coefficients are significant at the one percent level and have the expected sign. In the robust regression, the magnitude of foreign student enrollment corresponds closely to the results of the cross section for the OECD, indicating that a one percent increase in student flows is associated with about half a percent increase in immigration. As we cannot control for fixed country effects in both the cross section and the LMS regressions, these larger effect most likely represents the pure between-country variation.

Table 4: Migration to the United States (1971-2001)

Explanatory Variables	(1)	(2)	(3)	(4)	(5)
log (foreign students/ origin population), t-1	0.08 (3.94***)	0.09 (4.86***)	0.09 (4.81***)	0.09 (4.47***)	0.49 (23.74***)
GDP per capita ratio (origin/ U.S.)	-0.52 (3.07***)	-0.54 (3.28***)	-0.53 (3.22***)	-0.55 (3.15***)	-1.35 (18.26***)
Schooling years ratio (origin/ U.S.)	0.62 (1.84*)	0.89 (2.76***)	0.93 (2.89***)	0.94 (2.77***)	1.54 (10.70***)
Distance from U.S.	-0.0004 (8.29***)	-0.0004 (8.17***)	-0.0004 (8.18***)	-0.0004 (7.98***)	-0.0002 (15.88***)
Landlocked, Dummy	-0.98 (3.24***)	-1.05 (3.52***)	-1.05 (3.52***)	-1.07 (3.06***)	-0.27 (3.45***)
English speaking (percent of population)	0.01 (2.87***)	0.01 (2.64***)	0.01 (2.61***)	0.01 (2.74***)	0.004 (3.82***)
Immigrant stock (t-1)/ origin population (t-1)	11.59 (4.05***)	12.09 (4.80***)	12.03 (4.77***)		41.42 (22.15***)
Immigrants/ population (t-1) Squared	-24.71 (2.95***)	-29.36 (3.93***)	-28.99 (3.88***)		-175.96 (20.77***)
Democracy	-0.004 (0.59)				
U.S. unemployment rate (t-1)		-0.02 (2.91***)			
U.S. GDP growth (t-1)			0.002 (0.86)		
Immigrant stock (t-1)/ U.S. population (t-1)				104.04 (2.07**)	
Immigrants/ U.S. population (t-1), squared				-2936.71 (1.83*)	
Number of countries	75	78	78	78	78
Number of observations	1723	1903	1903	1903	1903
Country/ Year dummies	No	No	No	No	No
Method of Estimation	OLS AR(1)	OLS AR(1)	OLS AR(1)	OLS AR(1)	LMS
R ² (overall)	0.57	0.57	0.57	0.49	0.53

Notes:

dependent variable: log (immigrants admitted/source country population)

't'-statistics in parentheses:

***, **, * significant at the 1, 5 and 10 percent levels respectively

VII. Conclusion

In this paper, we discover a close link between student flows and migration flows. In our cross-section analysis of migration to nine OECD countries, we found that the share of foreign students was highly significant. In our panel (OLS) regressions for the United States, a ten percent increase in student flows leads to an increase in immigration of between 0.3-0.9 percent. Clearly, we cannot be sure whether it is the students actually staying in their host country or the impact of other student-related factors driving those results. As one example, students may tell others about their experiences and have networks that other migrants may use.

Student flows explain migration more consistently in different specifications than traditionally highlighted economic variables - like the GDP per capita ratio between the country of origin and the United States and the schooling years ratio between the two. When country and time dummies are included, the stock of students also generally retains its significance, and seems to be a more robust predictor than the stock of previous immigrants or its square. When measures of the percentage of English-speaking population, democracy in the source country, the unemployment rate, or GDP growth in the United States are included, both the stock of foreign students and that of previous immigrants and its square are significant at the one percent level.

One potential problem in explaining migration arises from the omitted variable bias. It is conceivable that both student flows and migration flows would reflect an underlying cultural and economic proximity, not captured by distance or sharing the English language when explaining migration to the United States. To reduce the impact of this possibility, we also include the lagged endogenous variable - migration in the previous year - to our panel analysis. Indeed, the lagged endogenous variable turns out to be the most powerful explanatory variable of current migration. Importantly, the past stock of foreign students still retains its significance in all cases. Repeating the analysis using the Least Median of Squares estimator suggests that the results are not driven by outliers.

Our findings have several policy implications for both origin and host countries of foreign students. In our OLS regressions for the United States, a ten percent increase in student flows leads to an increase in immigration of between 0.3-0.9 percent. This suggests that hosting foreign students is an efficient way of attracting future migration, independent of the previous immigrant stock. Therefore, European countries aiming to attract immigrants from developing countries might benefit from educating students from these countries. Governments would then be well advised to take this into account when considering tuition

fees for foreign students. The trade-off arising from the fees is the following. On the benefit side, charging tuition fees gives additional resources to the host country, and such resources could be used to improve the financing of universities. On the cost side, tuition may discourage students from entering, either due to liquidity constraints or risk aversion. This would then also pre-empt potential gains from future immigration.

From the perspective of origin countries, sending students abroad involves both brain exchange and brain drain. That part of the students remaining abroad, or later emigrating, is a cost these countries encounter in order to upgrade the human capital of their youth. However, emigrants may also serve as a source of remittances and investments to their home country, thanks to significantly higher productivity of labor in destination countries, as well as establishing academic and business networks. Given this and population aging in destination countries with labor often being abundant in source countries, international student flows may well generate mutual gains.

References

- Alberto Alesina; Easterly, William; Devleeschauwer, Arnaud; Kurlat, Sergio and Romain Wacziarg, 2003, Fractionalization, *Journal of Economic Growth* 8: 155-194.
- Barro, Robert J. and Jong-Wha Lee, 2000, International Data on Educational Attainment: Updates and Implications, manuscript, Harvard University.
- Borjas, George J., 1994, The Economics of Immigration, *Journal of Economic Literature* 32: 1667-1717.
- Clark, Ximena; Hatton, Timothy J. and Jeffrey G. Williamson, 2002, Where do U.S. Immigrants Come From, and Why? NBER Working Paper 8998.
- Fertig, Michael, 2001, The Economic Impact of EU-enlargement: Assessing the Migration Potential, *Empirical Economics* 26: 707-720.
- Gibson, Campbell J. and Emily Lennon, 1999, Historical Census Statistics on the Foreign-born Population of the United States: 1850-1990, Technical Working Paper 29, U.S. Census Bureau, Population Division.
- Haan, Jakob de and Jan-Egbert Sturm, 2003, Economic Freedom and Economic Growth; A Reassessment of Evidence Based on Cross-Country Growth Regressions.

- Hatton, Timothy J., 2003, Emigration from the UK, 1870-1913 and 1950-98, University of Essex, mimeo.
- Hufbauer, Gary C. and Barbara Oegg, 2003, *The Impact of Economic Sanctions on US Trade: Andrew Rose's Gravity Model*, International Economics Policy Briefs PB03-4, Institute for International Economics.
- Institute of International Education, 50 Years of Open Doors, CD-ROM.
- Institute of International Education, Open Doors Data Archive, <http://opendoors.iienetwork.org/?p=28636>.
- Marshall, Monty G. and Keith Jagers, 2000, Polity IV Project: Political Regime Characteristics and Transitions, 1800-2000, <http://www.cidcm.umd.edu/inscr/polity/>.
- Mitchell, James and Nigel Pain, 2003, The Determinants of International Migration into the UK: A Panel Based Modelling Approach, National Institute of Economic and Social Research, London.
- OECD, 2001, International Migration Statistics, CD-ROM, Paris.
- OECD, 2003, Education Database, CD-ROM, Paris.
- Pedersen, Peder J., Mariola Pytlikova M. and Nina Smith, 2004, Selection or Network Effects? Migration Flows into 27 OECD Countries, 1990-2000, IZA DP 1104.
- Rousseeuw, Peter J., 1984, Least Median of Squares Regression, *Journal of the American Statistical Association* 79: 871-880.
- Tremblay, Karine, 2002, Student Mobility Between and Towards OECD Countries in 2001: A Comparative Analysis, in: OECD, *International Mobility of the Highly Skilled*, OECD, Paris: 39-67.
- U.S. Census Bureau, various years, Annual Current Population Survey
- U.S. Immigration and Naturalization Service, 2000, Characteristics of speciality occupation workers (H-1B), Washington, DC.
- William Easterly and Mirvat Sewadeh, 2002, Global Development Network Growth Database, World Bank, Washington, DC.
- World Bank, 2003, World Development Indicators, CD-ROM, Washington, DC.
- Zavodny, M., 1997, Welfare and the Locational Choices of New Immigrants, *Economic Review* – Federal Reserve Bank of Dallas, Second Quarter 1997.

Appendix A: Variable Definitions and Sources
Cross Section (OECD countries)

Variable	Source	Definition
Migration to the OECD	OECD (2001)	All legal immigrants admitted, including refugees. Data is by country of origin. Data are for 1995.
Foreign students	OECD (2003)	All foreign students enrolled in tertiary education. Data are for the year 2000.
GDP per capita ratio	World Bank (2003)	Gross domestic product divided by midyear population. Data are in constant U.S. dollars. The value for the origin country is divided by those for the U.S.
Schooling years ratio	Barro and Lee (2000)	Average schooling years in the total population aged 25+. The value for the origin country is divided by those for the U.S. Linearly interpolated when missing.
Landlocked, dummy	Easterly and Sewadeh (2002)	Dummy equals one if country has no direct access to the sea, zero otherwise.
Immigrant stock	OECD (2001)	Number of foreign born immigrants. Data are for 1995.

Appendix A (continued)
Panel (United States)

Variable	Source	Definition
Migration to the U.S.	Clark et al. (2002)	All legal immigrants admitted, including refugees. Includes immigrants who applied from abroad and those who are already in the United States but change to permanent status. Data is by country of origin.
Foreign students	Institute of International Education, various years.	A non-immigrant class of admission. Someone coming temporarily to the United States to pursue a full course of study in an approved program in either an academic (college, university, seminary, conservatory, academic high school, elementary school, other institution, or language training program) or a vocational or other recognized non-academic institution.
GDP per capita ratio	World Bank (2003)	Gross domestic product divided by midyear population. Data are in constant U.S. dollars. The value for the origin country is divided by those for the United States.
Schooling years ratio	Barro and Lee (2000)	Average schooling years in the total population aged 25+. The value for the origin country is divided by those for the United States. Linearly interpolated when missing.
Distance from the U.S.	Hufbauer et al. (2003)	Distance between capitals in miles. Updated with data received from Howard J. Wall.
Landlocked, dummy	Easterly and Sewadeh (2002)	Dummy equals one if country has no direct access to the sea, zero otherwise.
English speaking (percent of population)	Alesina et al. (2003)	Percent of population that is English speaking. Data refers to 2001.
Immigrant stock	U.S. Census Bureau, various years, and Gibson and Lennon (1999)	Number of foreign born immigrants. Linearly interpolated when missing.

Appendix A (continued)

Variable	Source	Definition
Population	World Bank (2003)	Total number of residents in thousands regardless of legal status or citizenship – except for refugees not permanently settled in the country of asylum.
Democracy	Marshall and Jaggers (2000)	0-10 (0 = low; 10 = high) democracy score. Measures the general openness of political institutions.
U.S. unemployment rate	World Bank (2003)	The share of the labor force that is without work but available for and seeking employment.
U.S. GDP growth	World Bank (2003)	Annual percentage growth rate of GDP at market prices based on constant local currency.

Appendix B:**Descriptive Statistics (Cross Section, Estimation Sample)**

Variable	Min	Max	Mean	Std. Dev. (overall)
log (immigrants admitted to U.S./ origin population)	-20.70	-13.63	-15.99	1.21
log (foreign students/ origin population)	-18.66	-6.95	-10.84	1.96
GDP per capita ratio (origin/destination)	0.01	2.01	0.56	0.43
Schooling years ratio (origin/destination)	0.21	1.40	0.83	0.27
Landlocked, dummy	0	1	0.02	0.16
Immigrant stock/ origin population	2.61e-06	0.03	0.003	0.01

Descriptive Statistics (Panel, Estimation Sample)

Variable	Min	Max	Mean	Std. Dev. (overall)
log (immigrants admitted to U.S./ origin population)	-14.34	-4.14	-8.59	1.76
log (foreign students/ origin population)	-17.48	-5.66	-9.01	1.31
GDP per capita ratio (origin/ U.S.)	0.007	2.27	0.39	0.45
Schooling years ratio (origin/ U.S.)	0.006	1.10	0.49	0.24
Distance from U.S.	988.32	9683.95	4859.76	2245.17
Landlocked, dummy	0	1	0.09	0.29
English speaking (percent of population)	0	97.29	8.67	25.27
Immigrant stock/ origin population	0.00008	0.34	0.019	0.03
Immigrant stock/ U.S. population	1.36e-06	0.028	0.0009	0.002
Democracy	0	10	5.91	4.27
U.S. unemployment rate	4.2	9.7	6.50	1.32
U.S. GDP growth	-2.07	7.28	3.20	2.20

Appendix C: Country List

Cross Section (Immigration to OECD)

Countries of Origin			Destination Countries
Algeria	Greece	Peru	Belgium
Australia	Hungary	Philippines	Denmark
Austria	Iceland	Poland	Germany
Bangladesh	India	Portugal	Japan
Belgium	Iraq	Romania	Netherlands
Brazil	Italy	Somalia	Norway
Bulgaria	Jamaica	Spain	Sweden
Canada	Japan	Sri Lanka	Switzerland
Chile	Lebanon	Sweden	United Kingdom
China	Malaysia	Thailand	
Croatia	Morocco	Tunisia	
Denmark	Netherlands	Turkey	
	New	United	
Ethiopia	Zealand	Kingdom	
Finland	Norway	United States	
France	Pakistan	Vietnam	
Germany			

Panel (Immigration to the United States)

Algeria	El Salvador	Japan	Sierra Leone
Argentina	Fiji	Jordan	Singapore
Australia	Finland	Kenya	South Africa
Austria	France	Kuwait	Spain
Bangladesh	Germany	Liberia	Sri Lanka
Barbados	Ghana	Malaysia	Sudan
Belgium	Greece	Malta	Sweden
Bolivia	Guatemala	Mexico	Switzerland
Brazil	Guyana	Nepal	Syria
Cameroon	Haiti	Netherlands	Trinidad and Tobago
Canada	Honduras	New Zealand	Tunisia
Chile	Hong Kong	Nicaragua	Turkey
China, People's Rep.	Hungary	Norway	United Arab Emirates
Colombia	Iceland	Pakistan	United Kingdom
Congo, Dem. Rep.	India	Panama	Zambia
Costa Rica	Indonesia	Paraguay	Zimbabwe
Cyprus	Iran	Peru	
Denmark	Ireland	Philippines	
Dominican Republic	Israel	Poland	
Ecuador	Italy	Portugal	
Egypt	Jamaica	Senegal	