# SKILLED IMMIGRATION AND INNOVATION: EVIDENCE FROM ENROLLMENT FLUCTUATIONS IN U.S. DOCTORAL PROGRAMS*

ERIC T. STUEN AHMED MUSHFIQ MOBARAK KEITH E. MASKUS

Abstract

We study the contribution of foreign doctoral students to innovation at 2300 American science and engineering (S&E) departments from 1973 to 1998. Macroeconomic and policy shocks in source countries that differentially affect enrollments across fields and universities isolate exogenous variation in the supply of students. Both U.S. and international students contribute significantly to the production of knowledge at scientific laboratories. A theoretical model of scholarships helps us infer the productivity effects of student quality. Visa restrictions limiting entry of high-quality students are found to be particularly costly for academic innovation. Foreign students increasing the diversity of departments appears to be one mechanism by which students contribute to the productivity of laboratories.

March 4, 2011

*Stuen is assistant professor at the University of Idaho. Maskus is professor at the department of economics, University of Colorado at Boulder. Mobarak is an assistant professor at the Yale School of Management. Contact: ahmed.mobarak@yale.edu. We thank the National Bureau of Economic Research (NBER) *Innovation Policy and the Economy* program and the National Science Foundation *SciSIP* program for financial support, the National Science Foundation *Division of Science Resources and Statistics* for providing data on Ph.D. recipients under a licensing agreement, and Rich Beaudoin, James Choy and Christine Rohde for excellent research assistance. We are grateful to Martin Boileau, Lee Branstetter, Benjamin Jones, Fiona Scott-Morton, Stephen Redding and seminar attendees at the NBER Innovation Policy and the Economy meetings, the NBER International Trade and Investment meetings, NBER Economics of Education Meetings, NSF SciSIP Conference, American Economic Association 2011 Annual Meetings, Duke University, Yale University, and London School of Economics for helpful comments.

## 1. Introduction

This paper explores whether and how foreign graduate students contribute to developing knowledge in science and engineering (S&E) at U.S. universities. We exploit fluctuations in the supply of foreign students stemming from macroeconomic and policy shocks in source countries to examine the effects of doctoral students in specific disciplines and universities on scientific publications and citations produced in those labs. Since the advent of tighter restrictions on the issuance of U.S. education visas after September 11, 2001, it has been increasingly argued in the media[1] and in prominent science journals[2] that the ability of American universities to undertake scientific research is dependent on technically trained international graduate students. In turn, restrictive visa policies could cause "…a crisis in research and scholarship" and harm the nation's innovative capacity.[3] This debate has now spread to Great Britain where scientists, business leaders and universities have expressed grave concern over the effects of newly proposed immigration caps and increases in foreign student fees on UK science and innovation.[4] We conduct careful analysis with micro data to shed light on these claims.

Among economists the U.S. immigration policy debate has centered on the effects of immigration on natives' wages (Card 1990, Altonji and Card 1991, Borjas 1994, 2003,

[1] A letter to this effect was published by a broad coalition of U.S. professors and administrators as "Academics Warn of Crisis over Visa Curbs", *Financial Times* May 16, 2004. See also "Visas and Science: Short-Sighted," *The Economist*, May 8, 2004.

[2] "Security Restrictions Lead Foreign Students to Snub US Universities," *Nature,* September 15, 2004. See also "The Knowledge Economy: Is the United States Losing Its Competitive Edge?" The Report of the Task Force on the Future of American Innovation, 16 Feb 2005.

[3] Partly because of tighter limits on student visas since 2001, the number of foreign graduate students in the United States fell by eight percent in 2002 and by a further ten percent in 2003, reversing a 15-year trend of rapid growth. Computer science and other S&E disciplines experienced the largest relative declines, as the U.S. Department of Homeland Security instituted the lengthy *Visa Mantis* security clearance program. A key concern is whether these trends presage a diminution in U.S. leadership in science and innovation.

[4] The Times UK, "8 Nobel laureates, including immigrants Geim & Novoselov, write to Times to condemn immigration cap," accessed November 12, 2010. Prominent newspaper editorials have also warned of a brain drain to other countries: "Government Cuts will trigger brain drain," The Telegraph, October 1, 2010.

Borjas, Freeman and Katz 1997, Ottaviano and Peri 2005). In the skilled-labor category, Borjas (2005) points out that doctoral student immigration has a significant adverse effect on wages of competing high-skilled U.S. workers. On the other hand, the United States has a global comparative advantage in science and innovation and in creating technology-driven new products and markets. The U.S. trade deficit is smallest in high-tech industries (Freeman 2005). Large imports of foreign doctoral students in science and engineering may be an important reason the United States has sustained its primary position as developer of new scientific knowledge, even with deficiencies in math and science training among American secondary school students.[5]

There are clear indications at a descriptive level that foreign students are important contributors to knowledge creation at U.S. universities. Foreign enrollments have increased in absolute and relative (to American enrollments) terms since the 1970s, and publishing and patenting have grown in lockstep. Foreign students are disproportionately more likely to earn graduate degrees in S&E, and now outnumber Americans in U.S. engineering departments (Council of Graduate Schools). Black and Stephan's (2008) survey of articles published in *Science* finds that 86.5% of papers have a student or post-doc author, and 60% of those authors are foreign-born. The authors conclude that international graduate students and post-doctoral candidates staff laboratories and play lead roles in university research. Regarding subsequent innovation, studies document that patent applications in the United States are correlated with foreign student enrollments at the aggregate national level (Chellaraj, et al 2008) and with post-doctoral enrollments at the university level (Gurmu, et al 2009). These provocative

[5] OECD (2006) reports that students aged 15 in the United States ranked 24th in mathematics and 19th in science among 29 countries. Freeman (2009) notes that "the U.S. has come to rely extensively on the

correlations between enrollments and patent productivity could be driven by omitted variables (e.g. if student applications surge when departmental quality improves), and deserve further scrutiny.

In this paper we analyze detailed information on individual students to assess their impact on research outcomes. We assemble a database of student enrollment counts by nationality for 2300 U.S. science and engineering departments for 1973-2004 by aggregating individual records on each doctoral student maintained by the *National Science Foundation*. We combine these records with publications in scientific journals from each of those departments, which are compiled from publication and citation searches on the Institute for Scientific Information's *Web of Science*.

To isolate the causal impacts of doctoral students, we devise an instrumental-variables strategy using the idea that macroeconomic shocks and policy changes in source countries lead to plausibly exogenous variation in the supply of foreign students. For example, macroeconomic crises in East Asia or decisions by Chinese authorities to permit their students to enter graduate programs abroad (or to reinstitute restrictions) tended to exogenously alter student supplies in the United States. Moreover, these shocks would have differentially larger impacts on research output in fields of study that are traditionally more popular among Asians and on universities that have traditionally recruited more students from that region. One advantage of this last insight is that even if some relevant events in the United States happen to coincide with those shocks (e.g. the 1980 Bayh-Dole Act and the roughly simultaneous lifting of study-abroad restrictions in China), the event is unlikely to affect publishing along the same patterns of universities and disciplines as the foreign shocks. Our specifications add fixed effects for all 2300

---

immigration of highly educated persons to maintain a lead position in science and technology."

departments, and university-specific and field-specific trends, so that empirical inference is based only on changes in publishing in an academic department following fluctuations in student enrollments in that department.

We find that foreign doctoral students significantly and positively influence publications and citations produced by U.S. academic departments. Each additional foreign student leads to 0.9 extra S&E journal articles per year.[6] The marginal effects of foreign and American students are statistically comparable, which is consistent with an optimizing department that equates value at the margin.

We further show that the type (quality) of entering foreign doctoral students matters greatly for research productivity. To overcome a lack of direct measures of student quality, we develop a theoretical model of admissions decisions in which universities trade off student quality and tuition income. This model shows that, comparing across different sets of source-country shocks that differentially affect application propensities of students who are able to pay versus students who rely on scholarships, we may distinguish empirically the productivity effects of higher-quality and lower-quality students. Shocks that differentially increase enrollment decisions of foreign students reliant on scholarships are much more productive for U.S. academic departments than shocks that send students who pay for their own graduate education. These results suggest that from the perspective of U.S. science education and innovation policy, visa restrictions for foreign students should not be applied uniformly or on the basis of financial means; they ought to account for student-quality differences.

---

[6] Accounting for publication-inflation in our data due to multiple-authored articles sometimes being double-counted, this implies a contribution of about 4.5 extra articles in the department over the course of a doctoral student's 5 or 6 year graduate career. The estimate size is quite reasonable given publication rates in science and engineering.

While an increase in foreign-student share changes the mix of nationalities, it also increases geographical diversity in the average U.S. S&E department (where 69 percent of doctoral students are American during the sample period). We find limited evidence that diversity matters, controlling for nationality effects. Only in an OLS model do we find that departments that enroll student cohorts from a wider variety of global regions experience larger gains in research output.

Our analysis thus documents a key benefit of high-skilled immigration in the United States, which relies on innovation for growth. The mainstream media is inundated with quotes from universities and employers on the economic dangers of visa restrictions.[7] It is valuable to estimate the causal returns of foreign-student presence so that immigration policy can account for the tradeoff between knowledge gain and the costs of tuition subsidies and congestion claimed by Borjas (2002, 2004). Our results add to the literature documenting other benefits of immigration (Cortes 2009, Mishra 2005).

The paper proceeds as follows. We outline modeling frameworks in section 2, including a tuition income model to motivate our empirical analysis of heterogenous student quality. We present the IV-based empirical approach in section 3, data sources in section 4, and discuss results in section 5. We make concluding remarks in section 6.

## 2. Modeling Framework

We estimate a knowledge production function. The extensive literature on the determinants of university patenting (Thursby and Thursby 2002, Thursby and Kemp 2002, Jaffe and Trachtenberg 2002, Mowery, et al 2004 and Azoulay, et al 2005) is thus closely related. Regarding the research productivity of foreign students, papers relating patent applications and grants to students and immigrants at the national level (Chellaraj, et al 2008), state level (Hunt and Gauthier-Loiselle 2008), and university level (Gurmu, et al 2009) are also relevant.

*2a. Insights from Neoclassical Production Function Analysis*

The simplest theory by which to interpret our statistical results is a neoclassical production function for an academic laboratory with domestic and foreign graduate students as inputs.[8] A rational resource allocator wishing to maximize a single output with a budget constraint would choose inputs such that the marginal product per dollar spent on each were the same (Pritchett and Filmer 1997). In our setting, research departments would admit domestic and international graduate students to equate the contribution of each type, scaled by some measure of department-level cost of educating them. Such costs could include stipends and the opportunity costs of faculty time. Domestic graduate students likely are cheaper because of lower tuition costs for state residents at public universities and higher training costs for foreign students due to language difficulties. Further, scholarship students who are more costly to educate must provide greater marginal value in equilibrium.

University departments generate multiple outputs (research and undergraduate teaching), so they would admit graduate students until the difference in per-dollar marginal products in producing publications equals the difference in per-dollar marginal

[7] See *New York Times* March 3, 2009 on researchers and academic conferences moving abroad, *Wall Street Journal* March 11, 2009 arguing that productive foreign-born workers create more jobs, and *New York Times* April 12, 2009 on immigration laws making it more difficult for Google to recruit workers.
[8] Hanushek (1979) is the seminal reference. See also Johnson (1978) and de Groot, et al (1991) for examples. These models assume either one output (e.g., test scores) or multiple outputs (e.g., graduate diplomas and research) produced using a variety of inputs, such as faculty size and research funding.

products in teaching. If, for example, domestic students offer greater productivity as teaching assistants than foreign students, we would observe a greater marginal contribution to producing publications from international doctoral students in equilibrium. In this model all marginal products should be positive. And enrollment shocks should not have a substantial impact on the production of publications because departments would be able to substitute across inputs.

*2b.Outline of a Model of Knowledge Creation with Heterogenous Inputs*

This approach misses some important details of the reality of what happens in academic research, since it is inaccurate to think of universities as operating with a fixed budget constraint and highly divisible homogeneous inputs. In this section we outline a more suitable modeling framework to motivate our identification strategy based on student-supply shocks in the basic publication and citation regressions. We then solve a simpler version of the model to motivate how different types of macroeconomic shocks can help us separately identify the effects of scholarship (i.e., higher-quality) students versus paying students, even when we do not have direct measures of student quality.

Suppose that a department $d$ produces knowledge using as inputs "professors" $P$ (i.e., an index of faculty, grant funding and the like) and quality-weighted students from the United States, $u$ and two foreign regions, $g$ and $h$:

$$K_{dt} = K_d(P_{dt}, \sum_{u=1}^{U} q_{dt}^{u} + \sum_{g=1}^{G} q_{dt}^{g} + \sum_{h=1}^{H} q_{dt}^{h}) \qquad (1)$$

The creation of knowledge rises in both faculty quality and the quality-weighted number of graduate students. Departments limit enrollments due to resource constraints and for quality control. For simplicity, assume that a department's capacity to enroll graduate students at any time is strictly constrained:

$$U_{dt} + G_{dt} + H_{dt} = \overline{S_{dt}} \tag{2}$$

Foreign students are costly to admit but this cost declines as departments gain experience with students from different regions:

$$C_{dt} = c^g_{dt}(G_{d,t-1}) \cdot G_t + c^h_{dt}(H_{d,t-1}) \cdot H_t \text{, where } c^g, c^h \geq 0 \text{ and } c^{g'}, c^{h'} \leq 0 \tag{3}$$

We normalize the cost of training an American to zero. The cost of training foreign students diminish as departments gain more experience with students from that region.

Economic conditions can affect both the supply and quality of students available, because their incentives and ability to apply depend on labor market conditions (e.g., wage available in alternative employment), macroeconomic factors (e.g., exchange rate changes that affect ability to incur cost of travel and emigration limits or visa restrictions). Thus, a student from region *G* applies to department *d* at time *t* if $P_{dt} + \varepsilon^g_{dt} \geq w^G_t$, or if the benefits she gets from studying at department *d* (determined by department quality *P* plus her idiosyncratic preference to study there) exceed the "outside option" *w* (e.g., the wage a college graduate would get in her region).

Departments admit students with the highest value added (*K* – *C*), taking into account student quality and region-specific costs. Experience with regions will matter in these admissions decisions over time, as the cost of admitting students from a region decreases as universities invest in learning about students from that region (or conversely, the region's students learn about the university). Since students prefer to attend higher-quality universities and universities prefer better students, the matching process assigning applicants to departments will be assortative.

This framework highlights the endogeneity problem in estimating the effects of students: A positive shock to department quality *P* would increase knowledge production independent of any changes to student enrollment, but it would also attract both more Americans and more foreigners to apply. In a regression of publications on the foreign-student share, the direction of the resulting bias would depend on which type of students ultimately enrolls as a result of this shock. If departments show a preference for Americans over foreigners in the admissions decision (e.g., because $c^g, c^h \geq 0$), enrollments will shift in favor of the former and the coefficient on foreign student enrollment in the publications regression would be biased in the negative direction.

Shocks to the outside option (i.e., $w^G$, $w^H$) may be used as instruments to identify exogenous fluctuations in the numbers of foreigners enrolling in graduate programs in the United States. Moreover, shocks in particular regions may have differential effects on enrollments across different departments by virtue of the fact that a department's history with students from a region can matter in the admissions decision ($c^{g'}(G_{d,t-1}) \leq 0$). Interaction terms between region-specific shocks and department-region enrollment histories may yield powerful instruments that identify shock-induced, department-specific fluctuations. We will take advantage of these insights in developing an estimation strategy in section 3.

*2c. A Simple Model with Student Quality, Tuition and Scholarships*

A key visa policy question of interest is whether immigration restrictions should account for the quality of incoming students. Given the costs associated with admitting foreign students (Borjas 2002, 2004, 2005), the optimal policy may be to allow only students who are likely to contribute more to innovation and U.S. competitiveness. U.S. visa policy for foreign students has traditionally paid more attention to financial resources and assets held abroad to infer students' incentives to remain in the United States past graduation, which is curiously viewed as a negative outcome.[9]

We would like to estimate the differential contributions to innovation from high-quality and low-quality foreign students to inform this debate. Unfortunately, there are no good direct empirical measures of a student's "quality". Further, even if such measures existed, we would have to account for the endogenous placement of higher-quality students. Given these constraints, we now solve a simplified version of a Ph.D.-admissions model to highlight how student-quality effects can still be estimated under the reasonable premise that: (a) higher-quality students are more likely to receive scholarships, and (b) certain source-country macroeconomic conditions cause differentially larger fluctuations in non-scholarship students.

Suppose universities care about student quality *q* and tuition income *m* (per-student). The population of potential students is divided into $N^R$ rich applicants who can afford to pay tuition and $N^P$ poor applicants who are credit-constrained and can only attend with a scholarship. The distribution of student quality in both groups is *f(q)*. With a convex cost function $c(\cdot)$ that reflects the congestion costs of admitting more students, and marginal benefits of quality and tuition to the university denoted as *A* and *B* respectively, the university's objective function becomes:

$$\begin{aligned} U &= A(N^R + N^P)\int_{q_s}^{\infty} qf(x)dx + AN^R\int_{q_n}^{q_s} qf(x)dx + BN^R\int_{q_n}^{q_s} mf(x)dx \\ &- c\left[N^P\int_{q_s}^{\infty} f(x)dx + N^R\int_{q_n}^{\infty} f(x)dx\right] \end{aligned} \qquad (4)$$

[9] Applicants for F-1 student visas must demonstrate that they have enough readily available funds to meet all expenses during a course of study and agree to leave the United States after completing their education (http://www.usastudyguide.com/immigration.htm).

The university maximizes over $q_s$ and $q_n$, the lower quality bounds for admitted scholarship and non-scholarship students, respectively. The university cannot observe whether students are rich or poor, so it gives scholarships to all students with quality exceeding $q_s$, although some of those students are rich. For algebraic tractability, we will assume that $c(x) = \frac{C}{2}x^2$ (with a constant $C$) and that $f(q) \sim U[0, \overline{q}]$.[10] The first order conditions simplify to:

$$-AN^P q_s + BN^R m + N^P C\left[N^P \frac{\overline{q} - q_s}{\overline{q}} + N^R \frac{\overline{q} - q_n}{\overline{q}}\right] = 0 \tag{5a}$$

$$-Aq_n - Bm + C\left[N^P \frac{\overline{q} - q_s}{\overline{q}} + N^R \frac{\overline{q} - q_n}{\overline{q}}\right] = 0 \tag{5b}$$

We can obtain closed-form solutions to $q_s$ and $q_n$ by solving (5a) and (5b). This shows that $q_s > q_n$ , as expected. Taking derivatives with respect to the numbers of rich and poor students:

$$\begin{aligned}
\frac{\partial q_s}{\partial N^P} &= \frac{A^2\overline{q}^2(C(N^P)^2 - BmN^R) - 2ABCmN^R\overline{q}(N^P + N^R) - BC^2 mN^R(N^P + N^R)^2}{A(N^P)^2(A\overline{q} + C(N^P + N^R))^2} \\
\frac{\partial q_s}{\partial N^R} &= \frac{A^2CN^P\overline{q}^2 + Bm(A\overline{q} + C(N^P + N^R))^2}{AN^P(A\overline{q} + C(N^P + N^R))^2} \\
\frac{\partial q_n}{\partial N^P} &= \frac{\partial q_n}{\partial N^R} = \frac{AC\overline{q}^2}{(A\overline{q} + C(N^P + N^R))^2}
\end{aligned} \tag{6}$$

Since total graduate student quality is $Q = N^P \int_{q_s}^{\infty} qf(q)dq + N^R \int_{q_n}^{\infty} qf(q)dq$

$= \frac{1}{2}N^P(\overline{q} - \frac{q_s^2}{\overline{q}}) + \frac{1}{2}N^R(\overline{q} - \frac{q_n^2}{\overline{q}})$, the effects of a change in $N^R$ or $N^P$ on $Q$ are:

[10] A normal distribution for ability may be more realistic, but the uniform distribution with a large upper-bound $\overline{q}$ is a reasonable approximation to the far right tail of the normal distribution, where applicants to Ph.D. programs are likely to reside.

$$\frac{\partial Q}{\partial N^P} = \frac{1}{2}\overline{q} - \frac{q_s^2}{2\overline{q}} - \frac{N^P}{\overline{q}} q_s \frac{\partial q_s}{\partial N^P} - \frac{N^R}{\overline{q}} q_n \frac{\partial q_n}{\partial N^P}$$
$$\frac{\partial Q}{\partial N^R} = \frac{1}{2}\overline{q} - \frac{q_n^2}{2\overline{q}} - \frac{N^P}{\overline{q}} q_s \frac{\partial q_s}{\partial N^R} - \frac{N^R}{\overline{q}} q_n \frac{\partial q_n}{\partial N^R} \qquad (7)$$

Plugging the expressions from (6) into (7), we get:

$$\frac{\partial Q}{\partial N^P} - \frac{\partial Q}{\partial N^R} = \frac{B^2 m^2 (N^P + N^R)^2}{2A^2 (N^P)^2 \overline{q}} \qquad (8)$$

This expression is always positive, so a positive shock to the number of poor applicants increases graduate student quality by more than a positive shock to the number of rich applicants. This is the main insight we will take advantage of in our empirical analysis of the effects of student quality. Specifically, even in the absence of direct measures of quality, we can infer quality effects through this scholarship mechanism if we can identify separate sets of foreign-country shocks that differentially affect the application propensities of rich versus poor students.

## 3. Empirical Methodology

### *3a. Basic Specifications*

We run the following basic specification to examine the effects of foreign and domestic Ph.D. student enrollments on knowledge produced in specific fields of inquiry within U.S. universities over the period 1973-1998:

$$K_{f,u,t} = \alpha_{f,u} + \delta_t + \gamma_u (D_u{*}Trend) + \rho_f (D_f{*}Trend)) + \beta_1 * American\_Students_{f,u,t-l} + \beta_2 * Foreign\_Students_{f,u,t-l} + \varepsilon_{f,u,t} \qquad (9)$$

The variable $K$ refers to either publications or citations. The dataset has four identifiers – the students' region of origin (e.g., South Asia), the university $u$ at which students are enrolled, the field of inquiry $f$ (e.g., industrial engineering), and year $t$. We will refer to the university-field pair as an academic "department". Equation (9) explains variation in scientific publications or citations in a given department and year as a function of American and foreign enrollments in the previous year, while controlling for fixed effects for every department, linear trends specific to each university and each field, and year dummies. Only some specifications add terms of the form $\beta * R\&D\_Exp_{f,\, u,\, MA(t\text{-}l,t\text{-}5)}$ to control for departmental resources (equipment, capital and R&D expenditures, including faculty salaries) because these are potentially endogenous variables that we do not have a good instrument for. The fixed effects control for time-invariant differences in characteristics across departments. The field-specific and university-specific trends capture any linear changes in the norms regarding publishing at a particular university or within a field of inquiry.[11]

After controlling for department and year fixed effects and field and university trends, the remaining objects of concern are unobservable characteristics of academic departments that vary non-linearly over time and affect both the publications produced by those departments and student enrollments. For example, if the quality of a department improves, say through better faculty or funding in a way not fully captured by the R&D expenditure controls, it may attract greater numbers of foreign students and also have an independent effect on the department's output. Conversely, if an improvement in the quality of a department (and therefore a rise in students' earning potential) attracts high-quality American students away from business, law and other professional degrees and

[11]The average number of publications per department rose from 25 in the 1970s to 54 in the 1980s, while citations rose from 832 to 1,654. Thus, it seems reasonable to think of the dependent variables as continuous, and we run linear regressions. We have also run negative binomial fixed-effects count-data models of the following form $K_{f,u,t} = e^{X_{f,u,t} \cdot \eta} \cdot \alpha_{fu} + \varepsilon_{f,u,t}$ (where $X_{f,u,t}$ encompasses all variables in (9)), and verified that the results are qualitatively similar.

into S&E fields, we may observe drops in foreign-student enrollments when a department's quality improves.[12] This is likely to bias the $\beta_2$ coefficient downward.

*3b. Instrumental Variables Approach*

Our solution to this problem is to use an instrumental variable (IV) estimator that takes advantage of plausibly exogenous fluctuations in the supply of foreign students. We instrument for U.S. and foreign enrollments using economic and policy shocks in the students' countries (or regions) of origin. We choose shocks that influence students' decisions about whether to enter (or travel to the United States for) graduate studies, but that are plausibly uncorrelated with the publications produced at specific U.S. academic departments. For example, eliminating or reinstituting study abroad restrictions altered Chinese students' ability to enter graduate programs in the United States. These policy changes can cause fluctuations in Chinese enrollment at U.S. universities but may not affect publishing in specific academic departments through other channels. To illustrate, Figure 1 plots enrollments of doctoral students from India at U.S. universities against Indian GDP growth. The co-movement of enrollment counts and the instrument displayed provides some preliminary indication of its statistical power.

We use fluctuations in source-country policies (e.g., lifting of restrictions on Russian and Chinese students regarding study abroad) and in economic conditions (e.g., income growth) to instrument foreign student enrollments. These instruments vary by country and year, while our endogenous variables of interest (foreign students) have

[12] Under a preference for Americans in admission due to language skills, wider ranges of financial-aid options available for natives or greater productivity in teaching, foreign students may get crowded out in a department of limited size once more high-quality American students start applying. If the marketability or popularity of a particular field of study among students at a given point in time varies non-linearly, that would be another omitted variable that may bias the impact of foreign students in either direction, depending on how U.S. students respond to such changes.

richer dimensions of variation, at the level of university, field of study, years of study, and origin. In order to exploit variation across all four dimensions in the data, we use the idea that the vulnerability to a student-supply shock from a particular country will differ by field and university. For example, if Purdue University has traditionally recruited a larger share of Indian students into its graduate programs (and therefore has invested in developing an ability to identify good Indian students), a shock to supply from India is expected to have a differentially larger impact on research at that institution. Similarly, if Indians are more likely to study chemical engineering, then this shock would affect chemical engineering departments more (and perhaps that field at Purdue the most).

Our disaggregated micro-data approach to answering these research questions has the advantage that, in this example, the Indian student shock would manifest itself in disproportionately larger impacts on publishing at Purdue (an institution-specific effect) and at relatively strong chemical engineering departments (a discipline-specific effect). This allows us better to distinguish the effects of student enrollments from coincident changes in economic or policy conditions in the United States that may alter publishing behavior. For example, the general decline in U.S. high-technology industries in the late 1990s may have affected university research output, and it also happened to coincide roughly with the East Asian financial crisis – a student-supply shock that we exploit with our instruments. However, given our IV strategy's reliance on the disproportionate effects of the Asian shock to particular fields and universities, this coincidence would only be a concern if the decline in the high-technology industries just happened to have a greater effect on publishing in the departments that have traditionally relied on East Asian students more. Observing events in the United States that had such specific patterns of influence on academic departments is considerably more unlikely than just finding events that happened to coincide with a foreign-country policy change or economic shock, which increases our confidence in this estimation strategy.

Figures 2-4 demonstrate the empirical relevance of these ideas. Figure 2 shows that there was a tremendous increase in Chinese doctoral students in the United States after the partial (1981) and total (1984) lifting of restrictions on study abroad. Moreover, these enrollments rose with the subsequent growth in Chinese GDP. Figures 3 and 4 further indicate that the University of Texas benefited differentially more from this surge in Chinese enrollments than did the University of California at San Diego and that electrical engineering departments benefited more than biochemistry departments.

We implement these ideas in the statistical analysis using triple interaction terms in our list of instruments:

*[Shock in region r in year t] * [fraction of university u foreign students who are from region r at some initial date* $t_0$*] * [fraction of students in field f from region r at* $t_0$*]*

The second term in brackets measures the university's historical dependence on students from that region while the third term measures the regional dependence of that field of study. We employ aggregated regions rather than specific countries because to instrument at the country level would incorporate so many variables that the problem of having many weak instruments in an over-identified system would be severe. The initial date $t_0$ is defined to be the start of each decade (1970, 1980, and 1990). The use of start-of-decade dependence on foreign students may lead to a concern that those interaction terms subsequently affect departmental faculty quality (since Ph.D. students in 1980 become faculty in the late 1980s). In some of our regressions we will directly control for a measure of faculty resources.

We instrument American enrollments with regional unemployment rates in a few US regions, as these movements affect the opportunity cost of enrollment. As with the international shocks, these are interacted with the lagged fraction of students in each university and field that are from the corresponding region of the US. The unemployment rate (and more generally U.S. economic conditions) may affect university scientific output through other channels, since university funding is tied to state revenues. In a robustness table, we omit all U.S. instruments, and show that our main results are qualitatively and quantitatively similar when only foreign macro shocks are used to instrument both domestic and foreign enrollments.

## 4. Data

### *4a. Publications and Citations*

We create counts of all science and engineering publications associated with the 100 U.S. universities that granted the largest number of foreign doctorates for the period 1973-2001, with data from the Thomson/ISI *Web of Science* database of publications and citations. Using a procedure described more fully in the data appendix, we sort each university's publication records into 23 S&E fields. Although laboratories and departments are the actual S&E administrative units at U.S. universities, we define "departments" as fields of science and engineering within a university. We extracted 3.2 million individual publication records by writing *Perl* script on the internet-based *Web of Science* database. Using information on the authors' department affiliation(s), the publications' subject categories and the year of publication, each of these records was assigned to one or more of 66,700 (100 x 23 x 29) university-field-year cells.[13] Our final database is a count of publications and total citations in each university-field-year cell. Summary statistics are provided in Table 1.

*4b. Enrollment Counts*

We create Ph.D. enrollment counts for each data cell (incorporating a university, field, year and country of origin) by aggregating the National Science Foundation *Survey of Earned Doctorates* (SED) micro-database, which contains a record for each individual who received a Ph.D. in the United States between 1959 and the present. Doctoral recipients fill out this survey when they receive the Ph.D. degree, so the yearly enrollment counts we create are based on the graduation date and the date of entry into the doctoral program reported by the students, and reflect only those students who have finished the degree. We infer enrollment counts for the period 1960-1997 only, since there are likely to be many students who entered doctoral programs in 1998 or thereafter who still had not received their degree by 2004, and therefore would not appear in the SED database.

We assign each student to one of 23 fields of study based on the reported three-digit dissertation specialty. The student's assignment by country of origin is based on the reported country of citizenship. Further details are in the data appendix. We create university-field-year-country enrollment counts for foreign students from the 50 largest countries (those that have supplied at least 930 doctoral students to the United States since 1960) studying in the 100 largest universities (those with at least 2100 doctoral students since 1960), in 23 S&E fields (as defined by Goldberger et al, 1995, also used by

[13] Publications with multiple authors were assigned once to each cell with which an author was affiliated.

Lach and Schankerman, 2008) during the period 1960-1997. There are approximately 700,000 doctoral students in the sample we analyze. Our instrumental variables are defined by aggregated regions of origin. We define six regions on the basis of economic, geographic and cultural similarities between countries.

Total doctoral enrollment in the average university-field-year was 47 students, 30 of whom were American. The East Asia/Pacific region, including China, was the next largest supplier of students at 8.4, followed by South Asia with 2.2. Enrollments for U.S. and foreign students are summarized in Table 1. The sample period for all regressions is 1973 to 1998, with enrollments lagged one year in order to reflect the lag from research to publication.

*4c. Data on Instrumental Variables*

We describe next the instruments we use for the first-stage prediction of enrollments and the share of foreign students.

*(1) GDP per Capita in Source Countries*: In relatively poor countries, this variable captures long-term changes in foreigners' ability to pay for a U.S. education (Figure 1). GDP growth can have the opposite effect in rich countries as it increases employment opportunities in local markets (Sakellaris and Spilimbergo, 2000).

*(2) Policy Changes*: We create a variable measuring the proportion of the world's population with freedom to study abroad as $\sum_{i=1}^{n}\sum_{t=1}^{T} D_{it} Pop_{it} / \sum_{i=1}^{n}\sum_{t=1}^{T} Pop_{it}$ , where $D_{it}$ = 1 if country *i* permitted study in the U.S. in year *t*, and 0 otherwise. An example of a major change in this variable is the lifting of the ban on study abroad by Chinese S&E students between 1978 and 1984 following the death of Mao Zedong (Orleans 1988). Other countries for which this policy indicator is relevant within our sample period include

Russia, Romania, Cuba, Poland, Hungary, Bulgaria, Czechoslovakia and (East) Germany. The data appendix provides further details.

*(3) International Students at non-U.S. Hosts*: Using the UNESCO *Statistical Yearbooks* (1963 to present), we create counts of the number of tertiary (university plus post-graduate) foreign students from each source country studying abroad at other (non-U.S.) host countries, such as the United Kingdom, Australia, Singapore, and Canada. The idea is that fluctuations in the number of South Asian students in the United Kingdom and Australia are related to changes in financial conditions and policy changes in South Asia and in those host countries, but uncorrelated with changes in conditions in the United States. To the extent that this instrument explains variation in South Asian students in the United States, the correlation is driven by the commonality between the two variables, which are the economic and policy conditions in South Asia.

## 5. Results

### *5a. First-Stage Results*

Table 2 shows the first stage of the two-stage limited-information maximum likelihood (LIML) regressions. All regressions control for a comprehensive set of fixed effects for "departments" (university -field pairs) and years, along with university-specific and field-specific time trends, but these coefficients are not shown in the tables.

Columns 1 and 2 show the first-stage with the full set of instruments, which correspond to the second-stage estimates reported in Table 3. Columns 3 and 4 exclude GDP shocks as instruments, and the subset of instruments that are included (policy shocks and students at non-U.S. hosts) identifies fluctuations in both scholarship and paying students. This first stage corresponds to the second-stage structural estimates reported in columns 1 and 3 of table 5 (which track the publication effects of pay-neutral shocks). Columns 5 and 6 in table 2 show the first stage where (conversely) only GDP movements in source countries to identify foreign student fluctuations. Since GDP movements primarily affect "ability to pay" for education for students from developing countries (Sakellaris and Spilimbergo 2000), it should exert a greater effect on non-scholarship enrollments. The first-stage regressions also include the unemployment rate in two U.S. regions to instrument fluctuations in American enrollments.

The first-stage regressions in columns 1 and 2 of table 2 show that, as expected, the "foreign" instruments have stronger positive effects on foreign enrollments, while the U.S. instruments have a positive effect on American enrollments. Comparing across columns 1 and 2, source-country instruments generally have the opposite effect on foreign enrollments and US enrollments, which suggests that there are some displacement effects.

Removal of study abroad restrictions in sources countries (measured as one-year lag of the percentage of the world population free to study in the U.S.) significantly increases foreign student presence in U.S. science and engineering departments. At the mean values of the interaction terms (start-of-decade university and field fractions), a one percentage point increase in the population free to study abroad increases international enrollments per department by 0.028. This is a small marginal effect, but the cumulated increase in freedom over the course of the sample period from 68.6% to 98.6% is responsible for almost one extra foreign student per department-year.

Foreign students studying abroad at non-U.S. destinations is strongly positively correlated with foreign enrollments in the U.S., which suggests that the instrument picks up variation in source country conditions. In poorer regions, positive GDP movements are associated with more students in the U.S. since "ability to pay" is the relevant margin for the study abroad decision. In more developed regions (East Asia and Europe) the effect is negative, supporting the theory that the opportunity costs of home-country employment is a more important determinant of study abroad from those countries.[14]

The unemployment rate in the New England census region has a positive effect on enrollment of Americans in S&E Ph.D. programs, but the effect is negative for two-year lagged unemployment in the E. S. Central region (states MS, AL, TN, KY). We add average GSP in these regions to the list of instruments in columns 3-6 (because removing large subsets of the source-country shocks leads to a weak instrument problem), and these are positively correlated with American enrollment.

The F-statistics for joint significance of all instruments in explaining US and foreign enrollments are 10.94 and 11.52, respectively, showing that their power, while low, is adequate to produce IV estimates with little bias. We use a two-stage LIML estimator to be conservative against a weak instrument problem (Imbens and Wooldridge 2007). Although our two-stage system is over-identified, we do not report over-identification tests because the theoretical model in section 2c implies that the different source country macro/policy shocks *should* have different effects on the productivity of foreign students (since the different shocks alter the quality distribution of students in different ways). If the theory is correct, then we should fail over-identification tests.

---

[14] In a more diverse, aggregated region (e.g. the W. Hemisphere region, which includes Canada, Australia, New Zealand and all of Latin-America and the Caribbean), the effects are ambiguous and not significantly

5b. *Summary Effects of Foreign and Domestic Doctoral Students*

Table 3 reports the OLS and the two-stage instrumental variables estimates of the knowledge production function (equation 9). The dependent variables are the number of scientific publications and the number of citations associated with each S&E "department" (university-field pair) in each year.[15] The endogenous variables of interest are counts of the numbers of domestic and foreign students in the department. All regressions control for university-field fixed effects, year effects and linear trends for each university and field, eliminating any possible effects of department size or secular changes in the sizes of universities and fields of science and engineering. Some models additionally control for a measure of department faculty and funding (R&D) resources.

In OLS regressions with the fixed effects and trends (columns 1 and 2), positive shocks to a department's student enrollment is associated with a larger number of publications the following year. Each additional student leads to about 0.13-0.15 publications per year, and the productivities are not significantly different across Americans and foreigners. These estimates suggest that over the course of an average-length six year Ph.D. program, the expected contribution of any Ph.D. student is 0.81 to 0.92 publications. This number reflects both authorship by the students themselves, and their indirect contribution to the productivity of the laboratory. It is also important to note that our computerized *Web of Science* search procedure counts each paper once for

different from zero..

[15] Consistent with the approach in Jaffe and Trajtenberg (2002) we actually use (1 plus counts) as the variable in order to distinguish between observations with no publications (a value of zero) and those with one or more publications that are not cited (a value of one).

each department which hosted a coauthor. As a result, we estimate that the nominal publications numbers we report are inflated by about 20%.[16]

Columns 3 and 4 report the two-stage LIML estimates of the knowledge production function using the full set of instruments. The marginal effects are much larger under IV than the OLS estimates. Here the point estimate of the effect of foreign students is greater than that of domestic students, but again the difference is not statistically significant (p = 0.544). Accounting for the co-authorship inflation, the estimates suggest that the marginal US student adds 0.59 publications per year while the marginal international student adds 0.74. By these estimates, the contribution of each student to her department's productivity over the average six year Ph.D. tenure is 3.5 to 4.4 publications. Doctoral students in science and engineering laboratories often graduate with 4 or more publications themselves,[17] and these estimates therefore appear to be of reasonable magnitude.

Comparing across OLS and IV specifications, OLS appears to under-estimate the effects of both domestic and foreign students. This suggests that departments received more (and higher quality) student applications in years when the other inputs to department quality (faculty, research funding, financial resources) were low. This would be consistent with surges in student applications in recession years, when departmental budgets are tight. This interpretation also suggests that the instrument for U.S. students (regional unemployment rates) may be correlated with departmental productivity through

[16] We arrive at this inflation estimate by comparing our publications data with an overlapping dataset compiled by the NBER, the Adams and Clemmons (2008) database of scientific papers.
[17] Publication rates among graduate students in science and engineering are much higher than in economics and other social science disciplines partly due differences in co-authorship norms. We informally surveyed the Dean of the Engineering school and department chairs in Biology, Physics, Chemistry, Physiology and Math at the University of Colorado at Boulder – a representative university in our sample – and learnt that

another channel (funding at state universities may suffer during recessions). We therefore report robustness results below omitting the U.S. instruments.

Columns 5-8 of Table 3 present results using the second knowledge measure, citations. In OLS regressions, we find almost no effect of foreign students on citation production. As with publications, the IV regressions show much higher productivities, with a marginal contribution from Americans of 73-78 citations per year, and 46-48 citations from international students. Here too we cannot reject the hypothesis of parity between the US and foreign student coefficients ($p = 0.218$).

*5c. Inferring Student-Quality Effects*

We would like to examine heterogeneous productivity effects by the quality of incoming foreign doctoral students, because the optimal visa policy may be conditional on the quality of skilled immigrants. In the absence of a direct measure of student quality, we use the insight developed in the model in section 2c. Specifically, if universities trade off tuition income against research productivity, shocks to the supply of poorer students (who are differentially more reliant on scholarships) will have larger effects on knowledge creation since those shocks will pick up higher-quality students on the margin.

We implement this idea in Table 4 by defining two sets of instruments. The first includes those instrumental variables that should have neutral effects across students of varying income levels, because they should uniformly affect the ability of all students to go abroad. These are source-country policy restrictions and changes in the number of tertiary students studying outside the United States. The second set includes those

---

in Engineering, doctoral students typically graduate with 4-5 publications, in natural sciences 3-4 and in mathematics, less than 3. These numbers were likely smaller in earlier decades.

instrumental variables that affect students' ability to bear the *monetary* cost of studying in the United States. GDP per capita movements in the source countries are more likely to affect study-abroad decisions for the set of students paying for a U.S. education.

The idea from the theory that we are trying to capture is that "pay neutral" shocks should have little effect on the quality distribution of students, while positive "ability to pay" shocks could lower the average quality of students. The set of compliers in the IV regressions with instruments that affect "ability to pay" are more likely to be non-scholarship students. The latter set of shocks thus disproportionately shift enrollments toward lower-quality applicants, and comparing across the two types of IV regressions allows us to infer indirectly the effects of student quality on knowledge production.

Table 4 shows the results with these subsets of instruments. Since our set of US instruments is not amenable to such division, we include all US instruments in both sets and focus on the estimates of the coefficient on foreign enrollment. And to deal with a weak instrument problem that arises once the source country instruments are split, we add two additional US instruments: mean gross state product for the E. S. Central and Mountain census regions. The coefficient on foreign students identified by the pay-affecting shocks is half as large as that identified by the pay-neutral shocks: 0.8 publications per year versus 1.6. This suggests that a positive shock to the ability to pay reduces the measured contribution of students to publishing by reducing student quality, at the margin. In the model this comes about because the pay-affecting shocks leads to greater enrollment of paying non-scholarship students on whom departments set a lower admissions standard. With regard to their effects on citations, we find that the effect of the marginal foreign student identified with "pay-neutral" shocks is much greater, 49.5

citations versus 31.5.  A test based on Hausman (1978) reveals that these two coefficients are significantly different from each other (t-stat of 5.62).  This suggests the research impact of visa restrictions on foreigners will depend on how the restriction affects the quality margin, which in turn depends on how the immigration policy is implemented (i.e., whether the agency issuing visas screens for research ability or for 'ability to pay').

*5d. Diversity Effects*

Increasing the enrollment of foreign students also has the effect of diversifying production teams, since they have been the minority in U.S. doctoral programs, which implies that adding foreigners adds to diversity within the laboratory in terms of background and training.  Team members bringing complementary skills and training may have multiplicative effects on innovation.  The increasing dominance of laboratory teams and co-authoring in the production of scientific knowledge is well documented (Jones, Uzzi and Wuchty 2008), and diversity may result in positive spillovers from the exchange and mixing of ideas, training and methods if students from different regions bring complementary and heterogenous skills (Sparber 2009, Ottaviano and Peri 2006, Alesina and Ferrara 2005, Berliant and Fujita 2004, Fujita and Weber 2005, and Niebuhr 2006).  Or it could facilitate the trend towards cross-country research collaborations (Adams, et al 2005).

Table 5 incorporates a variable capturing diversity of international doctoral student enrollments, a Herfindahl index of regional shares.  We do not have strong instruments to separately identify both variation in the number of international students and the regional composition of those students.  So we report merely suggestive OLS regressions in this table.  We find that after controlling for US and foreign enrollments, diversity has a positive and significant effect on both publications and citations. The elasticity of publications and citations with respect to diversity is about 0.04. This effect is small but quantitatively significant. For example, a department with ten foreign students from five regions is expected to generate 0.76 more publications and 28.65 more citations per year than a department with ten foreign students from two regions. This provides suggestive evidence in favor of models linking diversity to productivity: that mixing of complementary ideas from different regions is the mechanism by which combinations of U.S. and foreign students affect high-quality scientific achievement of academic departments. However, our lack of suitable instruments for diversity and the potential for diversity itself to be correlated with unobservable productive factors means that these results are not conclusive evidence of a causal effect of diversity.

Finally, motivated by a concern that U.S. variables (regional unemployment rates and state GSPs) may be related scientific output from U.S. universities through other channels (e.g. availability of university funding), table 6 reports robustness results where only the subset of foreign macro and policy shocks are included as instruments. We lose some first stage power in the process, but the foreign shocks identify movements in both domestic and foreign enrollments possibly due to displacement effects. These results are both qualitatively and quantitatively similar to our main results from table 3.

## 6. Concluding Remarks

Research scientists in American universities continue to argue strongly that their ability to develop knowledge, raise grants, and ultimately expand technology-based innovation depends on unimpeded access to the highest-caliber graduate students they can attract from anywhere in the world. This claim is intuitively plausible but before now has not been tested carefully with micro data. In the absence of such a test such claims could legitimately be criticized for potentially mixing cause and effect: perhaps the rise in both research productivity and numbers and share of international doctoral students were simply coincident or caused by other factors. In any event these arguments have not resulted in changes in basic immigration policy with respect to foreign graduate students.

More broadly, American and European policy-makers concerned with the relative decline of their economies in the face of China's and India's ascent may worry less about sunset sectors in which they no longer hold a comparative advantage, and more about their continued ability to innovate and generate new products and new markets at the frontier.[18] If foreign-born skilled labor is an important input for the production for new ideas and products, then immigration policy may be intricately tied to innovation policy.

Thus, our purpose here was to assess the causal role of domestic and international doctoral students in the production of S&E knowledge. Using detailed data on the national origins of students, academic research output, and exogenous variation in student – department matches, we demonstrate the existence of such causal effects. Both domestic and foreign students significantly increase the numbers of publications and citations in U.S. S&E departments.

Two further important results emerged from this analysis. First, the positive contribution of foreign students is muted when the students arrive as a result of macroeconomic shocks that differentially increase the proportion of paying (non-

[18] There is evidence that innovation and patenting shifted to China during during the recent (2009-2010) global recession. See "Innovation Shifted to China During the Downturn: U.N. Report," *The Huffington Post*, September 16, 2010 and "China Shifts away from Low-cost Factories," *New York Times* September 15, 2010

scholarship) students. High quality scholarship students are particularly valuable from the perspective on U.S. innovation policy. Second, we find suggestive evidence that one mechanism by which foreign students contribute to scientific productivity is by increasing departmental diversity.

That the quality of international students has a significant impact at the margin implies that U.S. student-visa policy may be misguided if an important objective is to expand the research capacity of American universities. Rather than relying largely on a demonstration of financial wealth sufficient to support graduate study and return home, a key criterion for issuing a visa could be indicators of student quality (easily measured by admission with scholarship to top-ranked programs) independent of assets or incomes.

To summarize, proponents of increasing the numbers of foreign graduate students seem to be correct, in that they clearly have had a positive effect on the conduct of science. While it is difficult to place a dollar figure on the costs of skilled-student visa restrictions in terms of reduced innovation, the fact that the American economy is dependent on new products, new markets and innovation for growth is strong indication that the consequences of limiting foreign enrollments are potentially large.

## References


Adams, J. and Clemmons J. "The NBER-Rensselaer Scientific Papers Database: Form, Nature, and Function". NBER WP 14575 (2008).

Adams, J., G. Black, J.R. Clemmons and P. Stephan (2005). "Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981-1999" *Research Policy* 34: 259-285

Alesina, A. and E. La Ferrara (2005). "Ethnic Diversity and Economic Performance," *Journal of Economic Literature* 43(3): 762-800.

Altonji, J. and D. Card (1991). "The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives" in J. Abowd and R. Freeman eds, *Immigration, Trade and the Labor Market* (Chicago, the University of Chicago Press)

Azoulay, P., W. Ding and T. Stuart (2005)."The Determinants of Faculty Patenting Behavior: Demographics or Opportunities?" NBER Working Paper 11348, May.

Becker, G. and K. Murphy (1992). "The Division of Labor, Coordination Costs and Knowledge." *Quarterly Journal of Economics* 107: 1137-1160

Berliant M. and M. Fujita (2004). "Knowledge Creation as a Square Dance on the Hilbert Cube," Washington University, Department of Economics, manuscript.

Black, G. and P. Stephan (2008). "The Economics of University Lab Science and the Role of Foreign Graduate Students and Postdoctoral Scholars." Research Paper No. 09-02, Andrew Young School of Policy Studies

Borjas, G. (1994). "The Economics of Immigration" *Journal of Economic Literature* 32: 1667-1717

Borjas, G. (2002). "An Evaluation of the Foreign Student Program." Center for Immigration Studies Backgrounder.

Borjas, G. (2003). "The Labor Demand Curve *Is* Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." *Quarterly Journal of Economics* 118(4): 1335-1374

Borjas, G. (2004). "Do Foreign Students Crowd Out Native Students from Graduate Programs." in G. Ehrenberg and P. Stephan eds, *Science and the University*, (Madison: University of Wisconsin Press).

Borjas, G. (2005). "The Labor-Market Impact of High-Skill Immigration." *American Economic Review: Papers and Proceedings* 95(2): 56-60

Borjas, G., R. Freeman and L. Katz (1997). "How Much do Immigration and Trade Affect Labor Market Outcomes." *Brookings Papers on Economic Activity* 1: 1-90

Card, D. (1990). "The Impact of the Mariel Boatlift on the Miami Labor Market." *Industrial and Labor Relation Review* XLIII: 245-257

Chellaraj, G., K. E. Maskus, and A. Mattoo, A. (2008). "The Contribution of International Graduate Students to U.S. Innovation." *Review of International Economics*, 16(3): 444-462.

Council of Graduate Schools (various years). *Survey of Graduate Enrollment*.

De Groot, H., W. W. McMahon and J. F. Volkwein (1991). “The Cost Structure of American Research Universities,” *Review of Economics and Statistics* 73(3): 424-431.

Ethier, W. (1979). “Internationally Decreasing Costs and World Trade,” *Journal of International Economics* 9(1): 1-24.

Freeman, R. B. (2005). “Does Globalization of the Scientific/Engineering Workforce Threaten U.S. Economic Leadership?” NBER Working Paper No. 11457

Freeman, R. B. (2009). “What Does Global Expansion of Higher Education Mean for the US?” NBER Working Paper No. 14962

Fujita M. and S. Weber (2004). “Strategic Immigration Policies and Welfare in Heterogeneous Countries,” Fondazione Eni Enrico Mattei (FEEM) working paper number 2.2004.

Goldberger, M., B. Maher and P.E. Flattau Eds. (1995) *Research-Doctorate Programs in the United States: Continuity and Change.* Washington, D.C.: National Academy Press

Gurmu, S., G. C. Black, and P. E. Stephan (2009). "The Knowledge Production Function for University Patenting." *Economic Inquiry*, forthcoming (available at http://www3.interscience.wiley.com/cgi-bin/fulltext/121448132/PDFSTART).

Hanushek, E. A. (1979).“Conceptual and Empirical Issues in the Estimation of Educational Production Functions.“ *Journal of Human Resources* 14(3): 351-388.

Hausman, J. (1978). “Specification Tests in Econometrics.“ *Econometrica* 46: 1251-1271.

Hunt, J. and M. Gauthier-Loiselle (2008). “How Much Does Immigration Boost Innovation?“ McGill University, manuscript.

Ichniowski, C., K. Shaw and G. Prennushi (1997). “The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines.“ *American Economic Review* 87: 291-313

Imbens, G. And J. Wooldridge (2007).“What’s New in Econometrics?“, Mini-Course Lecture Notes, NBER Summer Institute 2007, http://www.nber.org/WNE/.

Jaffe, A.B. and M. Trachtenberg (2002). Patents, Citations and Innovations: A Window on the Knowledge Economy. Cambridge: MIT Press.

Johnson, J. L. (1978). “The Role of the Student in the Higher Education Production Function,” *Research in Higher Education* 9(2): 169-179.

Kremer, M. (1993). “The O-Ring Theory of Economic Development.” *Quarterly Journal of Economics* 108(3): 551-575

Lach, S. and M. Schankerman (2008). “Incentives and Invention in Universities.” *Rand Journal of Economics* 39(2): 402-433

Mas, A. and E. Moretti (2009). “Peers at Work.” *American Economic Review* 99(1): 112-149

Mishra, P.(2007). "Emigration and Wages in Source Countries: Evidence from Mexico." *Journal of Development Economics* 82: 180-199

Mowery, D.C., R.R. Nelson, B.N. Sampat and A.A.Ziedonis (2002). *Ivory Tower and Industrial Innovation: University-Industry Technology Transfer Before and After the Bayh-Dole Act* (Stanford: Stanford University Press).

Niebuhr, A. (2006). "Migration and Innovation: Does Cultural Diversity Matter for Regional R&D Activity?" IAB discussion paper number 14.

OECD (2006). *PISA 2006 Science Competencies for Tomorrow's World*, available at http://www.pisa.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html.

Orleans, L. (1988). *Chinese Students in America: Policies, Issues and Numbers* (Washington, D.C.: National Academy Press)

Ottaviano, G.I.P. and G. Peri (2005). "Rethinking the Gains from Immigration: Theory and Evidence in the U.S." CEPR Working Paper No. 5226

Ottaviano, G.I.P. and G. Peri (2006). "The Economic Value of Cultural Diversity: Evidence from U.S. Cities," *Journal of Economic Geography* 6(1): 9-44.

Pritchett, L. and D. Filmer (1997). "What Education Production Functions *Really* Show: A Positive Theory of Education Expenditures," World Bank, manuscript.

Sakellaris, P. and A. Spilimbergo. (2000) "Business Cycles and Investment in Human Capital: International Evidence on Higher Education." *Carnegie-Rochester Conference Series on Public Policy* V. 52 (June).

Sparber, C. (2009) "Racial Diversity and Aggregate Productivity in US Industries: 1980-2000" *Southern Economic Journal* 75 (3): 829-856

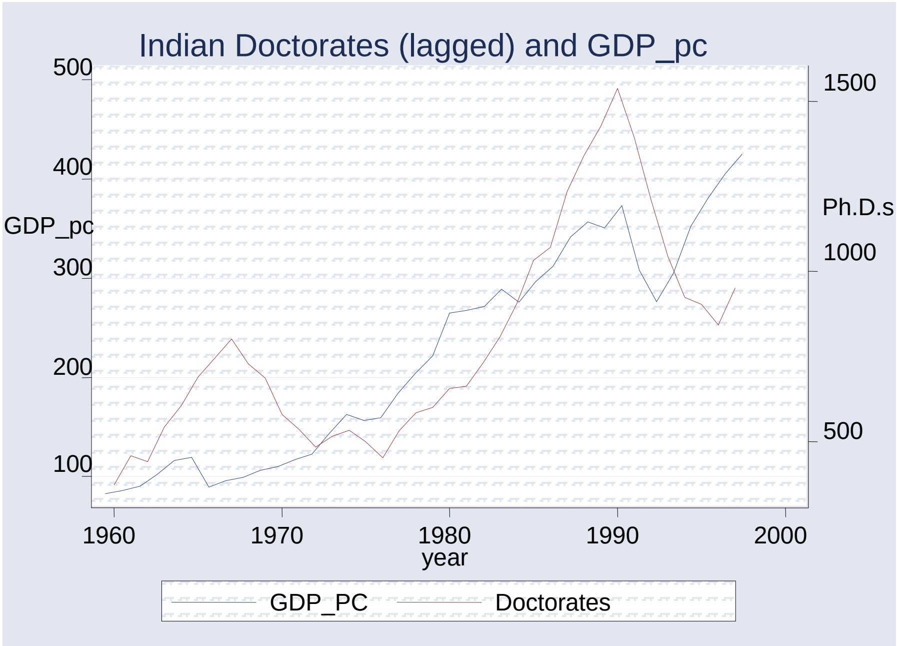
Thursby, J.G. and S. Kemp, (2002). "Growth and Productive Efficiency of University Intellectual Property Licensing." *Research Policy* 31:109-124.

Thursby, J.G. and M.C. Thursby, (2002). "Who is Selling the Ivory Tower? Sources of Growth in University Licensing." *Management Science* 48: 90-104.

Waldinger, F. (2009). "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany." mimeo. University of Warwick

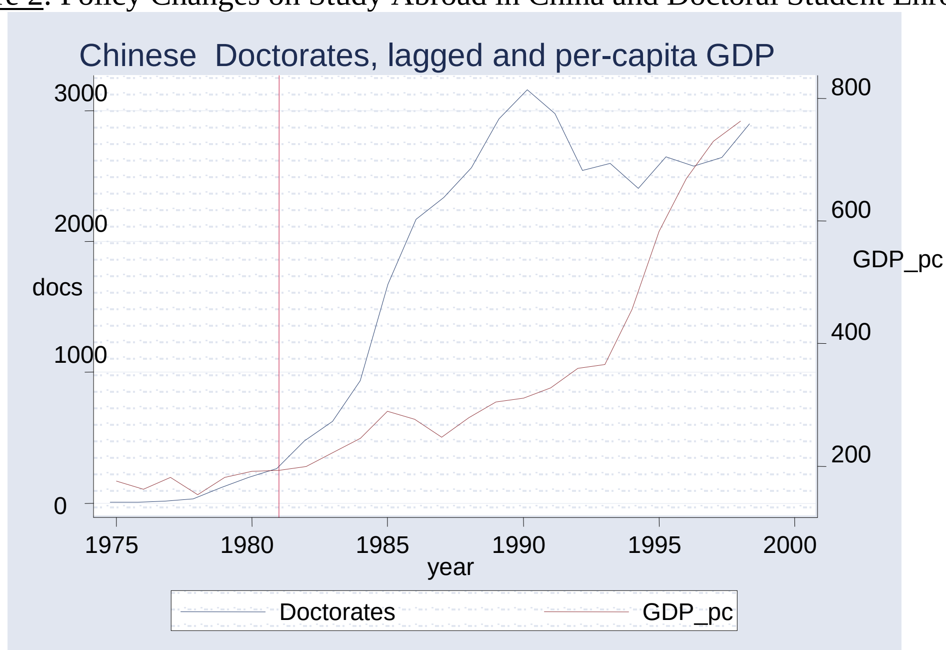
Wuchty, S., B.F. Jones and B. Uzzi (2007). "The Increasing Dominance of Teams in Production of Knowledge." *Science* 316: 1036-1039

Figure 1: Time Series of Indian Doctoral Student Enrollment and GDP per Capita



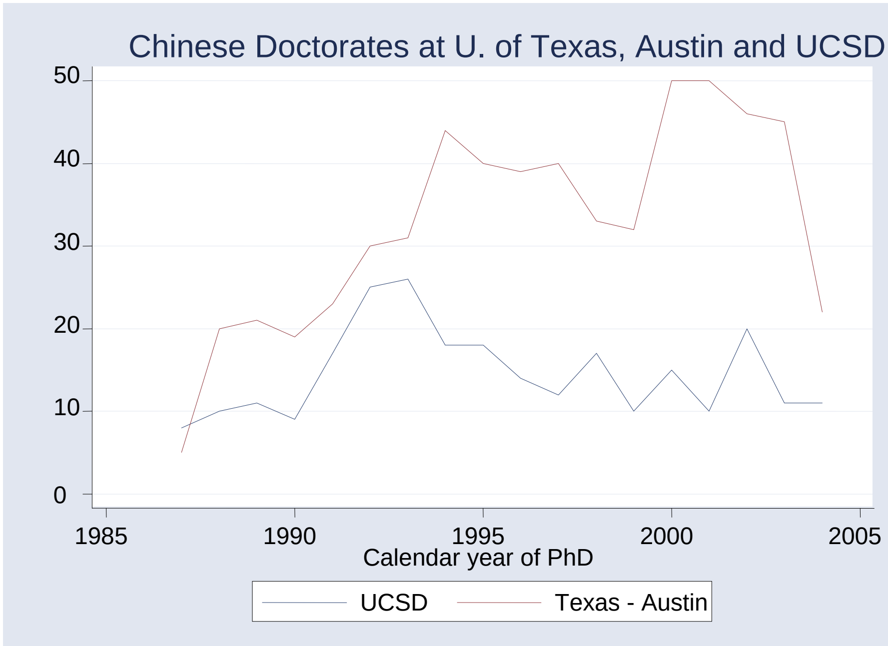

*Note: ' Lagged Doctorates' is the Series 'Number of Doctoral Degrees Received by Indians' set 6 years back (i.e. around the time those doctorate recipients were enrolling in graduate school)*

Figure 2: Policy Changes on Study Abroad in China and Doctoral Student Enrollment




*A policy shock in 1976 (Mao's Death) and normalization of relations in 1979 paved the way for the partial (1981) and total (1984) lifting of restrictions on Chinese study abroad (Orleans 1988). GDP growth in the 1980s may explain some of the magnitude in this spike of students. Partial restrictions on study abroad were re-imposed following the 1989 Tiananmen Square Protests*

Figure 3: Differential Response of Chinese Enrollment across Universities




*Note: In Figures 3 and 4 "Doctorates" refers to completed doctoral candidates in the year indicated.*

Figure 4: Differential Response of Chinese Enrollment across Disciplines

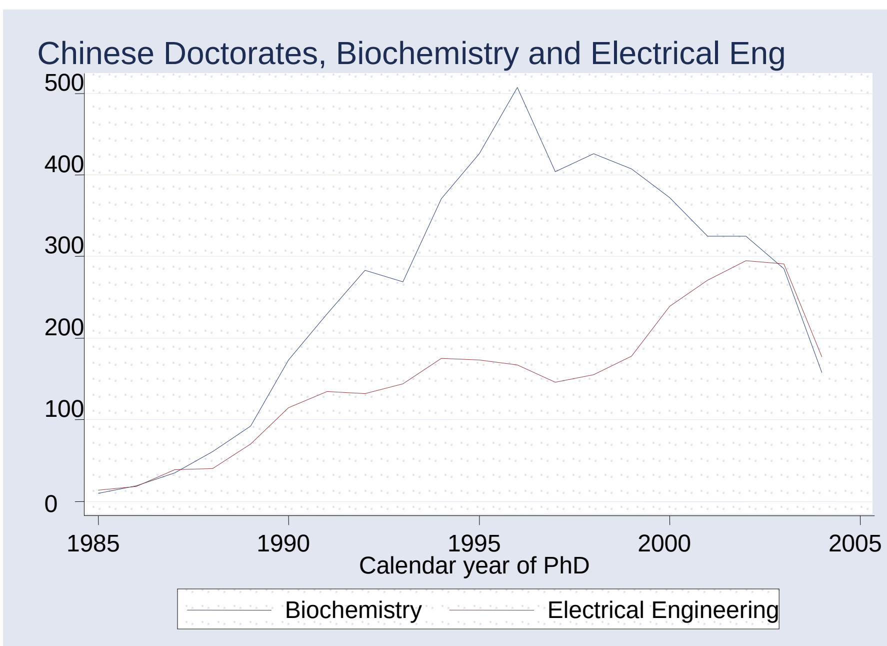


**Table1 : Summary Statistics for Panel with R&D controls**

| Variable | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|
| Publication Counts | 44.06 | 62.81 | 0 | 1159 |
| Publication Counts - 1970's | 27.56 | 39.09 | 0 | 526 |
| Publication Counts - 1980's | 42.22 | 58.03 | 0 | 807 |
| Publication Counts - 1990's | 58.17 | 77.41 | 0 | 1159 |
| Publication Citation Counts | 1400.16 | 2925.61 | 0 | 71051 |
| Citation Counts - 1970's | 929.61 | 1868.14 | 0 | 30056 |
| Citation Counts - 1980's | 1365.91 | 2636.96 | 0 | 53257 |
| Citation Counts - 1990's | 1780.14 | 3723.19 | 0 | 71051 |
| PhD Enrollment - Total | 46.79 | 55.32 | 0 | 498 |
| PhD Enrollment - Foreign | 16.87 | 23.33 | 0 | 317 |
| PhD Enrollment - U.S. | 29.92 | 38.41 | 0 | 380 |
| International Diversity Index | 0.64 | 0.27 | 0 | 1 |
| Equipment and Physical Plant Expenditures | 0.75 | 1.29 | 0 | 18.23 |
| Other R&D Expenditures | 3.68 | 6.01 | 0 | 103.53 |

N=47959, observations on university-field "departments"

Number of university-field groups: 2214

Note: Expenditures measured in $ millions, 5 year moving average.

**Table 2: First stage; PhD enrollments on instrumental variables**

| | | Full Set | | Pay-neutral Set | | Pay-affecting Set | |
|---|---|---|---|---|---|---|---|
| Dependent variables: one-year lags of US and Foreign enrollment | | US students (1) | Foreign Students (2) | US students (3) | Foreign Students (4) | US students (5) | Foreign Students (6) |
| Percent of world population free from study-abroad restriction | 1 year lag | -63.7388*** | 37.9390*** | -39.5978*** | 45.5437*** | | |
| | | (11.4382) | (8.8541) | (7.9553) | (5.8340) | | |
| | 2 year lag | 80.6168*** | 8.4193 | 30.9430*** | 6.6105 | | |
| | | (16.9055) | (15.6548) | (6.8043) | (6.1463) | | |
| Tertiary students abroad, non-US hosts | 1 year lag | 0.0024 | 0.0458*** | 0.0149* | 0.0157** | | |
| | | (0.0078) | (0.0076) | (0.0089) | (0.0075) | | |
| | 2 year lag | -0.0224** | -0.0129 | 0.0029 | -0.0860*** | | |
| | | (0.0099) | (0.0081) | (0.0161) | (0.0135) | | |
| Real average GDP per-capita, East Asia | 1 year lag | 0.0190*** | 0.0022 | | | 0.0019 | 0.0132*** |
| | | (0.0036) | (0.0031) | | | (0.0025) | (0.0022) |
| | 2 year lag | -0.0098** | -0.0190*** | | | 0.0105*** | -0.0210*** |
| | | (0.0042) | (0.0042) | | | (0.0022) | (0.0022) |
| Real average GDP per-capita, South Asia | 1 year lag | -1.7692*** | 1.2485*** | | | -1.4078*** | 2.0616*** |
| | | (0.4534) | (0.4448) | | | (0.3780) | (0.3678) |
| | 2 year lag | 0.3558 | 0.2569 | | | 0.2769 | -0.0885 |
| | | (0.2788) | (0.2947) | | | (0.2771) | (0.2905) |
| Real average GDP per-capita, Mid-East and Africa | 1 year lag | -0.0995** | 0.0849* | | | -0.0680* | 0.1732*** |
| | | (0.0463) | (0.0435) | | | (0.0378) | (0.0322) |
| | 2 year lag | -0.0292 | 0.0790** | | | -0.0115 | 0.0780* |
| | | (0.0318) | (0.0392) | | | (0.0330) | (0.0398) |
| Real average GDP per-capita, Europe and Russia | 1 year lag | -0.0048 | -0.0097*** | | | -0.0153*** | -0.0017 |
| | | (0.0035) | (0.0031) | | | (0.0035) | (0.0028) |
| | 2 year lag | -0.0128*** | -0.0042 | | | -0.0020 | -0.0042 |
| | | (0.0034) | (0.0031) | | | (0.0030) | (0.0026) |
| Real average GDP per-capita, Western Hemisphere | 1 year lag | -0.2532** | -0.0371 | | | -0.1948** | 0.0484 |
| | | (0.1084) | (0.0833) | | | (0.0947) | (0.0689) |
| | 2 year lag | 0.5446*** | 0.1111 | | | 0.5615*** | 0.1559** |
| | | (0.0855) | (0.0692) | | | (0.0862) | (0.0696) |
| Mean unemployment rate, New England census region | 1 year lag | 42.9613*** | 3.4698 | 41.2849*** | 6.0653 | 48.8791*** | -8.4576 |
| | | (7.8891) | (5.4648) | (7.8182) | (5.4854) | (7.9667) | (5.3693) |
| | 2 year lag | 19.4931* | -6.1023 | 27.9483*** | -13.3855* | 12.9317 | -4.1314 |
| | | (10.3964) | (6.9819) | (10.7540) | (7.2408) | (10.2040) | (6.9021) |
| Mean unemployment rate, East South Central census region | 1 year lag | 7.8207 | 31.7662*** | 12.4046 | 31.7676*** | 13.2604 | 22.5821*** |
| | | (8.8095) | (8.3616) | (8.8164) | (8.4936) | (8.9505) | (8.1981) |
| | 2 year lag | -38.7347*** | 1.3104 | -40.3514*** | -13.9126 | -44.0968*** | 1.2278 |
| | | (9.0992) | (9.5045) | (8.8077) | (9.4643) | (9.0714) | (9.4891) |
| Average GSP, E. S. Central census region | 1 year lag | | | 0.0108*** | 0.0116*** | 0.0109*** | 0.0124*** |
| | | | | (0.0036) | (0.0031) | (0.0036) | (0.0031) |
| Average GSP, Mountain census region | 1 year lag | | | 0.0284*** | 0.0026 | 0.0285*** | 0.0051 |
| | | | | (0.0094) | (0.0081) | (0.0094) | (0.0081) |
| Observations | | 47954 | 47954 | 47954 | 47954 | 47954 | 47954 |
| No. of Field-University Pair FE | | 2209 | 2209 | 2209 | 2209 | 2209 | 2209 |
| R-squared | | 0.25 | 0.57 | 0.24 | 0.57 | 0.25 | 0.57 |

Robust standard errors in parentheses, Errors clustered by field-university pairs

* significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include field-university pair and year fixed effects, university and field specific trends and R&D measures.

All instruments interacted with region's start-of-decade university and field fractions of enrollment

**Table 3: OLS and LIML estimates of PhD student research productivity, full sample**

| Dependent Variable: | Publications / Dept / year | | | | Citations to dept.'s pubs, / year | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | OLS | OLS | LIML | LIML | OLS | OLS | LIML | LIML |
| U.S. students | 0.164*** | 0.1537*** | 0.8365* | 0.7449 | 6.721*** | 6.6520*** | 78.0567*** | 73.7838** |
| | (0.032) | (0.0320) | (0.5080) | (0.4717) | (1.587) | (1.5793) | (29.0457) | (29.8487) |
| International students | 0.152*** | 0.1347*** | 0.9667*** | 0.9244*** | 0.473 | 0.3518 | 46.3826*** | 47.7693** |
| | (0.033) | (0.0332) | (0.3261) | (0.3441) | (1.444) | (1.4453) | (16.5979) | (19.1215) |
| Equipment and physical plant expenditure (5 yr MA, $millions) | | -0.1740 | | -0.3870 | | -16.0653 | | -34.5360 |
| | | (0.3028) | | (0.3919) | | (17.2549) | | (27.0504) |
| R&D expenditure (5 yr MA, $ millions) | | 0.4776*** | | -0.2837 | | 4.4977 | | -58.2408* |
| | | (0.1543) | | (0.4640) | | (6.0384) | | (30.2283) |
| Observations | 47959 | 47959 | 47954 | 47954 | 47959 | 47959 | 47954 | 47954 |
| No. of Field-University Pair FE | 2214 | 2214 | 2209 | 2209 | 2214 | 2214 | 2209 | 2209 |
| R-squared: | 0.51 | 0.52 | | | 0.34 | 0.34 | | |
| First-Stage Statistics | | | | | | | | |
| F-stat of excluded IV on U.S. students: | | | 10.72 | 10.94 | | | 10.72 | 10.94 |
| F-stat of excluded IV on International: | | | 12.42 | 11.52 | | | 12.42 | 11.52 |
| F- or Chi-squared stat. for parity of US and International Coefficients: | 0.09 | 0.21 | 0.11 | 0.37 | 9.78 | 9.94 | 1.05 | 1.52 |
| P-value of test: | 0.7669 | 0.6451 | 0.737 | 0.544 | 0.0018 | 0.0016 | 0.305 | 0.218 |
| Elasticity at means: | 0.111 | 0.104 | 0.513 | 0.457 | 0.144 | 0.142 | 1.658 | 1.567 |
| Elasticity at means: | 0.058 | 0.052 | 0.329 | 0.315 | 0.006 | 0.004 | 0.547 | 0.564 |

Robust standard errors in parentheses

Errors clustered by field-university pairs

* significant at 10%; ** significant at 5%; *** significant at 1%

All regressors lagged one year.

All regressions have field-university pair and year fixed effects, and university and field specific trends.

Full set of instruments used in all limited-information maximum likelihood regressions

**Table 4: LIML estimates of PhD student research productivity, by IV set**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable: | Publications | | Citations | |
| IV set: | Pay-neutral | Pay-affecting | Pay-neutral | Pay-affecting |
| U.S. students | 1.4086 | 0.6245** | 36.9038 | 54.2989** |
| | (1.0439) | (0.3065) | (22.5384) | (24.2585) |
| International students | 1.6205* | 0.8147*** | 49.4975*** | 31.5138* |
| | (0.8478) | (0.2522) | (17.4995) | (17.2043) |
| Equipment and physical plant expenditure (5 yr MA, $millions) | -0.6004 | -0.3506 | -28.1470 | -28.8377 |
| | (0.5921) | (0.3623) | (21.4317) | (23.0143) |
| R&D expenditure (5 yr MA, $ millions) | -1.0314 | -0.1573 | -39.3599 | -38.6324 |
| | (1.1108) | (0.3376) | (24.3344) | (24.7653) |
| Observations | 47954 | 47954 | 47954 | 47954 |
| No. of Field-University Pair FE | 2209 | 2209 | 2209 | 2209 |
| First-Stage Statistics | | | | |
| F-stat of excl. IV on U.S. students: | 15.96 | 10.47 | 15.96 | 10.47 |
| F-stat of excl. IV on International: | 17.93 | 12.19 | 17.93 | 12.19 |
| Chi-squared stat. for parity of US and International Coefficients: | 0.40 | 1.60 | 0.87 | 5.97 |
| P-value of test: | 0.525 | 0.206 | 0.351 | 0.015 |

Robust standard errors in parentheses

Errors clustered by field-university pairs

* significant at 10%; ** significant at 5%; *** significant at 1%

Both instrument sets include mean unemployment rate and mean Gross State Product by US census region.

All regressions have field-university pair and year fixed effects, and university and field specific trends.

All instruments interacted with region's university and field fractions

| **Table 5: Inclusion of Diversity Index** | (1) OLS | (2) OLS |
|---|---|---|
| Dependent Variable: | Publications | Citations |
| U.S. students | 0.1511*** | 6.5550*** |
| | (0.0319) | (1.5662) |
| International students | 0.1347*** | 0.3511 |
| | (0.0331) | (1.4442) |
| Index of International Diversity | 2.5433*** | 95.5378** |
| | (0.7541) | (45.5329) |
| Equipment and physical plant expenditure (5 yr MA, $millions) | -0.1627 | -15.6389 |
| | (0.3024) | (17.2526) |
| R&D expenditure (5 yr MA, $ millions) | 0.4775*** | 4.4948 |
| | (0.1545) | (6.0508) |
| Observations | 47959 | 47959 |
| Number of group(univ field) | 2214 | 2214 |
| R-squared | 0.52 | 0.34 |
| Elasticity of publications with respect to International Diversity, at mean: | 0.04 | 0.04 |

Robust standard errors in parentheses

Errors clustered by field-university pairs
* significant at 10%; ** significant at 5%; *** significant at 1%

All regressors lagged one year.

All regressions have field-university pair and year fixed effects, and university and field specific trends.

International Diversity Index is Herfindal index of shares of international enrollment for seven foreign regions.

| **Table 6: LIML estimates, excluding regional US instruments** | | | | |
|---|---|---|---|---|
| Dependent Variable: | Publications | | Citations | |
| | (1) | (2) | (3) | (4) |
| | LIML | LIML | LIML | LIML |
| U.S. students | 1.7658* | 1.6134* | 141.8275* | 138.3362* |
| | (0.9195) | (0.8289) | (74.1915) | (78.9811) |
| International students | 1.1692*** | 1.1617** | 77.0840** | 85.0233* |
| | (0.4435) | (0.4650) | (36.6860) | (44.5904) |
| Equipment and physical plant expenditure (5 yr MA, $millions) | | -0.4291 | | -46.7472 |
| | | (0.5667) | | (43.8125) |
| R&D expenditure (5 yr MA, $ millions) | | -0.8850 | | -112.1993* |
| | | (0.7255) | | (68.0223) |
| Observations | 47936 | 47936 | 47936 | 47936 |
| No. of Field-University Pair FE | 2174 | 2174 | 2174 | 2174 |
| First-Stage Statistics | | | | |
| F-stat of excluded IV on U.S. students: | 9.01 | 9.39 | 9.01 | 9.39 |
| F-stat of excluded IV on International: | 13.65 | 12.20 | 13.65 | 12.20 |

Robust standard errors in parentheses
Errors clustered by field-university pairs
* significant at 10%; ** significant at 5%; *** significant at 1%
All regressors lagged one year.
All regressions have field-university pair and year fixed effects, and university and field specific trends.
All instruments interacted with region's university and field fractions

**Data Appendix for "Skilled Immigration and Innovation: Evidence from Enrollment Fluctuations in U.S. Doctoral Programs"**

**Not for Publication**

***Fields of Science and Engineering***

1 Mathematics
2 Computer Science
3 Statistics/Biostatisics
4 Chemistry
5 Physics
6 Astrophysics/Astronomy
7 Geosciences
8 Oceanography
9 Biochemistry/Molecular Biology
10 Genetics
11 Neurosciences
12 Pharmacology
13 Physiology
14 Cellular and Development Biology
15 Ecology, Evolution and Behavior
16 Aerospace Engineering
17 Biomedical Engineering
18 Chemical Engineering
19 Civil Engineering
20 Electrical Engineering
21 Industrial Engineering
22 Materials Engineering
23 Mechanical Engineering

***Dependent Variables: Publication Counts and Citations***

We chose 100 universities based on (highest) total doctoral degrees granted to foreign students. Ninety of these universities also were the top total Ph.D.-granting institutions. We collected data on all publications by those universities in S&E fields from 1973 to 2001. The data were downloaded from Thomson ISI's *Web of Science*, using a Perl script. Each publication record included the university ID, year, number of times cited, subject category or categories and department affiliation(s). Using an algorithm (described below), we sorted the publication records into 23 fields of science and engineering. We then constructed the number of publications per university/field/year and the sum(1+times cited) per university/field/year.

Since *Web of Science* does not standardize department abbreviations, we started with typical abbreviations, which were closely aligned to the 23 fields (e.g., the typical abbreviation for a mathematics department is "dept math").

- Searching with typical abbreviations, we identified the 5,000 most highly cited publications within each field.
- Using *Web of Science's* assignment of publications to subject categories, we identified all subject categories referenced by at least one percent (50) of those publications, for each field.

- In order to ensure that all publications related to the core literature of each field were assigned to the correct field, we designated categories identical or very close to the field name as unique, and removed them from the other fields' listings. Categories that were truly unique were also designated.

The sorting algorithm is as follows:

**1.** If there is only one subject category listed by the publication and:
**a)** it is a unique category, it is assigned directly to the associated field;
**b)** it is a non-unique category, but the associated typical department is listed and matches a field, then it is assigned to the associated field;
**c)** it is a non-unique category, and the department does not match a field, it is assigned to the highest ranking field (see below) that is associated with the subject category.

**2.** If there are multiple subject categories listed and:
**a)** the department listing matches a field, it is assigned to that field;
**b)** the department does not match a field, and there is only one unique subject category, it is assigned to the field associated with that subject category;
**c)** the department does not match a field, and there are multiple unique subject categories, it is assigned to the field associated with the highest ranked unique subject category;
**d)** the department does not match a field, and there are no unique subject categories, then it is assigned to the field associated with the most subject categories listed;
**e)** the department does not match a field, there are no unique subject categories, and several fields are tied for the most subject categories, then of the tied fields, assign the publication to the highest ranked field.

**3.** If there are no subject categories listed and:
**a)** the department listing matches a field, it is assigned to that field;
**b)** the department listing does not match a field (or there is no department listing), the publication cannot be assigned.

In all, some 3.2 million records were collected, of which 290,000 could not be assigned with this algorithm. The distribution of records among fields is not uniform, but not heavily skewed either. Computer science has the least records, around 40,000, while ecology, evolution and behavioral biology has the most, around 520,000 records. Priority in ranking fields was given to fields with specific topics of inquiry, such as neuroscience and aerospace engineering, over fields with methods of inquiry, such as biochemistry and mechanical engineering. Of the 3.2 million records, many are duplicates, having been assigned to multiple universities on account of co-authorship by researchers at several institutions.

.

***Independent Variables***

**a) Graduate Student Enrollment Counts**

Data on graduate student enrollments were compiled from the NSF's *Survey of Earned Doctorates*, a survey requested of every doctorate recipient upon completion of that degree. The survey has been consistent in its core questions from 1959 to the present. For key identifying variables the NSF inferred responses from the location and time of the survey, so that doctoral institution and year of graduation are identified with a response rate of 100 percent. Other key variables, such as country of citizenship, year of graduate entry and dissertation field had response rates on the order of 90-95 percent.

Students were assigned to fields based on their indicated three-digit dissertation specialty. The SED uses 340 of these titles to categorize specific areas of study, of which 189 are related to science and engineering. We matched these 340 specialties to the 23 fields of science and engineering used in the National Research Council's 1993 *Survey of Graduate Faculty*, and to a twenty-fourth field, which we call non-science. This matching, although ad hoc, was for the most part obvious. When not obvious, assignment was made using information from the list of subject categories (discussed above to match publications to fields). In cases where a dissertation specialty seemed to match more than one product category, students indicating that specialty were randomly distributed to the product categories.

Using information in the SED on year of graduation and year of entry,[1] we assumed that the respondent was enrolled at his doctoral institution for the intervening years before completion. We thus created an inferred enrollment count, whereby each Ph.D. recipient was counted in a university/field/year observation for each year of enrollment. This assumption may slightly overstate enrollments due to breaks in attendance. However, since the SED does not record people who leave before completing their doctorate, the enrollment counts may as likely be an underestimate. One difficulty with inferred enrollments is that since the SED only goes to 2004, such counts for the most recent years underestimate total enrollment. Since the observed average time to degree is six years, a student entering in 1999 would graduate in 2004. To be conservative, we use inferred enrollment counts only through 1997, although counts for 1995-1997 will have some slight truncation because students finishing in 2005 and 2006 would not be included.

**b) Elite Universities**

To create an indicator of elite and non-elite universities of enrollment, we considered U.S. institutions as elite if the 25$^{th}$ percentile of undergraduate SAT (ACT) scores was at least 1220 (26), as reported by *US News and World Report*, 2005.

**c) Control Variables: R&D and Equipment Expenditures**

Two variables were created with data from the NSF's *Survey of R&D Expenditures at Universities and Colleges*, available online through WebCaspar (webcaspar.nsf.gov). The survey contains total R&D by university, field and year, and has sub-totals by federal/non-federal funding and equipment (capital) expenditures. The fields in this survey were slightly different than those contained in the *Survey of Graduate Faculty*, and our correspondence between them is available upon request. The first measure we create from this data is real non-equipment R&D expenditures, which includes both federal and local funding, but is net of equipment and capital (physical plant) expenditures. This sum includes mainly administrative costs and payments to professors, post-docs and graduate students. The second variable is real equipment & physical plant expenditures, which again includes both federal and local funding.

### *3. Instrumental Variables*

**a) Per-capita GDP**

[1] To be precise, the SED includes multiple variables indicating year of entry. We used the one with the highest rate of response and if omitted used the next most common, and so on.

GDP data were constructed from the World Bank's *World Development Indicators* (WDI) series of real GDP (in year 2000 U.S. dollars), divided by the WDI series of population. Data for Taiwan were taken from the *Penn World Tables*, while figures for the U.S.S.R. prior to its breakup were from estimates compiled by Angus Maddison. For instruments at the regional level of aggregation, the median per-capita GDP of each region was used.

**b) Percentage Change in Exchange Rate**

Exchange rate data were constructed from the IMF's *International Financial Statistics* series of domestic currency/SDR exchange rates. Our variable is the annual percentage change in the exchange rate. At the regional level of aggregation we used both the median percentage change and maximum percentage change of each region.

**c) Total International Students to Non-U.S. Hosts**

Data on international student enrollment at the tertiary (undergraduate and graduate) level came from UNESCO's *Statistical Yearbooks* 1963-1998, and UNESCO's online database for post-1998. The data are reported as a count, with observations by origin/host/year. Our variable is total students per origin per year, which we made by first linearly interpolating missing values in the origin-host series, then summing across non-U.S. hosts to create the aggregate variable. At the regional level the sum of students from the region is used.

**d) OECD**

The OECD variable is a dummy for OECD membership at the beginning of our panel. It is interacted with per-capita GDP as another instrument. At the regional level, OECD membership is averaged.

**e) State Control Policy**

This dummy variable takes the value of unity if official state policy in the given year prohibited citizens from studying in the United States and zero otherwise. At the regional level, it is averaged. It has a value of one in the following cases: China (pre-1978), Russia (pre-1986), Poland (pre-1972), Germany, Romania, Bulgaria, Czechoslovakia and Hungary (all pre-1990), and Cuba (whole sample). Detailed documentation of such policies is available upon request. In brief, the seven Soviet and Eastern Bloc countries dictated student enrollment and prohibited travel. While Poland reformed its policies and relaxed restrictions in the early1970s, the others remained autocratic. With the introduction of *glasnost* in the U.S.S.R. in 1986, small exchanges of students with the United States began, but the other Eastern Bloc countries resisted this change. Only with the revolutions of late 1989 was state control relaxed in those countries. Germany is considered a state-control country because, post-reunification, East Germans are counted among all German students, and so the East German policy effectively restricted the numbers of German students in our panel. Cuban students have been restricted from studying in the United States for the whole period, while China officially changed its study abroad policy in 1978, two years after the death of Mao Zedong.

**f) U.S. Instruments**

State Gross Domestic Product (GSP) data was obtained from the Bureau of Economic Analysis, to create variables of average GSP within census regions.