# FOREIGN PHD STUDENTS AND KNOWLEDGE CREATION AT U.S. UNIVERSITIES: EVIDENCE FROM ENROLLMENT FLUCTUATIONS*

ERIC T. STUEN AHMED MUSHFIQ MOBARAK KEITH E. MASKUS

Abstract

We study the contribution of foreign science and engineering talent to the creation of new knowledge in the U.S. economy, a subject that has received considerably less attention than the impact of immigration on wages and employment. This paper uses panel data on 2300 science and engineering (S&E) departments at 100 large American universities from 1973 to 1998 to estimate the impact of foreign and domestic graduate students on the publications produced by those departments. Since the supply of students is endogenous to department quality, we use macroeconomic shocks and policy changes in source countries that differentially affect enrollments across fields and universities to instrument for the supply of students by region. We outline a student-department matching model, where the decisions to apply and admit are endogenous, in order to identify the source and direction of bias in the OLS estimates and to devise our instrumental variable strategy.

We aggregate micro data on every Ph.D. recipient at all major S&E departments in the United States to create student enrollment counts by source country. We locate the numbers of publications and citation-weighted publications for each academic department by automating *Web of Science* searches. The empirical analysis shows that both foreign and domestic graduate students are central inputs into knowledge creation, and that OLS estimates of the foreign student contribution is biased downwards. The impact of an additional foreign doctoral student varies by type of shock. The impact of more restrictive immigration policies depends on how they affect the quality margin of incoming foreigners.

September 10, 2007

Preliminary and Incomplete; Please do not Quote without Permission.

*Stuen is doctoral candidate and 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: maskus@colorado.edu or ahmed.mobarak@yale.edu. We thank the National Bureau of Economic Research (NBER) *Innovation Policy and the Economy* 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 and Christine Rohde for excellent research assistance. We are grateful to Benjamin Jones, Martin Boileau and participants at the 2006 NBER Summer Institute Innovation Policy and the Economy meetings and the 2006 NBER Summer Institute International Trade and Investment meetings for helpful comments.

## 1. Introduction

In this paper we explore statistically the roles that domestic and foreign graduate students play in developing knowledge in science and engineering (S&E) at U.S. universities. Knowledge is measured by scientific publications, defined at the level of disciplinary field in individual universities. It is increasingly argued in the media that the ability of American universities to undertake scientific research has become more dependent on the presence of technically trained international graduate students. However, this basic proposition has not been examined empirically at the detailed level of specific student, discipline, and university.

Since the advent of far tighter restrictions on the issuance of U.S. education visas after September 11, 2001, visa policy for foreign graduate students has become the subject of intense debate. Many argue that a more restrictive policy will harm the nation's research and innovation capacity. For example, American university officials are concerned that these restrictions could cause "…a crisis in research and scholarship…"[1] The point is made also in editorials.[2] Lawrence Summers, former president of Harvard, warned the U.S. State Department that the decline in foreign students threatens the quality of research performed at U.S. universities.[3] The problem reached the top levels of policy debate and the Bush administration recently partially relaxed visa limits. Concerns about the risk of a declining U.S. advantage in developing and deploying new technologies clearly underlie these debates.[4]

---

[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.
[2] "Visas and Science: Short-Sighted," *The Economist,* May 8, 2004.
"Security Restrictions Lead Foreign Students to Snub US Universities," *Nature,* September 15, 2004.
[3] *Financial Times*, April 28, 2004.
[4] An example of this concern is in the report "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.

There are well-known deficiencies in the U.S. secondary education system. Indeed, among OECD countries, the United States ranks near the bottom in mathematics and science achievement among eighth graders (TIMMS 2003). Despite this fact, the United States has sustained an unparalleled position as developer of new scientific knowledge, and continues to be a world leader in innovation and technology. The large number of foreign graduate students that enrolled at U.S. universities over the last 20 years may help explain this seeming inconsistency.[5] Foreign students are disproportionately more likely to earn graduate degrees in S&E, and in recent years foreign graduate students studying engineering in the United States have outnumbered their American counterparts (Council of Graduate Schools, various years)

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.[6] This reversed a 15-year trend in which foreign graduate students increased by four percent per year on average. Computer science and other S&E disciplines experienced the largest declines, as the U.S. Department of Homeland Security instituted the lengthy *Visa Mantis* security clearance program for students and researchers working in fields the government considers sensitive.

Given the concerns of university officials and researchers, in conjunction with the reduction in foreign enrollments, it is important to study: (a) whether international graduate students are, in fact, significant contributors to the development of new technological knowledge, and (b) whether domestic graduate students are effective

[5] 586,000 foreign graduate students were enrolled at U.S. universities in 2002 compared to 270,000 in Britain, the next most popular destination among foreign students (*New York Times* 2004).

substitutes for foreign students. Borjas (2002, 2004) has pointed out some potential costs of the U.S. student visa program, including the crowding-out effect on American students. Thus, it is important for policy purposes to estimate the research benefits of foreign students and their contribution relative to domestic graduate students.

We assemble a database of student enrollment counts by source-country 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 *Web of Science* publication and citation searches. Publications and citations form our measures of knowledge creation.

A recent paper (Chellaraj, et al, 2007) was the first to document that annual patent applications in the United States are strongly correlated with aggregate foreign student enrollment. This result was widely publicized by organizations concerned with universities and technology (Ehrenberg 2005; The National Academies 2005, pp. 53-59), and discussed in the media (*Financial Times* 2004; *Economic Times of India* 2005; Anderson 2006). Although this correlation between enrollments and patent productivity is a provocative finding, it could be driven by omitted variables (e.g. if student applications surge when departmental faculty quality improves), and causality is difficult to establish.

Our solution is to devise an instrumental variables strategy using the idea that macroeconomic shocks and policy changes in foreign countries can lead to some quasi-random variation in the supply of foreign students. For example, macroeconomic crises,

[6] The decline in foreign student applications was actually much larger at 28 percent, which raises the possibility that the students now enrolled in U.S. universities are of lower average ability than their

the collapse of the Soviet Union, or the opening up of China to trade and investment tended exogenously to expand student supplies, which possibly affected knowledge production in the United States. Moreover, such a shock would have differentially larger impacts on fields of study that are traditionally more popular among Russians and Chinese, and on universities that have traditionally recruited more such students.

Using data variation on the differential effects of foreign macro shocks across universities and disciplines has the advantage that even if some U.S. events that affect research happen to coincide with the foreign shock (e.g. the 1980 Bayh-Dole Act and the roughly simultaneous lifting of study abroad restrictions in China), they would be a concern for our empirical identification only if they have differential effects on publishing along the same patterns of universities and disciplines as the foreign shock that our IV strategy exploits. This would in general be much less likely than the mere coincidence of a foreign macro shock with a U.S. event.

While our approach has this important advantage, it requires us to add to our list of instruments an interaction term between the foreign macro shock and each U.S. university's and academic discipline's susceptibility to the shock.[7] This need raises the possibility that the historical composition of graduate students in a field of study affects the current faculty quality in that field, which in turn influences knowledge produced. We therefore include control variables for each department's faculty resources (annual R&D expenditures, including faculty salaries) to purge that type of correlation between our instruments and the error term in the second-stage publications equations. We also

antecedent cohorts.

[7] We create measures of the historical dependence of each university and field on students from that region in the form of the fraction of students at that university or in that discipline who were from that region at the start of each decade in our sample.

control for fixed effects for all 2300 departments, so that empirical inference is based only on changes in publishing in an academic department following fluctuations in student enrollment in that department. We add university-specific and field-specific trends to capture any linear changes in the norms regarding publishing at a particular university or a field of inquiry.

To help interpret the empirical results, we outline a theoretical model of matching between academic departments and students, where the size of each department is constrained and the decisions to apply and to admit are endogenous. The model predicts that ignoring the endogeneity of department-student matching would generally under-estimate the contribution of foreigners to research output, and it suggests the instrumental variables strategy that we employ. The model also predicts that the marginal contributions of foreign and domestic students are not necessarily equated in equilibrium if either the costs to admitting each type of student are not the same or if students are valued for some characteristic other than research productivity.

We find that instrumentation makes a significant difference in the estimated impacts of international graduate students. Both foreign and domestic students are positive inputs in departmental knowledge production (even after controlling for the fixed effects, trends, and departmental resources), and exogenous shock-induced changes in foreign-student enrollments generally have a greater positive impact on departmental output. Further, the nature of the shock matters. Enrollments induced by shocks that send more non-scholarship students to the United States have a smaller positive impact on knowledge production. Overall, the marginal foreign student is neither clearly better nor clearly worse than the marginal American. Foreign students contribute relatively more in terms of citations and at elite universities. Finally, there are significant variations in the marginal productivity of students across source regions, which is consistent with a model of search behavior between universities and graduate students of differential quality levels.

The paper proceeds as follows. In the next section a brief literature review is offered. In Section 3 we sketch a matching model that may be used to inform the interpretation of econometric results. In Section 4 we develop the methodology for instrumenting students and identifying shocks to enrollments, noting the performance of the instrumental variables. In Section 5 we describe our data and in Section 6 we interpret the econometric results. We offer concluding remarks in Section 7.

## 2. Prior Literature

While the presumption that graduate students in S&E are central inputs into the development of new knowledge is intuitive and sufficiently powerful for graduate departments to advocate policy changes, it has not been rigorously tested in statistical terms. We are unaware of studies that have linked the presence of graduate students to the number of publications by university and field, or examined issues arising from endogenous student supply, or substitution possibilities between foreign and domestic graduate students.

An alternative knowledge output is innovation and the determinants of university patenting are the subject of extensive recent inquiry (e.g., Thursby and Thursby, 2002; Thursby and Kemp, 2002; Jaffe and Trachtenberg, 2002). An informative history of this process is in Mowery, et al (2004). Also studied are the determinants of individual faculty patenting behavior, such as prior publications and patent stock of the scientist's university (Azoulay, et al, 2005) and, in reverse, the impact of faculty patenting on scientific productivity measured by publications and citations (Breschi, et al, 2005). On the international migration of students, Rosenzweig (2006) examines the reason why students from poor countries seek schooling in rich countries, building on the brain-drain literature (Bhagwati and Hamada, 1974; Docquier and Marfouk 2006).

Two recent studies try to statistically link the presence of foreign graduate students to future patenting. Chellaraj, et al (2007) was the first academic paper to address this issue, and it documents a strong positive correlation between the presence of such students and patenting activity in the United States. The paper shows, using regressions of the total number of patent applications and patents awarded in aggregate annual U.S. data, that a 10 percent rise in (lagged) foreign student presence increases patent applications by 4.7 percent and patents granted to universities by 5.3 percent. These findings survive a variety of sensitivity checks, including estimation of co-integration relationships, but causal inference is difficult because the correlations are based on aggregate annual data, which leaves open the possibility that other unobserved factors may be driving both patenting behavior and foreign graduate student enrollment.

A follow-on comment by Stephan, et al (2006) improves this specification by regressing the number of patent grants made to individual universities on measures of domestic and foreign PhD recipients and post-doctoral students, controlling for the number of faculty per institution and the presence of a technology transfer office. They find that international post-doctoral students contribute positively to university patenting but the impact of foreign doctorates depends on visa status. Isolating causation is difficult in this setting as well, since enrollments can respond to events correlated with patenting.

## 3. Modeling Framework

*3a. Insights from Neoclassical Production Function Analysis*

The simplest modeling framework within which to interpret our statistical results would be to consider domestic and foreign graduate students to be inputs into a neoclassical production function. There is an extensive literature on specification and estimation of educational production functions.[8] These models generally assume either one output (e.g., student test scores) or multiple outputs (e.g., graduate diplomas and research outputs) produced using a variety of inputs such as faculty size, university budgets, and research funding. 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, if university research departments truly were attempting to maximize production of publications, they would admit domestic and international graduate students to the point where the contribution of each type, scaled by some measure of department-level cost of educating them, were equalized. Such costs could include tuition subventions, living stipends, and the opportunity costs of faculty time in training them. 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, there is likely to be more uncertainty about

[8] Hanushek (1979) is the seminal reference. See also Johnson (1978) and de Groot, et al (1991) for examples.

student ability for applicants from more remote foreign areas. Accordingly, we should observe higher domestic enrollments and lower marginal products.

However, university departments generate multiple outputs, including publications 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 products in teaching, with the latter objective valued at some cost relative to publications (Pritchett and Filmer, 1997). If, for example, university departments found domestic students to offer greater productivity as teaching assistants than foreign students, and a relatively high value were placed on teaching, they would admit larger numbers of U.S. graduate students. Again, we would observe a greater marginal contribution to producing publications from international doctoral students in equilibrium.

Thus, the production-function approach would predict that departments optimally choose students with different productivities, equalizing their marginal products in publications per dollar of cost, or allocating inputs to achieve equalized differential marginal efficiencies in multiple outputs. All marginal products should be positive. And enrollment shocks should not have a substantial impact on the production of publications, because of the ability to substitute across inputs.

*3b. Outline of a Matching Model*

We doubt that the standard approach captures the reality of what happens in graduate education, since it is inaccurate to think of research departments as operating with a fixed budget constraint and highly divisible inputs. We therefore set out the following framework, which seems more suitable for explaining graduate enrollments across departments and thinking about the estimated productivity impacts we develop. First, graduate students are indivisible and enter an enterprise that likely limits enrollments for purposes of quality control. Thus, rather than the budget being fixed, it may be more accurate to think of the number of students as being admitted subject to a ceiling, implying that the enrollment of another domestic student might crowd out a foreign student. Second, students are themselves heterogeneous in their ability to be trained and contribute to knowledge creation.

Third, students are not the only input to production, and university departments differ in terms of other resources (e.g. faculty, equipment, and research budgets) available. Students prefer to attend better universities, and through some form of matching, students of lesser quality end up at worse departments. Fourth, economic conditions can affect both the supply and quality of students available, because their incentives to apply depend on labor market conditions, such as the wage available in alternative employment.[9] Finally, there are costs of admitting students, including financial costs, dealing with immigration procedures, and language training, and students with better outside opportunities are costlier for universities to attract.

Suppose that a department *d* produces knowledge using as inputs professors and quality-weighted students from the United States (*u*) and two foreign regions, *m* and *n*:

$$K_{dt} = K_d(P_{dt}, \sum_{u=1}^{U} q_{dt}^{u} + \sum_{m=1}^{M} q_{dt}^{m} + \sum_{n=1}^{N} q_{dt}^{n}) \qquad (1)$$

$$\frac{\partial K}{\partial P} > 0, \frac{\partial K}{\partial q} > 0$$

Here, the input *P* that we call "professors" is an index of faculty size and quality and other research inputs, such as grant funding and department facilities, that jointly determine the quality of a department. The creation of knowledge increases in both faculty quality and the quality-weighted number of graduate students. Assume next that a department's capacity to enroll graduate students at any time is constrained:

$$U_{dt} + M_{dt} + N_{dt} = \overline{S_{dt}} \qquad (2)$$

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

$$C_{dt} = c^m_{dt}(M_{d,t-1}) \cdot M_t + c^n_{dt}(N_{d,t-1}) \cdot N_t \qquad (3)$$

$$c_m, c_n \geq 0 \qquad c'_m, c'_n \leq 0$$

Note that we permit these cost functions to vary between regions. The notion is that universities find it more expensive to admit students from regions where U.S. visa restrictions may be more difficult to manage or where language differences are significant and require remedial training. Thus, the costs of admitting, say, Canadian students are likely to be smaller than those of admitting Chinese or Turkish applicants. However, as departments gain more experience with students from a particular region (indexed here by lagged enrollments), these costs diminish. This will impart a tendency toward partial specialization in regional admissions over time.

Students differ in two ways. First, they have differential quality characteristics in terms of their ability to act as inputs in performing research and preparing publications.

[9] For all these reasons, it may be possible to observe negative marginal products for some students, if they supplant other students with higher inherent efficiency in producing knowledge. Indeed, we observe this outcome in some regressions where we disaggregate graduate students by region.

Each student gets her individual ‘quality’ draw from region-specific distributions of quality:

$$q^u \sim F(u), q^m \sim G(m), q^n \sim H(n) \tag{4}$$

These distributions are defined over positive support levels, since we assume that no student would offer directly negative productivity. We assume that departments are aware of the quality of each individual applicant.

Second, students are heterogeneous in their tastes for entering graduate school. Each potential applicant has an idiosyncratic taste shock ε that reflects her preference to enter a particular graduate department. A higher ε does not necessarily imply a higher degree of quality *q*. A student from a particular region chooses to apply to department *d* at time *t* if the following condition is met:

$$\begin{aligned} &P_{dt} + \varepsilon_{dt}^{u} \geq w_t^U \\ &P_{dt} + \varepsilon_{dt}^{m} \geq w_t^M \\ &P_{dt} + \varepsilon_{dt}^{n} \geq w_t^N \end{aligned} \tag{5}$$

Thus, a student only applies if the benefit she gets from studying at department *d* (which we assume is simply given by department quality *P*), plus her idiosyncratic preference to study there, must exceed the “outside option” *w* (e.g. the wage a college graduate would get in her region). Note that the decision to apply is quality independent, since for a “higher quality” person both the returns to graduate education and the outside option might be larger.

The timing of this matching process is as follows. First, based on department quality and taste shocks all potential students from each region rank every department in their intended field of study and decide where to apply. For any ε, higher-quality departments will meet the inequalities in equations (5) more than lower-quality departments. Thus, the former will attract more applications at all quality levels than the latter, though individual taste shocks could induce some high-quality students to apply at lower-ranked schools. Second, departments admit the student with the highest value added in the sense of maximizing the difference between knowledge and costs (K – C), taking into account student quality in (4) and region-based costs in (3). There is a tradeoff for departments: they wish to admit students of highest quality, regardless of source, but must account for admission and training costs. Note that experience with regions matters in these decisions over time. The conditions in inequalities (5) guarantee that the higher-ranked departments have more students to choose from and will admit higher-quality candidates.

Since students will be admitted by multiple departments, they are matched to departments by the Gale-Shapley (1962) algorithm. In our setup, departments rank all applicants from their highest to lowest contribution to (K - C) and offer admission based on this ranking, up to maximum capacity. In return, students rank all departments to which they are admitted and choose the best match. This matching algorithm assigns students to departments such that there is no department-student pair who would rather be matched than where they currently are. Put differently, no unilateral defection by either side can make both sides better off. The algorithm involves iterating matches until this property holds, making it a stable allocation of students to departments.[10]

Finally, once the matches are established, departments incur costs of training students and use them, in combination with other inputs, to produce publications.

[10] The algorithm was developed to analyze marriage matches but has been applied in other matching contexts, including allocating medical students to hospitals (Irving, 1998).

In this model, departments would bias their admission decisions in favor of applications from cheaper regions, which means those students may not contribute as much at the margin to knowledge creation as would higher-quality students from costly regions. The ability of departments (and fields) to reduce costs through experience with particular regions imparts some tendency toward specialization of those units in admission decisions.

This framework clarifies endogeneity problems arising in regression models that relate enrollments to departmental output and the nature of the resulting bias. It also suggests possible solutions. Suppose that in a particular year a specific department experiences a positive shock to *P*. This positive shock would increase a department's knowledge production independent of any changes to its enrollments of students. The shock would also attract both more Americans and more foreigners to apply. In a regression of publications on domestic and foreign student enrollments, the direction the resulting bias would depend on which type of students ultimately enroll as a result of this shock. If departments show a preference for Americans over foreigners in the admissions decision (e.g. because $c_m, c_n \geq 0$), 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.

The model also suggests that shocks to the outside option (i.e. $w^M$, $w^N$) 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 ( $c_m^{'}(M_{d,t-1}) \leq 0$ ). Thus the use of 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 the next section.

## 4. Methodology

### *4A. Basic Specifications*

The empirical analysis we conduct examines the impacts of foreign students in the aggregate, and those from different regions of origin enrolled at U.S. science and engineering Ph.D. programs, on knowledge produced in specific fields of inquiry at those U.S. universities over the period 1973-1998. The dataset has four identifiers – the students' region of origin (e.g. South Asia), the university at which students are enrolled, the field of inquiry (e.g. industrial engineering), and year. We explain variation in scientific publications and citations to them as a function of foreign and domestic student enrollments, as well as fixed effects for each field in each university, linear trends specific to each university and each field, year dummies and, in some specifications, controls for departmental equipment, capital and R&D expenditures. The fixed effects control for any time-invariant differences in characteristics across "academic departments" (i.e. university-field pairs) that may be correlated with the output produced at those departments, including, for example, any fixed level differences in faculty quality across departments within a field. The field and university specific trends can capture any linear changes in the norms regarding publishing at a particular university or a field of inquiry.

Linearized forms of the basic specifications we run are as follows:

$$Output_{, u, t} = \alpha_{f,u} + \delta_t + \gamma_u (D_u{*}Trend) + \rho_f (D_f{*}Trend) ) + \beta_1 * R\&D\_Exp_{f, u, MA(t\text{-}l,t\text{-}5)} +$$

$$+\beta_2 * U.S.\_Students_{f, u, t\text{-}l,} + \beta_3 * Foreign\_Students_{f, u, t\text{-}l} + \varepsilon_{f, u, t} \quad (6a)$$

$$Output_{, u, t} = \alpha_{f,u} + \delta_t + \gamma_u (D_u{*}Trend) + \rho_f (D_f{*}Trend) ) + \beta_1 * R\&D\_Exp_{f, u, MA(t\text{-}l,t\text{-}5)} +$$

$$+\beta_2 * Total\_Students_{f, u, t\text{-}l,} + \beta_3 * Foreign\_Share_{f, u, t\text{-}l} + \varepsilon_{f, u, t} \quad (6b)$$

where *f, u, t* index the field of study (e.g. biochemistry), university (e.g. Yale), and year (e.g. 1985), respectively, $D_f$ and $D_u$ are a set of dummy variables for each field and each university, *Trend* is a linear time trend, and $\varepsilon$ is a mean-zero error term.

*Foreign_Students* and *U.S._Students* measure international and domestic enrollments. *Total_Students* measures total enrollment of both foreign and domestic students, while *Foreign_Share* is the fraction of students that are foreign. These student variables are entered with a one-year lag and are almost always instrumented with foreign country shocks. In some later specifications we also disaggregate foreign enrollments into eight source regions, such as Western Europe and China. In others we disaggregate foreign and domestic enrollment by the quality of undergraduate institution previously attended by those doctoral students. The two proxies for departmental resources in *R&D_Exp* (capital and equipment expenditures and other R&D expenditures) are measured as a 5-year moving average.

Our main approach is to undertake linear regression in order to facilitate the two-stage least squares estimation. Recall from Table 1 that 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. However, to answer the objection that our dependent variables are counts of publications or citations, we also run negative binomial fixed effects count-data models of the following form:

$$Output_{f,u,t} = e^{X_{f,u,t} \cdot \eta} \cdot \alpha_{fu} + \varepsilon_{f,u,t} \qquad (7)$$

The vector $X_{f,u,t}$ in equation (7) encompasses all the variables in the summation between $\delta_t$ and *Foreign_Students* (inclusive) in equation (6a) above or between $\delta_t$ and *Foreign_Share* (inclusive) in equation (6b). Parameter $\alpha_{f,u}$ is a separate indicator for each field-university pair.

Given the fixed-effects negative binomial specification with time dummies and specific field and university trends, the estimates of the coefficients on the student enrollment variables ($\beta$) are consistent even in the presence of correlation between those variables and (a) time-invariant, field-university specific unobservables; (b) time-variant unobservables that are either constant across fields and universities (such as any macro shocks or changes in U.S. patent law); or (b) unobservables that follow a linear trend for each field or university (Hausman, Hall and Griliches, 1984).

The remaining objects of concern are therefore unobservable characteristics specific to academic departments that vary non-linearly over time and affect both the publications produced by those departments and their foreign student enrollments. For example, if the quality of a department improves (e.g. through greater funding and better faculty recruitment) in a dimension 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. This is likely to bias the $\beta_2$ coefficient in (6b) upward.[11] Conversely, if an improvement in the quality of a department (and therefore students' earning potential) attracts high-quality American students away from business, law and other professional degrees and into S&E fields, we may observe drops in foreign student enrollments when a department's quality improves. Under any preference for Americans in admission (e.g. due to a wider range of financial aid options available for natives, or due to their native language skills), foreign students may get crowded out in a department of limited size once more high-quality American students start applying. This is likely to bias the $\beta_3$ coefficient in (6b) downward.

*4B. Instrumental Variables Estimates*

In the presence of any unobserved time-variant department characteristic coupled with the self-selection of students into academic fields, estimates of $\beta$ would no longer be consistent, since that would imply $E(\eta|X) \neq E(\eta)$ in equation (7). In other words, both foreign and domestic student enrollments can endogenously respond to some unobserved time-variant characteristic of a university-field correlated with its output, which would lead to biased estimates of the effects of changes in enrollment.

Our solution to this problem is to use an instrumental variable estimator, where we instrument for foreign student enrollments using economic and policy shocks in the students' individual countries of origin. We try to identify shocks that influence foreigners' decisions of whether to travel to the United States for graduate study, but that are plausibly uncorrelated with the publications produced at specific academic

[11] 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.

departments. For example, an exchange-rate shock in Thailand (e.g., a currency devaluation) would affect Thai students' ability to pay for a U.S. education, and can lead to fluctuations in their enrollment at U.S. universities, but may not affect publishing in specific academic departments through any other channel. To illustrate, Figures 1 and 2 plot enrollments of doctoral students from India and Nigeria at U.S. universities against two relevant instruments (Indian GDP growth and Nigerian oil dependence). The co-movement of enrollment counts and each instrument displayed in these figures provides some preliminary indication of the power of these instruments.

Our instrumental variables estimates, therefore, only use the variation in foreign student enrollments that are a result of economic and policy shocks in students' source countries for the purpose of identifying variation in research outputs from departments where those students enroll. Any shocks to the supply of foreign students that are uncorrelated with factors related to publications in the United States allow us to identify the causal impact of changes in foreign students on department-specific outputs. Our estimation strategy uses a set of first-stage regressions where we instrument foreign-student supply with fluctuations in source-country policies (e.g., lifting of restrictions on Russian and Chinese students regarding study abroad) and economic conditions (e.g., extra revenues generated in oil-producing countries from oil price shocks, or fluctuations in the SDR – local currency exchange rates in East Asia and Latin America).

The instruments as described above vary by country-year, while our endogenous variables of interest (foreign students) have 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 are motivated by the theoretical framework in section

3b to 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, a shock to the supply of Indian students 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 are exploiting 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. Finding events in the United States that had such specific patterns of influence on academic departments is considerably more difficult 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 3-5 demonstrate the empirical relevance of these ideas. Figure 3 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, and with the subsequent growth in Chinese GDP. Figures 4 and 5 further indicate that the University of Texas benefited differentially more from this surge in Chinese enrollments than did the University of California – San Diego, and that electrical engineering departments benefited more than biochemistry departments. Even though the policy shift in China happened to coincide with the passage of the Bayh-Dole Act and the beginning of the U.S. biotechnology boom, the empirical inference on the productivity effects of student enrollments in our regressions is based on the differential effects of the China surge by department. Thus, as long as the Bayh-Dole Act or the biotechnology boom did not disproportionately benefit departments that also took advantage of the China surge[12] our estimates of the effects of foreign students remain causal.

We implement these ideas in the statistical analysis by using the following types of 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 foreign students in field $f$ from region $r$ at $t_0$]

Note that we employ aggregated regions rather than specific countries because to instrument at the country level would incorporate so many variables that multicollinearity would be a severe problem. The notion in this specification of interaction terms is that university-field pairs with a high initial share of students from, say, Latin America would be particularly susceptible to a subsequent macroeconomic shock in that region. We define three initial periods, 1970, 1980, and 1990, which updates these dependency parameters in each decade. Overall, then, our list of instruments includes both region-specific shocks and those shocks interacted with university-field shares in each decade.

We primarily rely on these foreign shocks to instrument both total enrollments and the foreign share of enrollments. For the regressions reported in one table below (Table 6) we add regional (e.g., northeastern United States) averages of state GDPs, unemployment rates, and university-aged populations (aged 20-29) to the list of instruments since these U.S. economic conditions predict American enrollments better in the first stage. However, these U.S. instruments are excluded from most of our analysis since it could be argued that American economic conditions are endogenous to departmental research output.

In our main specifications, we perform two-stage least squares using the instruments described above. In the negative binomial estimation we carry out a two-stage estimation procedure by first predicting foreign students and domestic students using an OLS first-stage regression of *Foreign_Students*, by region, on the global set of instruments and a regression of *U.S._Students* either on the global set or on U.S. regional shocks. We then use these predicted values of foreign and domestic students in the second-stage negative binomial regression (7). This two-stage procedure produces consistent estimates of $\beta$ in the negative binomial model (Mullahy, 1997), although the

[12] Figure 5 shows that this is unlikely, as electrical engineering departments took greater advantage of the Chinese student boom compared to biochemistry departments, which arguably had stronger links to the

covariance matrix estimates have to be adjusted to account for the sampling variation introduced by first-stage prediction of foreign students. [13]

## 5. Data

*5A. Publications*

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, taking 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. 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 of 66,700 (100 x 23 x 29) university-field-year cells.[14] Our final database is a count of publications and total citations in each university-field-year cell. Summary statistics are provided in Table 1. Note that both publications and citations rise over time.

*5B. Enrollment Counts*

We create Ph.D. enrollment counts for each university-field-year-country of origin cell 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

biotechnology industrial boom.

[13] We have bootstrapped standard errors for the second-stage regressions, but this is computationally intensive and, given the small standard errors on our coefficients of interest, in practical terms makes little difference in terms of statistical significance.

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 country of origin assignment 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 Lach and Schankerman, 2003) during the period 1960-1997. There are approximately 700,000 doctoral students in the sample we analyze.

Although we generally exploit country-level variation in the instrumental variables, our second-stage regressions use enrollment counts from all aggregated foreign students and from aggregated regions of origin rather than the specific country of origin. We define eight regions on the basis of economic, geographic and cultural similarities between countries, taking into account each country's relative importance as a supplier of students to American universities.

Total doctoral enrollment in the average university-field-year was 42 students, 27 of whom were American. The East Asia/Pacific region (including East and South-East

---

[14] The Data Appendix has details on the algorithm used for this assignment

Asia, Australia and New Zealand, but excluding China) was the next largest supplier of students at 4.8, followed by China and then South Asia. Enrollments for U.S. and foreign students are summarized in Table 1.

*5C. Instrumental Variables*

We describe below the instruments we use for the first-stage prediction of PhD student enrollments, and then discuss the power of the instruments and report the first-stage results.

*(1) Fluctuations in the Special Drawing Right – Foreign Currency Exchange Rate*:

This measure exploits the idea that movements in, say, the Baht – SDR exchange rate during the Asian financial crisis altered Thais' ability to pay for a U.S. education. Currently, 70 percent of foreign students in the United States are from Asian and Latin American countries, suggesting that financial-crises-related indicators could serve as appropriate instruments. The raw data are strongly suggestive that this instrument is likely to be powerful. Student enrollments from Korea, Thailand, Malaysia and Indonesia grew by 41 percent between 1992 and 1997, but dropped by 15.5 percent during the financial crisis years (1997-1999).

*(2) GDP per Capita in Source Countries*:

This variable should capture long-term changes in foreigners' ability to pay for a U.S. education. To illustrate, Figure 1 shows the evolution of Indian GDP per capita and Indian doctorate recipients from 1960-2004. However, GDP growth can have the opposite effect in relatively rich countries as it increases employment opportunities in local markets (Sakellaris and Spilimbergo, 2000). To capture such non-monotonic effects, we interact GDP measures with an OECD country indicator.

*(3) Oil Dependence (Oil Share of GDP)*:

Worldwide fluctuations in the price of oil can have a powerful impact on the ability to pay of students from oil-exporting countries, as indicated in Figure 2, which plots Nigerian doctorate recipients and its oil share of GDP. We avoid using data on oil price shocks directly, since a commodity price shock can affect U.S. economic activity, which may in turn be related to research activity in the United States.

*(4) Policy Changes*:

We create an indicator for country-years where official state policies prohibited students from studying in the United States. As a specific example, this indicator captures the gradual lifting of the ban on study abroad by Chinese S&E students between 1978 and 1984 following the death of Mao Tse Tung (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 has further details.

*(5) International Students at non-U.S. Hosts*:

Using the UNESCO *Statistical Yearbooks* (1963 – present), we create counts of the number of 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.

## 6. Results

### *6A. First-stage Instrumental Variables Regressions*

Table 2 reports the first-stage instrumental variables regressions for foreign students, where we predict enrollments from each region of origin using the sets of instruments described above.  In each case the foreign-shock variable (e.g. exchange rate movements, policy shifts, GDP per capita changes) appears: (a) by itself, (b) interacted with the university's dependence on that foreign region (i.e., fraction of university's foreign students from that region at the beginning of the decade), (c) interacted with the field's dependence on that foreign region (i.e., fraction of foreign students studying in that field at the beginning of the decade who were from that region), and (d) a triple interaction of the foreign-shock variable with the field's dependence and the university's dependence.

This list of instruments may be invalid if the past composition of graduate students in a department has an impact its publication output today.  Since all of our regressions control for a fixed effect for every department, this is a concern only if the start-of-decade composition of graduate students is correlated with some unobserved time-varying factor that affects the department's research output.  Since many former graduate students become university faculty, it is possible that the past composition of graduate students affects faculty quality today.  Note that in our fixed-effects specifications, inference is based only on changes in publications over time within a department, so this argument raises a concern for our IV strategy only if departments hire their own former students, a rare practice. Even so, for each department we add a control for real (inflation-adjusted) non-equipment R&D expenditures, which include faculty salaries paid, in order to help purge any correlation between the past composition of graduate students and the error term in the second-stage publications regressions. As a group, our instruments are quite powerful.

In general, exchange-rate devaluations are associated with reductions in student enrollments, though this is more evidently the case for maximal changes in the region than for median changes. Increases in GDP per capita generally expand enrollments from non-OECD countries (where changes in ability-to-pay might be key), but reduce them from OECD countries (where the opportunity cost of a domestic labor market may dominate the decision). Positive oil shocks increase the supply of students from oil-producing nations in the Middle East and Africa.

Given all the interaction terms included in Table 2, it is difficult to see the direction of effects for each instrument, since those effects are heterogeneous across fields and universities, conditional on each field's and each university's historical dependence on students from a particular region. In Table 3, we construct some examples of the magnitude of impacts for particular university-field and region-of-origin combinations. The first row in this table indicates that, computed at the mean values of all variables, a one-percent increase in GDP per capita in China in the 1980s was predicted to increase the number of Chinese industrial engineering students at the University of Maryland, College Park by seven percent. This roughly translates into an increase of 0.7 Chinese students in this particular department for every $100 increase in

Chinese GDP per capita in the mid-1980s. Conversely, our first-stage estimates indicate that a one-percent increase in Western European GDP per capita in the 1980s was expected to reduce the enrollment of Western Europeans in the Physiology department at Columbia University by 2.8 percent.

*6B. Publications and Citations Regressions*

The sample period for analyzing the impact of doctoral students on scientific publications is 1973 to 1998, with enrollments lagged one year in order to reflect the lag from research to publication. The reason for ending the period at this date is that after 1997 our count of student enrollments falls off since many were still in graduate school and not counted in the SED database by 2004. All regressions control for a comprehensive set of fixed effects for years and university-field pairs, along with time trends specific to each university and each field.

Regressions presented in Tables 4a and 4b study, respectively, the determinants of the number of scientific publications and the number of citations to those publications in each university-field pair.[15] The attempt here is to see if contributions to higher-quality publications (those with more citations) are similar to contributions to publications *per se*. We examine the relative contributions of domestic and foreign students by defining the enrollment variables as the total number of enrollees (domestic+foreign) and the *share* of foreign students (foreign/domestic+foreign). The first column (1a and 1b) in each table reports OLS regressions without instrumentation. This specification finds that total enrollments contribute positively to creating publications, while the negative

[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). The publication citations were taken from the ISI Web of Science database with our *Perl* program.

coefficient on foreign share suggests the productivity of international students is lower. We translate these findings into marginal impacts and elasticities, computed at sample means, in the next two columns. These calculations find that both domestic and international doctoral students have positive productivities, but the marginal contribution of the former is somewhat larger. Each extra foreign student in the 'average' department raises publications by 0.38 percent (or 0.15 articles published per year), while each extra American raises publications by 0.48 percent (0.19 publications). These effects appear small (though not insignificant) relative to the average number of publications per department per year of 38. In the citations regressions, the coefficient on foreign share is not significant, suggesting that the productivities of both domestic and international students are the same.

Regressions in columns (2a and 2b) use foreign-country shocks to instrument for both total enrollments and foreign share in two-stage least squares estimation. Instrumenting enrollments more than doubles the coefficient on total enrollments and raises the estimated productivity of U.S. students to 1.13 percent, or 0.45 articles. The coefficient on foreign share switches signs but is not significant, suggesting that the causal marginal productivities of both domestic and foreign students are the same. However, in the citations case, 2SLS estimates find that the contribution of Americans is doubled, to 0.82 percent, or over 10 additional citations. Here, the marginal productivity of foreign doctoral students is even larger, at 1.27 percent, amounting to 16 citations.

In the last two columns of each table we add as controls each department's real equipment and physical plant expenditures and non-equipment real research expenditures, including salaries. This reduces slightly the coefficient on total enrollments in Table 4a, but it remains highly significant. The coefficients are reduced also in the citations equations in Table 4b and the impact of foreign share becomes insignificant. The 2SLS coefficient on equipment is negative in the publications equations but positive in the citations equations. In any case, these results verify that the estimated positive contributions of both American and foreign students to the production of publications and citations is robust to controlling for other departmental resources.

We provide the negative binomial regressions for the same specifications in Tables 4c and 4d. The results in the publications case without instrumental variables are similar to the linear regressions, though the estimated impacts are smaller. In contrast, using the IV approach finds a considerably higher productivity of U.S. students than of foreign students in publications, though both are significant. The impacts are the same in the citations regressions. Here, both forms of research expenditures positively affect publications and citations.

The instrumental variables results in the negative binomial specifications should be treated with caution because we simply insert the first-stage predicted enrollments into the second-stage negative binomial regressions, which is not a full 2SLS approach. We report them here primarily for comparison purposes, noting the extensive similarity between results in the linear and non-linear cases. From this point forward we do not report negative binomial regressions, which are available on request.

Tables 5a and 5b provide different cuts at the data. All regressions are run with both U.S. and foreign students instrumented with the source-country shocks and include the research controls. In the first two columns of each table we show the basic regressions when U.S. departments are split into “large” (those with above-median enrollments) and “small” (below-median enrollments). The results for publications are similar to those in Table 4a, though the coefficient magnitudes are larger. However, the negative coefficient on foreign share is insignificant for small universities, suggesting that international students make a somewhat larger contribution to publishing in those departments. This result holds also in the citations regressions in Table 5b. It is of interest to note that the contribution of research funds is considerably higher in the smaller departments.

The next pair of columns in each table breaks the department samples into “early” (1973-86) and “late” (1986-97). In Table 5a we find that the negative coefficient on international share becomes smaller in magnitude in the later period, suggesting that the contribution of foreign students to publishing became higher over time. This does not seem to be the case for publication citations, however, where the foreign-student share remained insignificant throughout the period.

The final pair of regressions in each table breaks the sample into departments at “elite” research universities (those for which the 25$^{th}$ percentile undergraduate SAT score was greater than 1250 or ACT score greater than 25) and other research universities. Here an interesting difference emerges. Specifically, U.S. students contribute approximately the same to publications in both university types, with the marginal impacts being 0.38 (elite) and 0.32 (non-elite). In contrast, the marginal impact of international students in elite universities (0.52 publications) is considerably larger than that in lower-quality institutions (0.32 publications). Turning to citations in Table 5b, we find that both U.S. and international students make a substantial marginal contribution to work that is cited, with the domestic impact being 10.8 citations (elite) and 5.2 citations (non-elite) and the foreign impact being 18.2 citations (elite) and 5.2 citations (non-elite). Here the contribution of international students is notably larger in elite research universities.

In Tables 6a and 6b we consider the impacts of fluctuations in student enrollments when they are induced by different types of shocks. In particular, we define two sets of instruments. The first includes those instrumental variables that should have neutral effects across students of varying income and wealth patterns, because they should uniformly affect the ability of all students to go abroad. These are changes in the number of students studying outside the United States and the policy restrictions. The second set includes those instrumental variables that could affect the ability of students of varying means to study abroad. These are shocks in GDP per capita, the OECD dummy interacted with GDP per capita, the oil share of GDP and exchange rates. Thus, we run the 2SLS estimation separately for each set of instruments.

The idea we are trying to capture is that "pay neutral" shocks should have little effect on the quality distribution of graduate students, while positive "ability to pay" shocks could lower the average quality of students. The reason is that the latter shocks would be more likely to increase the opportunities for relatively higher-income students from poor countries to apply, while it would not much increase applications from lower-income applicants, who would need financial aid. Because departments are likely to offer financial aid only to applicants of higher quality, such shocks would shift enrollments toward lower-quality applicants on average.

In Table 6a the first column shows that pay-neutral shocks result in a significantly positive coefficient on the foreign student share. Translating this into impacts, we find that the marginal foreign contribution is 1.87 percent, or 0.74 publications. However, in regression (12a) we find that ability-to-pay shocks reduce this impact to 0.82 percent, or 0.33 publications. In short, a reduction in the average quality of international applicants does reduce the ability of departments to engage in publishable research. Interestingly, it also slightly reduces the productivity of U.S. students, suggesting that there may be a small degree of complementarity between domestic and international candidates as the quality of the latter changes. These results carry over to regressions (13a) and (14a), where we include the R&D controls. Table 6b finds analogous results for publication citations. A shock that reduces average international student quality substantially reduces the marginal contribution of such students in comparison with a pay-neutral shock. In this case, however, the marginal productivity of domestic candidates rises.

These results suggest that the type of source-country shock that sends students to the United States can matter a great deal, since it can change the quality distribution of applicants and enrollees. The research impact of a visa restriction on foreigners will crucially 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 quality or for ‘ability to pay’).

In Tables 7a and 7b we account for the fact that students come from undergraduate institutions of varying quality. Using information on the undergraduate university of each Ph.D. recipient in the *SED* micro data, we classify each as a high-quality institution or other institution. American colleges and universities were allocated this status on the basis of their selectivity, using the same criteria for determining elite and non-elite research universities. International universities were considered to be of high quality if they were listed in the top 200 in the global rankings by Shanghai Jiao Tong University in China or were listed by that group as one of the top two institutions in their country. We are more confident in our rankings of U.S. institutions, in part because the Chinese ranking heavily weights extreme scientific research accomplishments over teaching effectiveness.

In Table 7a we find what might seem a counterintuitive result: students from U.S. lower-ranked BA institutions make positive marginal contributions to publications, while those from higher-ranked colleges have negative productivities. One potential explanation for this result is that top-quality students from higher-ranked institutions have more opportunities to go into more lucrative professions (that is, they have a higher outside opportunity) than those from lower-ranked colleges. In this context, the top students admitted from the lower-quality universities may have stronger research capabilities than more average students from higher-quality universities. In contrast, foreign students coming from top research universities have a much higher marginal productivity in publications at the average U.S. department, with an elasticity of 1.23, while the elasticity of publications with respect to students from other institutions is 0.27.

Additional perspective is offered in the final pair of columns, where we investigate these effects on research performed in departments at elite versus non-elite U.S. universities. Here we find that the marginal contribution of domestic students from highly ranked colleges remains negative at elite universities but is strongly positive at other institutions. However, the strongly positive productivity of U.S. students from lower-ranked colleges exists solely at elite research universities. Foreign students from higher-ranked BA institutions have stronger productivities in publishing at both types of

U.S. universities that do those from lower-ranked institutions. Overall, put in these terms, international students from excellent research universities offer the strongest research contributions in American departments.

The analysis of publication citations in Table 7b tells a similar story. Here we find that students from higher-quality U.S. undergraduate institutions tend to diminish citations to departmental research, though the effect holds only for elite universities. Foreign students raise citations of research done in elite universities, and the marginal impact is higher for students from higher-ranked BA institutions. Thus, it appears that students from lower-ranked U.S. colleges are more likely to contribute to publications and citations at departments in elite universities but those from higher-ranked institutions are more productive at non-elite locations. International students contribute strongly to both publications and citations at elite American universities, with a stronger impact characterizing students from higher-ranked BA institutions. In this regard, such students seem to be sorted effectively across departments.

Tables 8a and 8b repeat the basic analysis by using both the global and U.S. regional shocks as instruments for total enrollments and the foreign share. Here we find that both domestic and international students have substantial impacts on publishing journal articles. The marginal contribution of a U.S. student is 0.43 publications, while that of international students is 0.51 publications. In Table 8b, however, we find a considerably larger impact of foreign students on generating citations. These results are sustained when we incorporate the research controls.

To this stage the analysis generally has found that the marginal research contributions of Americans are smaller than that of foreign students, though this is not the case in all specifications. The general result is consistent with the assumption that foreigners are more expensive for departments to admit. Table 9 offers additional perspective on this based on specifications where we break down foreign students into seven regions of origin, using negative binomial regressions. Comparing the un-instrumented (regression 23) to the instrumented (regression 24) specification indicates that instrumenting generally corrects for a negative bias. The striking finding is the large variation in the marginal contributions of students from various regions. Every additional American student leads to an increase of only 0.11 publications, compared to the 0.8 publication increase from each additional East Asian and a 2.97 increase from an additional South Asian.[16] Similar stories emerge in the citation count regressions in the next 2 columns (labeled 25 and 26), where instrumenting increases the estimated productivities, and there is high variability across regions.

## 7. Concluding Remarks

In this paper we report our initial findings regarding the contributions of domestic and foreign graduate students in science and engineering programs to knowledge creation (articles published) in the United States. To identify these impacts we undertook a first-stage instrumental variables approach to explaining shifts in foreign enrollments by

[16] At the other extreme, Eastern Europeans have a negative impact on departmental output. Our model explains how marginal productivities can be negative. If a shock in the U.S. make more domestic students apply (e.g., a decrease in wages of college graduates) there would be an increase in the quality of the applicant pool, raising publications (*K* in our model). Since Americans now present a better quality-cost tradeoff, the effect would be to squeeze some students from other regions out of those admitted. In turn, the higher publications would be associated with a lower number of enrollments from some foreign regions and OLS would find a negative coefficient on the latter. Even in the IV estimates, if shocks in two different regions are correlated (e.g. U.S. and Canada), a positive shock in both U.S. and Canada may drive *K* up while reducing Canadian enrollments, thereby leading to a negative coefficient.

region. The instruments plausibly are uncorrelated with domestic U.S. factors that influence university publishing.

Interest arises in this question for a number of reasons. If foreign students are strong contributors to knowledge creation in universities, U.S. visa restrictions could have significantly negative impacts on research. Perhaps more significantly, many other countries and international universities are actively recruiting high-quality doctoral students to study outside the United States. Such environmental changes may again be considered shocks to American universities that could negatively affect prospects for research.

Our findings are fairly clear at the most basic levels. We find that instrumentation makes a significant difference in the estimated impacts of international graduate students. Both foreign and domestic students are positive inputs in departmental research production, even after controlling for the fixed effects, trends, and departmental resources, and exogenous shock-induced changes in foreign-student enrollments generally have a greater positive impact on departmental output. Further, the nature of the shock matters. Enrollments induced by shocks that send more non-scholarship students to the United States have a smaller positive impact. Overall, the marginal foreign student is neither clearly better nor clearly worse than the marginal American. Foreign students contribute relatively more in terms of citations and at elite universities. Finally, there are significant variations in the marginal productivity of students across source regions, a fact that is consistent with a model of search behavior between universities and graduate students of differential quality levels.

The magnitude of the foreign student contribution relative to Americans is important to discuss since the impacts of visa restriction policies will depend on whether it would be easy to substitute the foreigner who is denied entry with an American. The relative contribution of foreigners and Americans appear to depend on the type of foreign student (from lower-quality institution versus higher-quality) and on the type of shock in source countries that sends students here. Foreign enrollments induced by shocks that are more likely to send non-scholarship students to the U.S. have a smaller impact of university productivity. Worries about aggregate visa restrictions or growth in opportunities abroad that reduce the overall supply of foreign students may be warranted only in certain contexts. The statements made in editorials by journalists and university administrators about declining enrollments leading to a "crisis in research and innovation" should, at best, be conditional statements.

## References


Anderson, Stuart (2006). "The Debate over Quotas on Highly Skilled Immigrants," *International Educator*, November-December.

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

Bhagwati, Jagdish and Koichi Hamada (1974). "The Brain Drain, International Integration of Markets for Professionals and Unemployment," *Journal of Development Economics* 1(1): 19-42.

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

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).

Breschi, S., F. Lissoni and F. Montobbio (2005). "The Scientific Productivity of Academic Inventors," CESPRI Working Paper, Bocconi University.

Chellaraj, G., Maskus, K., and Mattoo, A. (2007). "The Contribution of Skilled Immigration and International Graduate Students to U.S. Innovation." *Review of International Economics*, forthcoming.

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

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.

Docquier, Frederick and Abdesalam Marfouk (2006). "International Migration by Education Attainment," in Caglar Ozden and Maurice Schiff (eds). *International Migration, Remittances & the Brain Drain*, Washington: World Bank and Palgrave.

Economic Times of India (2005). "Visa Curbs to Topple U.S. Technology Edge," *Economic Times of India*, 14 March.

Education Testing Service (2004). *Education Testing Services Media Release*

Ehrenberg, Ronald G. (2005). "Graduate Education, Innovation, and Federal Responsibility," Council of Graduate Schools, *Communicator*, 38(6).

Financial Times (2004). "Opening the Door," *Financial Times*, October 15.

Gale, D. and L. S. Shapley (1962), “College Admissions and the Stability of Marriage,” American Mathematical Monthly 69: 9-14.

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

Hausman, Jerry, Bronwyn Hall and Zvi Griliches (1984). “Econometric Models of Count Data with an Application to the Patents-R&D Relationship,” Econometrica 52(4): 909-938.

Irving, R. W. (1998), “Matching Medical Students to Pairs of Hospitals: A New Variation on a Well-Known Theme,“ Proceedings of ESA ’98 (Sixth Annual European Symposium on Algorithms, published in Lecture Notes in Computer Science 1461: 381-392.

Jaffe, A.B. and Trachtenberg, M. (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.

Lach, S. and Schankerman, M. (2003). “Incentives and Invention in Universities.” Working Paper No. 9727, NBER, May.

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).

Mullahy, John (1997). “Instrumental Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior,” *The Review of Economics and Statistics* : 586-593.

National Academies (2005). *Policy Implications of International Graduate Students and Postdoctoral Scholars in the United States*. Washington: The National Academies Press.

New York Times (2004). “U.S. Slips in Attracting the World’s Best Students,” *New York Times,* December 21.

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

Rosenzweig, Mark (2006). “Global Wage Differences and International Student Flows,” Brookings Trade Forum *Global Labor Markets?* May 11-12, 2006

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

Stephan, P., Black, A. and Gurmu, S. (2006). "The Role of Foreign-Born Graduate Students in University Patenting." manuscript, Georgia State University.

TIMMS (2003). Trends in International Mathematics and Science Study, http://timss.bc.edu/timss2003.html.

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.

## Data Appendix

This initial table lists our fields of science and engineering (used in the university-field pairs) and the patent classification. More detail on patent allocation is provided in the next section.

Data Appendix Table One

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 |

### *1. Independent 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 such.

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 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 should be 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.

.

### *2. 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). We also matched dissertation specialties to our patent subcategories, which was more difficult due to a lack of congruence between categories of products and scientific disciplines. 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,[17] 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) High-Quality Graduate Student Enrollment Counts**

To create an indicator of high-quality student enrollment, we used information on the institutions where each student received his bachelor's degree. If that college or university was regarded as high-quality, the student was similarly considered to be of high-quality. We considered U.S. institutions, both research universities and other colleges and universities, as high-quality if the 25th percentile of undergraduate SAT (ACT) scores was at least 1220 (26), as reported by *US News and World Report*, 2005. We considered foreign institutions to be high-quality if they were either (a) listed in the world top 100 universities by the Institute of Higher Education, Shanghai Jiao Tong University, 2005 or (b) were one of the top two universities of each country listed in the world top 100-500 (by the same group). We then followed a process similar to

[17] 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.

subsection (a) above to create a count of high-quality students per university/field/year observation.

**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**

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 was 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) Oil Share of GDP**

The oil share of GDP is the ratio of real oil revenues to real GDP. Real oil revenues were calculated as production quantity multiplied by the real oil price. Production data for crude oil were taken from the U.S. Department of Energy, Energy Information Agency publication *International Energy Annual*. Oil price data came from OPEC's *Annual Statistical Bulletin*, deflated with CPI data from *International Financial Statistics*. At the regional level, mean oil share is used.

**d) Total International Students to Non-US 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/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.

**e) 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.

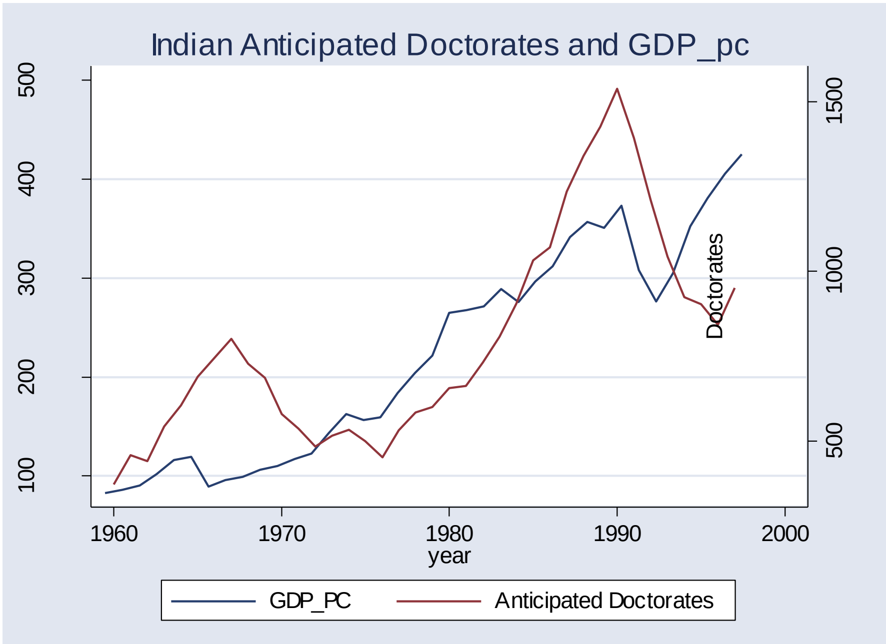
**f) State Control Policy**

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.

**g) U.S. Instruments**

Three variables were compiled at the level of U.S. Census Division (9 divisions) to attempt to account for domestic enrollment with regional economic fluctuations. This included gross state product (per capita), unemployment rates and population ages 20-29. All series were collected at the state level, then averaged or summed to the Census Division level. Data on gross state product were obtained from the U.S. Bureau of Economic Analysis, on unemployment from the Bureau of Labor Statistics and on population from the Census Bureau.

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




*Note: 'Anticipated 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: Time Series of Nigerian Doctoral Student Enrollment and Oil Revenues

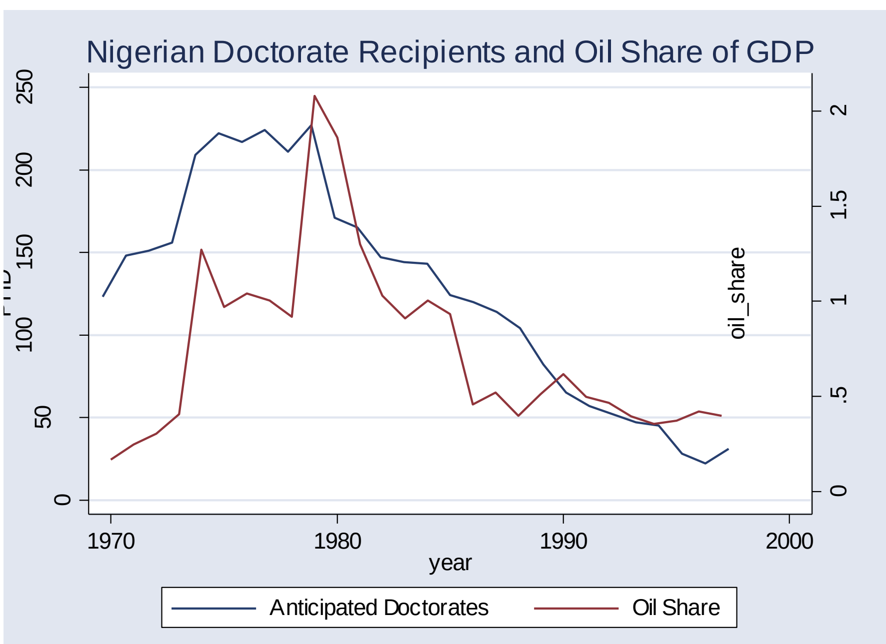
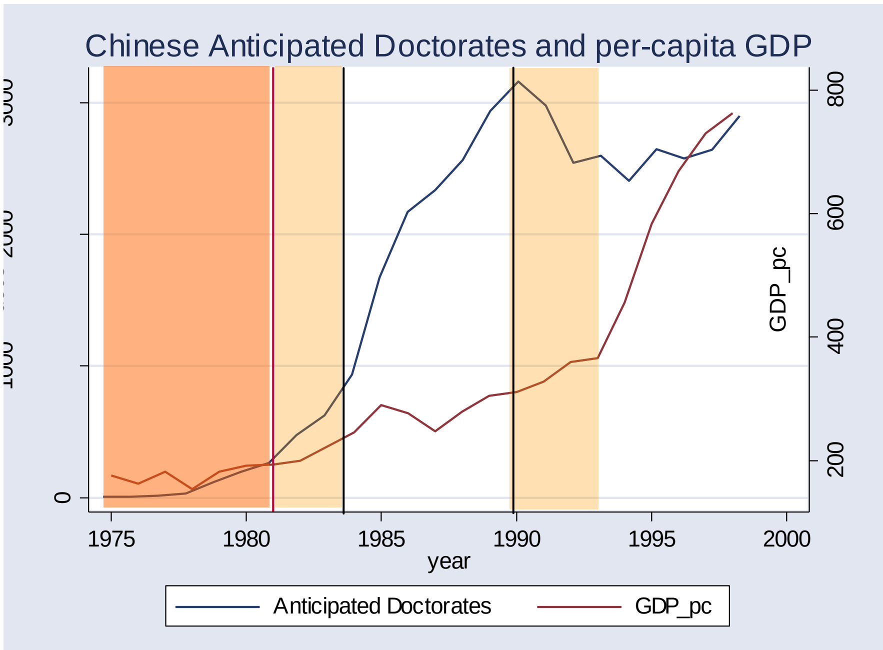


Figure 3: Policy Changes on Foreign Study Abroad in China and Doctoral Student Enrollment




*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 4: Differential Response of Chinese Enrollment across Universities

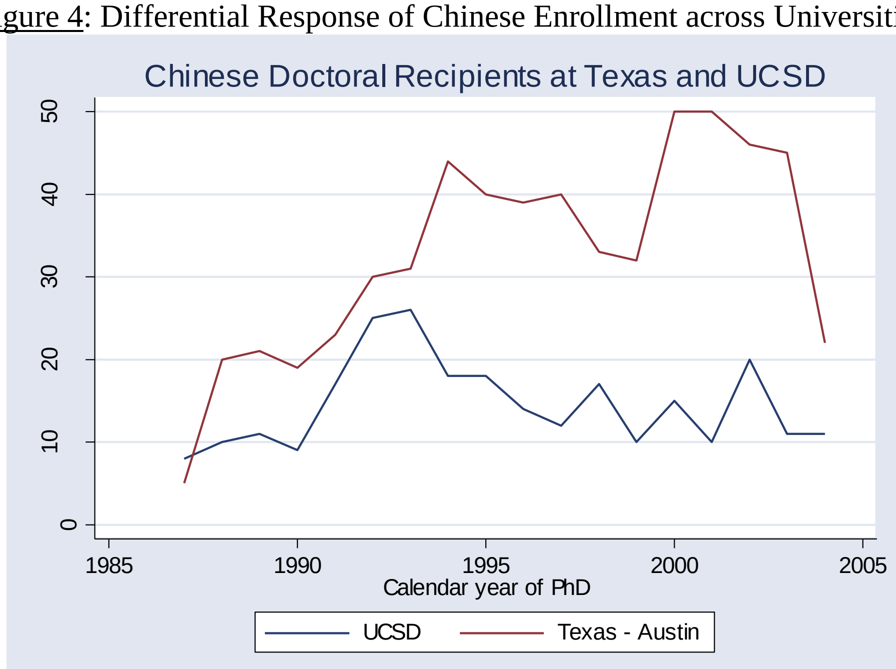


Figure 5: Differential Response of Chinese Enrollment across Disciplines

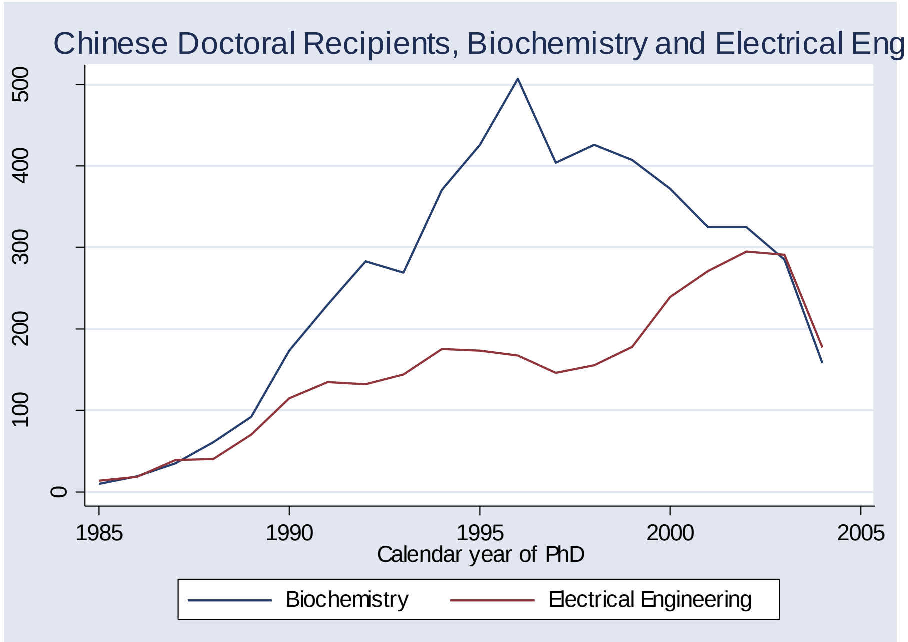


**Table 1. Summary Statistics for Main Publications Dataset**

| Variable | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|
| Publication Counts | 39.84 | 59.84 | 0 | 1159 |
| Pub. Counts - 1970's | 24.87 | 37.71 | 0 | 526 |
| Pub. Counts - 1980's | 38.99 | 56.51 | 0 | 807 |
| Pub. Counts - 1990's | 54.00 | 74.55 | 0 | 1159 |
| Elite | 54.08 | 71.44 | 0 | 1108 |
| Non-Elite | 39.75 | 58.18 | 0 | 1159 |
| Publication Citation Counts | 1271.65 | 2797.24 | 0 | 71051 |
| Pub. Citations - 1970's | 832.08 | 1814.92 | 0 | 37860 |
| Pub. Citations - 1980's | 1273.47 | 2630.17 | 0 | 53257 |
| Pub. Citations - 1990's | 1654.01 | 3550.75 | 0 | 71051 |
| Elite | 2104.15 | 4075.86 | 0 | 71051 |
| Non-Elite | 1096.69 | 2185.71 | 0 | 48900 |
| Enrollment - Total | 42.01 | 53.58 | 0 | 498 |
| Enrollment - US | 27.02 | 37.20 | 0 | 380 |
| High-Rank BA Institution | 7.56 | 14.68 | 0 | 207 |
| Low-Rank BA Institution | 19.46 | 26.25 | 0 | 312 |
| Enrollment – Foreign | 14.99 | 22.25 | 0 | 317 |
| High-Rank BA Institution | 1.21 | 3.26 | 0 | 98 |
| Low-Rank BA Institution | 13.77 | 20.12 | 0 | 281 |
| Equipment and Physical Plant Expenditures ($ million) | 0.74 | 1.28 | -0.005 | 18.23 |
| Other R&D Expenditures | 3.52 | 5.95 | 0.00089 | 103.53 |
| N=57475 | Number of university-field groups: 2299 | | | |

## Table 2. First-Stage Instrumental Variables Regressions Predicting Enrollment Counts by Region of Origin

| Region | China | E. Asia | S. Asia | ME & Africa | E. Europe | W. Europe | W. Hem |
|---|---|---|---|---|---|---|---|
| | 0.0278 | 0.0002 | 0.0085 | 0.0012 | -0.0001 | -0.0002 | 0.0000 |
| GDP per capita | (16.82)*** | (1.17) | (9.02)*** | (6.87)*** | (6.64)*** | (3.88)*** | (0.07) |
| Interacted with University's dependence | -0.0083 | -0.0045 | -0.0461 | -0.0120 | -0.0005 | -0.0010 | 0.0041 |
| on region | (0.72) | (12.98)*** | (6.39)*** | (11.70)*** | (0.43) | (6.34)*** | (3.12)*** |
| Interacted with Field's dependence | -0.0829 | -0.0017 | -0.0207 | -0.0041 | -0.0020 | 0.0000 | 0.0018 |
| on region | (10.42)*** | (6.94)*** | (4.89)*** | (5.75)*** | (2.21)** | (0.12) | (1.72)* |
| Triple interaction: both University's | -0.3130 | 0.0158 | 0.3831 | 0.0300 | 0.5990 | 0.0116 | -0.1148 |
| & Field's dependence on region | (2.96)*** | (8.04)*** | (4.18)*** | (1.86)* | (3.45)*** | (3.35)*** | (3.94)*** |
| | | -0.0001 | | -0.0001 | -0.0000 | 0.0000 | 0.0001 |
| GDP per capita * OECD nation | | (0.84) | | (0.93) | (0.59) | (3.58)*** | (0.88) |
| Interacted with University's dependence | | 0.0028 | | 0.0036 | -0.0262 | 0.0004 | 0.0003 |
| on region | | (2.46)** | | (1.10) | (5.82)*** | (2.10)** | (0.16) |
| Interacted with Field's dependence | | 0.0019 | | 0.0000 | 0.0020 | 0.0001 | 0.0012 |
| on region | | (1.91)* | | (0.01) | (1.13) | (0.34) | (0.63) |
| Triple interaction: both University's | | -0.0492 | | 0.0567 | 4.9078 | -0.0077 | -0.0582 |
| & Field's dependence on region | | (4.31)*** | | (1.28) | (12.79)*** | (1.62) | (1.11) |
| Percentage change in SDR exchange rate | | | | | -0.0415 | | -0.0414 |
| (Region Median) | | | | | (1.86)* | | (0.33) |
| Interacted with University's dependence | 81.4627 | -5.0376 | 6.5642 | 16.9729 | -2.8568 | 18.3986 | 2.6629 |
| on region | (5.73)*** | (0.18) | (0.50) | (1.87)* | (0.82) | (1.68)* | (0.69) |
| Interacted with Field's dependence | 53.2826 | 22.4271 | 14.4241 | 23.0192 | 1.2057 | 10.6179 | 0.7015 |
| on region | (5.62)*** | (1.34) | (1.92)* | (3.75)*** | (0.60) | (1.31) | (0.24) |
| Triple interaction: both University's | -780.0054 | 625.9146 | 350.4066 | 757.2024 | -744.0548 | -1,083.6089 | 55.7838 |
| & Field's dependence on region | (6.08)*** | (3.00)*** | (1.70)* | (5.41)*** | (2.17)** | (4.01)*** | (0.60) |
| Foreign students from region studying at | 0.0000 | -0.0000 | 0.0000 | -0.0000 | -0.0000 | 0.0000 | 0.0000 |
| Non-US hosts | (3.19)*** | (3.07)*** | (3.21)*** | (3.39)*** | (1.49) | (3.93)*** | (2.86)*** |
| Interacted with University's dependence | -0.0004 | 0.0000 | -0.0007 | 0.0001 | -0.0001 | -0.0001 | -0.0006 |
| on region | (2.25)** | (0.73) | (2.97)*** | (3.75)*** | (2.17)** | (3.81)*** | (3.44)*** |
| Interacted with Field's dependence | 0.0000 | 0.0000 | -0.0004 | 0.0000 | -0.0001 | -0.0000 | -0.0003 |
| on region | (0.35) | (0.17) | (3.08)*** | (2.79)*** | (3.91)*** | (2.82)*** | (1.87)* |
| Triple interaction: both University's | 0.0061 | 0.0015 | 0.0279 | 0.0015 | 0.0019 | 0.0023 | 0.0187 |
| & Field's dependence on region | (3.42)*** | (4.87)*** | (8.17)*** | (5.17)*** | (0.31) | (5.05)*** | (4.33)*** |
| Policy restrictions on study | 0.2616 | | | | 0.8704 | -0.8062 | 3.6078 |
| abroad | (4.89)*** | | | | (44.14)*** | (5.30)*** | (1.55) |
| Interacted with University's dependence | -0.0512 | | | | 60.5353 | 21.5546 | 3.0403 |
| on region | (0.02) | | | | (16.21)*** | (4.47)*** | (0.04) |
| Interacted with Field's dependence | 15.1254 | | | | 27.2640 | 15.8089 | -130.5287 |
| on region | (9.13)*** | | | | (13.00)*** | (4.14)*** | (2.15)** |
| Triple interaction: both University's | 200.6083 | | | | 3,417.5069 | -456.9912 | 1,059.6823 |
| & Field's dependence on region | (6.39)*** | | | | (9.39)*** | (3.87)*** | (0.68) |
| Percentage change in SDR exchange rate | | | -0.1979 | 0.0214 | -0.0076 | -0.0891 | -0.0024 |
| (Region Maximum) | | | (1.20) | (3.94)*** | (3.43)*** | (4.00)*** | (1.67)* |
| Interacted with University's dependence | | -32.7045 | 1.7015 | -0.4159 | 0.5261 | 1.8033 | 0.0866 |
| on region | | (4.34)*** | (0.39) | (3.51)*** | (0.87) | (2.37)** | (1.96)* |
| Interacted with Field's dependence | | -16.4295 | -0.9475 | -0.3903 | 0.9535 | 1.7816 | 0.0351 |
| on region | | (3.72)*** | (0.40) | (4.65)*** | (2.67)*** | (3.22)*** | (1.02) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Triple interaction: both University's | | 476.7333 | -2.3553 | 2.8981 | -156.6729 | -30.8467 | -1.5905 |
| & Field's dependence on region | | (8.31)*** | (0.04) | (1.56) | (1.45) | (1.61) | (1.53) |
| Oil Share of GDP | | | | | | | |
| Interacted with University's dependence | | | | 24.5665 | -192.9500 | | -125.8367 |
| on region | | | | (7.60)*** | (4.44)*** | | (3.77)*** |
| Interacted with Field's dependence | | | | 4.7928 | -65.7356 | | -66.8567 |
| on region | | | | (2.53)** | (2.47)** | | (2.52)** |
| Triple interaction: both University's | | | | 219.3639 | 60,883.5863 | | 3,572.3831 |
| & Field's dependence on region | | | | (3.93)*** | (9.63)*** | | (4.41)*** |
| | | 2.1228 | | 1.3348 | | 1.1328 | 0.1166 |
| OECD country indicator | | (3.67)*** | | (6.92)*** | | (12.05)*** | (0.48) |
| Interacted with University's dependence | | -27.8052 | | -8.5849 | | -13.7152 | 17.2458 |
| on region | | (3.65)*** | | (1.51) | | (4.51)*** | (2.47)** |
| Interacted with Field's dependence | | -25.1727 | | -7.5112 | | -7.7804 | 10.1142 |
| on region | | (3.84)*** | | (2.48)** | | (3.25)*** | (1.68)* |
| Triple interaction: both University's | | 321.0165 | | 75.4939 | | 620.9804 | -308.6823 |
| & Field's dependence on region | | (4.11)*** | | (0.95) | | (8.34)*** | (1.78)* |
| Constant | 229.4667 | -475.4711 | -7.5219 | -148.9375 | -31.3172 | -156.0519 | -57.9441 |
| | (7.20)*** | (5.01)*** | (0.54) | (11.51)*** | (15.67)*** | (5.55)*** | (2.74)*** |
| Number of observations | 59800 | 59800 | 59800 | 59800 | 59800 | 59800 | 59800 |
| Number of university-field fixed effects | 2300 | 2300 | 2300 | 2300 | 2300 | 2300 | 2300 |
| R-squared | 0.62 | 0.33 | 0.25 | 0.19 | 0.50 | 0.30 | 0.16 |
| F-Stat on excluded instruments | 272.64 | 74.92 | 56.82 | 122.25 | 810.11 | 312.42 | 43.22 |
| P-value | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Absolute value of t statistics in parentheses; *significant at 10%; **significant at 5*; ***significant at 1%

All regressions control for university-field pair fixed effects, year dummies, and field- and university-specific time effects

**Table 3. Examples of the Effects of Instrumental Variables on Enrollment Counts**

| University, Field and Decade | Region of Origin | Instrument | Marginal Effect | Elasticity |
|---|---|---|---|---|
| U. of Maryland, Industrial Engineering, 1980s | China | GDP per capita | 0.027 | 7.01 |
| U. of Virginia, Statistics, 1990s | E & SE Asia | Percentage change in exchange rate, max in region | -1.596 | -0.57 |
| Texas A&M U., Electrical Engineering, 1980s | South Asia | GDP per capita | 0.0064 | 0.162 |
| U. of Oklahoma, Molecular and Cellular Biology, 1980s | Mid East & Africa | Oil Share of GDP | 3.82 | 4.22 |
| Columbia U., Physiology, 1990s | Western Europe | GDP per capita | -0.00023 | -2.8 |
| Cornell U., Ecology & Evolution, 1990s | Western Hem. | Students enrolled in other host countries | 0.000053 | 0.055 |

Notes: Effects vary by university-field pairs and by decade due to the inclusion of the interaction terms in the set of first-stage regressors. The marginal effects and elasticities are computed on the basis of the model in Table 3, at the means of all variables for the region. These are only intended to represent the typical effect of the strongest instrument by region.

**Table 4a. Linear Fixed Effects Regressions**

Dependent Variable: Publications by Department

| | OLS | Publication effect at sample means of adding one more | | 2SLS | Publication effect at sample means of adding one more | | 2SLS | 2SLS |
|---|---|---|---|---|---|---|---|---|
| | (1a) | American | Foreigner | (2a) | American | Foreigner | (3a) | (4a) |
| Total Students | 0.1778*** | 0.482% | 0.382% | 0.4496*** | 1.129% | 1.129% | 0.4594*** | 0.4250*** |
| | (0.0055) | 0.19 pubs | 0.15 pubs | (0.0208) | 0.45 pubs | 0.45 pubs | (0.0224) | (0.0230) |
| Foreign Student Share | -1.7201*** | ε = 0.13 | ε = 0.057 | 2.4330 | ε = 0.305 | ε = 0.169 | 2.3017 | 1.5258 |
| | (0.4897) | | | (1.6834) | | | (1.8374) | (1.8704) |
| Real Equipment & Phys. Plant Exp. ($ millions) (5 yr MA) | | | | | | | -0.2063** | -0.2615** |
| | | | | | | | (0.1045) | (0.1056) |
| Real R&D Exp. ($ millions) (5 yr MA) | | | | | | | | 0.1680*** |
| | | | | | | | | (0.0456) |
| Observations | | 57500 | | | 57500 | | 49025 | 47959 |
| No. of Field-University Pair FE | | 2300 | | | 2300 | | 2240 | 2214 |

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

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

In two-stage least squares regressions, total students and foreign share are instrumented with foreign shocks.

**Table 4b. Linear Fixed Effects Regressions**

Dependent Variable: Citations to a Department's Publications

| | OLS (1b) | | Citation effect at sample ans of adding one more American | Foreigner | 2SLS (2b) | | Citation effect at sample neans of adding one more American | Foreigner | 2SLS (3b) | 2SLS (4b) |
|---|---|---|---|---|---|---|---|---|---|---|
| Total Students | 5.0620*** | | 0.398% | 0.398% | 12.4440*** | | 0.815% | 1.274% | 12.3225*** | 10.8377*** |
| | (0.3147) | | 5.06 cites | 5.06 cites | (1.1679) | | 10.36 cites | 16.19 cites | (1.2526) | (1.2812) |
| Foreign Student Share | -43.9624 | | ε = 0.108 | ε = 0.06 | 250.7964*** | | ε = 0.219 | ε = 0.192 | 181.2389* | 65.2541 |
| | (28.0114) | | | | (94.7386) | | | | (102.6393) | (104.0897) |
| Real Equipment & Phys. Plant Exp. ($ millions) (5 yr MA) | | | | | | | | | - 18.0665*** | - 18.2146*** |
| | | | | | | | | | (5.8376) | (5.8781) |
| Real R&D Exp. ($ millions) (5 yr MA) | | | | | | | | | | -3.6704 |
| | | | | | | | | | | (2.5386) |
| Observations | 57500 | 57500 | 57500 | 57500 | 57500 | 57500 | 57500 | 57500 | 49025 | 47959 |
| No. of Field-University Pair FE | 2300 | 2300 | 2300 | 2300 | 2300 | 2300 | 2300 | 2300 | 2240 | 2214 |

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

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

In two-stage least squares regressions, total students and foreign share are instrumented with foreign shocks.

Citation Counts = $\sum_i (1 + \# \; papers\_citing\_paper_i)$

**Table 4c. Negative Binomial Fixed Effects Regressions**

Dependent Variable: Publications by Department

| | No IV (1c) | Publication effect at sample means of adding one more American | Foreigner | IV (2c) | Publication effect at sample means of adding one more American | Foreigner | IV (3c) | IV (4c) |
|---|---|---|---|---|---|---|---|---|
| Total Students | 0.0023*** | +0.27% | +0.21% | 0.0088*** | +1.29% | +0.17% | 0.0088*** | 0.0086*** |
| | (0.0001) | 0.11 pubs | 0.09 pubs | (0.0003) | 0.51 pubs | 0.07 pubs | (0.0003) | (0.0003) |
| Foreign Student Share | -0.0344*** | ε=0.08 | ε=0.03 | -0.5079*** | ε= 0.39 | ε =0.03 | -0.4899*** | -0.5057*** |
| | (0.0102) | | | (0.0274) | | | (0.0353) | (0.0283) |
| Real Equipment & Phys. Plant Exp. ($ millions) (5 yr MA) | | | | | | | 0.0118*** | 0.0113*** |
| | | | | | | | (0.0013) | (0.0013) |
| Real R&D Exp. ($ millions) (5 yr MA) | | | | | | | | 0.0049*** |
| | | | | | | | | (0.0005) |
| Observations | 57475 | | | 57475 | | | 49011 | 47942 |
| No. of Field-University Pair FE | 2299 | | | 2299 | | | 2234 | 2205 |

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

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

All regressions estimated using negative binomial count data models

In IV regressions, total students and foreign share are instrumented with foreign shocks

**Table 4d. Negative Binomial Fixed Effects Regressions**

Dependent Variable: Citations to a Department's Publications

| | No IV | Citation effect at sample means of adding one more | | IV | Citation effect at sample means of adding one more | | IV | IV |
|---|---|---|---|---|---|---|---|---|
| | (1d) | American | Foreigner | (2d) | American | Foreigner | (3d) | (4d) |
| Total Students | 0.0030*** | +0.20% | +0.48% | 0.0118*** | +1.18% | +1.18% | 0.0106*** | 0.0104*** |
| | (0.0001) | 2.54 cites | 6.11 cites | (0.0005) | 15 cites | 15 cites | (0.0005) | (0.0005) |
| Foreign Student Share | 0.1208*** | ε=0.05 | ε=0.07 | 0.0157 | ε= 0.31 | E=0.18 | 0.0040 | -0.0462 |
| | (0.0149) | | | (0.0448) | | | (0.0468) | (0.0470) |
| Real Equipment & Phys. Plant Exp. ($ millions) (5 yr MA) | | | | | | | 0.0175*** | 0.0149*** |
| | | | | | | | (0.0021) | (0.0021) |
| Real R&D Exp. ($ millions) (5 yr MA) | | | | | | | | 0.0082*** |
| | | | | | | | | (0.0008) |
| Observations | 57475 | | | 57475 | | | 49011 | 47942 |
| No. of Field-University Pair FE | 2299 | | | 2299 | | | 2234 | 2205 |

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

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

All regressions estimated using negative binomial count data models

In IV regressions, total students and foreign share are instrumented with foreign shocks.

Citation Counts = $\sum_i (1 + \#\, papers\_citing\_paper_i)$

**Table 5a. Impacts on Publishing by Size, Period and Elite Status**
Dependent Variable: Publications by Department

| | Large (5a) | Small (6a) | Early (7a) | Late (8a) | Elite (9a) | Marginal Effects: American | Marginal Effects: Foreign | Non-Elite (10a) | Marginal Effects: American | Marginal Effects: Foreign |
|---|---|---|---|---|---|---|---|---|---|---|
| Total Students | 0.4147*** | 0.4704*** | 0.3622*** | 0.3284*** | 0.4252*** | 0.698% | 0.966% | 0.3183*** | 0.801% | 0.801% |
| | (0.0289) | (0.0281) | (0.0304) | (0.0516) | (0.0297) | 0.38 pubs | 0.52 pubs | (0.0287) | 0.32 pubs | 0.32 pubs |
| Share Foreign | -8.4956** | -0.4221 | -6.7394*** | -5.7868** | 8.2177** | ε = 0.261 | ε = 0.178 | -2.4252 | ε = 0.267 | ε = 0.176 |
| | (3.4707) | (1.3325) | (2.3294) | (2.4489) | (3.4839) | | | (2.0060) | | |
| Real equip & phys. plant exp, 5 yr MA ($ millions) | 0.1355 | -0.7751*** | -0.1435 | 0.6324*** | -0.0003 | | | -0.4892*** | | |
| | (0.1467) | (0.1389) | (0.1129) | (0.1903) | (0.1566) | | | (0.1454) | | |
| Real R&D expenditures, 5 yr MA | 0.0831 | 0.6348*** | 0.0814 | 0.4508*** | 0.2955*** | | | 0.2057*** | | |
| | (0.0571) | (0.0716) | (0.0573) | (0.0913) | (0.0590) | | | (0.0679) | | |
| Observations | 25515 | 22444 | 25589 | 22370 | 14464 | | | 33495 | | |
| Number of group(univ field) | 1148 | 1066 | 2157 | 2137 | 648 | | | 1566 | | |

Standard errors in parentheses, * significant at 10%; ** significant at 5%; *** significant at 1%.

All Regressions use Foreign Instruments in two-stage least squares regressions, university-field pair and year fixed effects, and university and field specific trends.

Large defined as above-median total PhD enrollment.

Early Sample defined as 1973-1986.

Late Sample defined as 1987-1998.

Elite defined as institutions with 25th percentile undergraduate SAT>1210 or ACT>25.

**Table 5b. Impacts on Publishing by Size, Period and Elite Status**

| Dependent Variable: Citations to Publications | | | | | | Marginal Effects | | | Marginal Effects | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Large | Small | Early | Late | Elite | American | Foreign | Non-Elite | American | Foreign |
| | (5b) | (6b) | (7b) | (8b) | (9b) | | | (10b) | | |
| Total Students | 9.1017*** | 11.4392*** | 11.1813*** | -0.1694 | 13.2718*** | 0.515% | 0.867% | 5.2019*** | 0.475% | 0.475% |
| | (1.6490) | (1.2294) | (1.8119) | (2.8787) | (2.0558) | 10.8 cites | 18.2 cites | (1.3461) | 5.20 cites | 5.20 cites |
| Foreign Student Share | -387.6340* | -29.8941 | 33.5420 | -67.7143 | 420.7577* | ε = 0.192 | ε = 0.16 | 22.0593 | ε = 0.158 | ε = 0.104 |
| | (197.8693) | (58.2095) | (138.7528) | (136.5159) | (241.0307) | | | (94.2134) | | |
| Real equip & phys. plant exp, 5 yr MA ($ millions) | -2.9205 | -24.5299*** | -1.5094 | -3.6427 | -13.3826 | | | -35.4963*** | | |
| | (8.3619) | (6.0692) | (6.7221) | (10.6073) | (10.8337) | | | (6.8295) | | |
| Real R&D expenditures, 5 yr MA | -6.2766* | 18.3109*** | -2.5621 | 9.4148* | 0.8468 | | | -1.0190 | | |
| | (3.2549) | (3.1289) | (3.4128) | (5.0888) | (4.0821) | | | (3.1899) | | |
| Observations | 25515 | 22444 | 25589 | 22370 | | 14464 | | | 33495 | |
| Number of group(univ field) | 1148 | 1066 | 2157 | 2137 | | 648 | | | 1566 | |

Standard errors in parentheses, * significant at 10%; ** significant at 5%; *** significant at 1%.

All Regressions use Foreign Instruments in two-stage least squares regressions, university-field pair and year fixed effects, and university and field specific trends.

Large defined as above-median total PhD enrollment.

Early Sample defined as 1973-1986.

Late Sample defined as 1987-1998.

Elite defined as institutions with 25th percentile undergraduate SAT>1210 or ACT>25.

**Table 6a. Instrument Sets: Affects vs Does Not Affect Ability to Pay**

Dependent Variable: Publications by Department

| | Enrollment Induced by Shocks that are Pay Neutral | | | Enrollment Induced by Shocks Affecting Paying Students More | | | Pay-Neutral Shocks | Shocks that Affect Paying Students More |
|---|---|---|---|---|---|---|---|---|
| | | Publication effect at sample means of adding one more | | | Publication effect at sample means of adding one more | | | |
| | (11a) | American | Foreigner | (12a) | American | Foreigner | (13a) | (14a) |
| Total Students | 0.6280*** | 1.42% | 1.87% | 0.4443*** | 1.28% | 0.82% | 0.5552*** | 0.4027*** |
| | (0.0457) | 0.56 pubs | 0.74 pubs | (0.0221) | 0.51 pubs | 0.33 pubs | (0.0463) | (0.0254) |
| Foreign Student Share | 7.7529** | ε = 0.381 | ε = 0.281 | -7.9461*** | ε = 0.347 | ε = 0.122 | 17.3618*** | -11.3884*** |
| | (3.2407) | | | (2.1033) | | | (3.5137) | (2.2310) |
| Real Equipment & Phys. Plant exp. ($ millions) (5 yr MA) | | | | | | | -0.2974*** | -0.2587** |
| | | | | | | | (0.1103) | (0.1056) |
| Real R&D ($ millions) (5 yr MA) | | | | | | | 0.0171 | 0.1988*** |
| | | | | | | | (0.0638) | (0.0471) |
| Observations | | 57500 | | | 57500 | | 47959 | 47959 |
| No. of Field-University Pair FE | | 2300 | | | 2300 | | 2214 | 2214 |

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

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

Foreign IV set 1 in (11a) and (13a) include instruments that may affect paying and scholarship students equally: total # students studying abroad from home region, restriction policy dummy. Foreign IV set 2 in (12a) and (14a) include instruments that may affect paying students more than scholarship students: GDP, OECD dummy, GDP*OECD, Oil Revenue/GDP and exchange rate shocks.

**Table 6b. Instrument Sets: Affects vs Does Not Affect Ability to Pay**

Dependent Variable: Citations to Publications

| | Enrollment Induced by Shocks that are Pay Neutral | | | Enrollment Induced by Shocks Affecting Paying Students More | | | Pay-Neutral Shocks | Shocks that Affect Paying Students More |
|---|---|---|---|---|---|---|---|---|
| | | Citation effect at sample means of adding one more | | | Citation effect at sample means of adding one more | | | |
| | (11b) | American | Foreigner | (12b) | American | Foreigner | (13b) | (14b) |
| Total Students | 14.2560*** | 0.63% | 2.00% | 12.9253*** | 1.02% | 1.02% | 9.4307*** | 11.0602*** |
| | (2.4979) | 8.05 cites | 25.43 cites | (1.2423) | 12.93 cites | 12.93 cites | (2.5071) | (1.4165) |
| Foreign Student Share | 747.4936*** | ε = 0.169 | ε = 0.303 | -79.8663 | ε = 0.275 | ε = 0.153 | 947.3539*** | -414.1737*** |
| | (177.1241) | | | (118.3218) | | | (190.3052) | (124.4271) |
| Real Equipment & Phys. Plant exp. ($ millions) (5 yr MA) | | | | | | | -17.4867*** | -18.4395*** |
| | | | | | | | (5.9741) | (5.8881) |
| Real R&D ($ millions) (5 yr MA) | | | | | | | -2.5531 | -3.6817 |
| | | | | | | | (3.4547) | (2.6282) |
| Observations | 57500 | | | 57500 | | | 47959 | 47959 |
| No. of Field-University Pair FE | 2300 | | | 2300 | | | 2214 | 2214 |

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

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

Foreign IV set 1 in (11a) and (13a) include instruments that may affect paying and scholarship students equally: total # students studying abroad from home region, restriction policy dummy. Foreign IV set 2 in (12a) and (14a) include instruments that may affect paying students more than scholarship students: GDP, OECD dummy, GDP*OECD, Oil Revenue/GDP and exchange rate shocks.

**Table 7a. Accounting for Quality of Undergraduate Institution**

Dependent Variable: Publications by Department

| | (15a) | (16a) | (17a) | Elasticities, (17a) | Elite (18a) | Non-Elite (19a) |
|---|---|---|---|---|---|---|
| US student count, high-rank BA institution | -0.9067*** | -0.5653*** | -0.5400*** | ε = -0.37 | -0.6735*** | 1.8953*** |
| | (0.1649) | (0.1743) | (0.1696) | | (0.1544) | (0.3002) |
| US student count, other BA institution | 1.0280*** | 0.8405*** | 0.7971*** | ε = 0.54 | 1.0071*** | 0.0648 |
| | (0.0859) | (0.0871) | (0.0869) | | (0.1055) | (0.0885) |
| Foreign student count, high-rank BA institution | 1.7359*** | 1.7631*** | 1.8039*** | ε = 1.23 | 1.5920*** | 1.1903*** |
| | (0.1238) | (0.1293) | (0.1330) | | (0.1592) | (0.2032) |
| Foreign student count, other BA institution | 0.3743*** | 0.4183*** | 0.3904*** | ε = 0.27 | 0.4815*** | 0.1990*** |
| | (0.0382) | (0.0400) | (0.0410) | | (0.0615) | (0.0477) |
| Real equip & phys. plant exp, 5 yr MA ($ millions) | | -0.1237 | -0.1560 | | 0.1362 | -0.2051 |
| | | (0.1084) | (0.1099) | | (0.1661) | (0.1547) |
| Real R&D expenditures, 5 yr MA | | | 0.1027** | | 0.2069*** | -0.0339 |
| | | | (0.0495) | | (0.0641) | (0.0810) |
| Observations | 57500 | 49025 | 47959 | | 14464 | 33495 |
| Number of groups (univ-field) | 2300 | 2240 | 2214 | | 648 | 1566 |

Standard errors in parentheses, * significant at 10%; ** significant at 5%; *** significant at 1%.

All regressions have Field-University Pair and year fixed effects, and university and field specific trends.

Foreign Regional Instruments used in two-stage least squares regressions for US and foreign student counts.

Elite defined as institutions with 25th percentile undergraduate SAT>1210 or ACT>25.

High-rank baccalaureate institution defined the same as Elite, if in U.S., and by the Institute of Higher Education's (Jiao Tong Univ., Shanghai) World Rankings if BA institution foreign.

**Table 7b. Accounting for Quality of Undergraduate Institution**

Dependent Variable: Citations to Publications

| | (15b) | (16b) | (17b) | Elasticities, (17b) | Elite (18b) | Non-Elite (19b) |
|---|---|---|---|---|---|---|
| US student count, high-rank BA institution | -50.1017*** | -28.8769*** | -27.1562*** | ε = -0.58 | -37.8098*** | 44.3223*** |
| | (9.2024) | (9.6688) | (9.3504) | | (10.5438) | (13.7421) |
| US student count, other BA institution | 45.7840*** | 36.7972*** | 34.1628*** | ε = 0.73 | 50.4582*** | 0.0150 |
| | (4.7908) | (4.8289) | (4.7922) | | (7.2047) | (4.0533) |
| Foreign student count, high-rank BA institution | 15.5950** | -0.4178 | 5.1797 | ε = 0.00 | 27.0607** | -19.5309** |
| | (6.9072) | (7.1693) | (7.3320) | | (10.8678) | (9.3012) |
| Foreign student count, other BA institution | 10.7136*** | 10.0728*** | 7.2285*** | ε = 0.15 | 12.1936*** | 3.4994 |
| | (2.1312) | (2.2194) | (2.2597) | | (4.1973) | (2.1854) |
| Real equip & phys. plant exp, 5 yr MA ($ millions) | | -19.2808*** | -19.6767*** | | -9.6654 | -31.1286*** |
| | | (6.0118) | (6.0600) | | (11.3368) | (7.0817) |
| Real non-equipment R&D expenditures, 5 yr MA | | | -0.2725 | | -3.0804 | -1.2702 |
| | | | (2.7302) | | (4.3750) | (3.7082) |
| Observations | 57500 | 49025 | 47959 | | 14464 | 33495 |
| Number of groups (univ-field) | 2300 | 2240 | 2214 | | 648 | 1566 |

Standard errors in parentheses, * significant at 10%; ** significant at 5%; *** significant at 1%

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

Foreign regional instruments used for US & foreign, all regressions

Elite defined as institutions with 25th percentile undergraduate SAT>1210 or ACT>25

High-rank baccalaureate institution defined the same as Elite, if in U.S., and by the Institute of Higher Education's (Jiao Tong Univ., Shanghai) World Rankings if BA institution foreign

**Table 8a. Adding Instruments of US Conditions**

Dependent Variable: Publications by Department

| | 2SLS - w/ US Regional IV | Publication effect at sample means of adding one more | | 2SLS - w/ US Regional IV | 2SLS - w/ US Regional IV |
|---|---|---|---|---|---|
| | (20a) | American | Foreigner | (21a) | (22a) |
| Total Students | 0.4617*** | 1.09% | 1.29% | 0.4221*** | 0.3898*** |
| | (0.0206) | 0.43 pubs | 0.51 pubs | (0.0220) | (0.0224) |
| Foreign Student Share | 3.4164** | ε = 0.29 | ε = 0.19 | 6.2343*** | 5.4280*** |
| | (1.7362) | | | (1.9304) | (1.9970) |
| Real Equipment & Phys. Plant exp. ($ millions) (5 yr MA) | | | | -0.1597 | -0.2565** |
| | | | | (0.1038) | (0.1052) |
| Real R&D ($ millions) (5 yr MA) | | | | | 0.2646*** |
| | | | | | (0.0435) |
| Observations | | 57500 | | 49025 | 47959 |
| No. of Field-University Pair FE | | 2300 | | 2240 | 2214 |

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

All Regressions use Foreign and Domestic Instruments in two-stage least squares regressions.

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

**Table 8b. Adding Instruments of US Conditions**

Dependent Variable: Citations to Publications

| | 2SLS – w/ US Regional IV | Publication effect at sample means of adding one more | | 2SLS - w/ US Regional IV | 2SLS - w/ US Regional IV |
|---|---|---|---|---|---|
| | (20b) | American | Foreigner | (21b) | (22b) |
| Total Students | 12.2374*** | 0.66% | 1.51% | 10.5978*** | 9.2094*** |
| | (1.1604) | 8.38 cites | 19.18 cites | (1.2368) | (1.2544) |
| Foreign Student Share | 464.4165*** | ε = 0.18 | ε = 0.23 | 535.4355*** | 459.6407*** |
| | (97.6343) | | | (108.4483) | (111.7232) |
| Real Equipment & Phys. Plant exp. ($ millions) (5 yr MA) | | | | -16.0935*** | -17.6151*** |
| | | | | (5.8340) | (5.8842) |
| Real R&D ($ millions) (5 yr MA) | | | | | -0.7129 |
| | | | | | (2.4330) |
| Observations | | 57500 | | 49025 | 47959 |
| No. of Field-University Pair FE | | 2300 | | 2240 | 2214 |

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

All Regressions use Foreign and Domestic Instruments in two-stage least squares regressions.

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

**Table 9. Publication and Citation Counts - Students Broken Down by Region of Origin**

| Negative Binomial Regressions | Publication Counts | | Citation Counts | |
|---|---|---|---|---|
| | (23) | (24) | (25) | (26) |
| | Non-IV | IV | non-IV | IV |
| Enrollment from Region of Origin | | | | |
| United States | 0.0015 | 0.0029 | 0.0030 | 0.0035 |
| | (15.17)*** | (27.91)*** | (20.87)*** | (25.15)*** |
| China | 0.0016 | 0.0009 | -0.0049 | 0.0083 |
| | (4.52)*** | (0.46) | (9.41)*** | (2.61)*** |
| East & South-East Asia, Oceania | -0.0003 | 0.0200 | 0.0019 | 0.0217 |
| | (0.96) | (9.66)*** | (4.00)*** | (6.52)*** |
| South Asia | 0.0061 | 0.0757 | 0.0162 | 0.1692 |
| | (9.78)*** | (9.33)*** | (16.74)*** | (13.50)*** |
| Middle East and Africa | 0.0023 | 0.0219 | 0.0104 | 0.0547 |
| | (3.51)*** | (5.10)*** | (10.02)*** | (7.95)*** |
| Eastern Europe | 0.0148 | -0.0172 | -0.0065 | -0.0472 |
| | (11.94)*** | (5.71)*** | (3.00)*** | (9.88)*** |
| Western Europe | 0.0045 | 0.0047 | 0.0054 | 0.0477 |
| | (5.64)*** | (1.71)* | (3.86)*** | (10.54)*** |
| Western Hemisphere | 0.0033 | 0.0098 | 0.0045 | 0.0367 |
| | (4.49)*** | (1.62) | (3.59)*** | (3.56)*** |
| Observations | 62073 | 59774 | 62073 | 59774 |
| No. of University-Field Pair Fixed Effects | 2300 | 2300 | 2299 | 2299 |