PRIPE
7/24/06
3:00 PM

Foreign Graduate Students and Knowledge Creation at U.S. Universities: Evidence from Enrollment Fluctuations

Eric T. Stuen*

Ahmed Mushfiq Mobarak*

Keith E. Maskus*

Abstract

First draft: July 17, 2006.

Preliminary and Incomplete; Please do not Cite or Quote Without Permission.

Prepared for NBER Summer Institute, meetings on Innovation Policy and the Economy

*Eric Stuen is doctoral candidate, Ahmed Mushfiq Mobarak is Assistant Professor, and Keith E. Maskus is Professor, all at the department of economics, University of Colorado at Boulder. Contact: maskus@colorado.edu, 303-492-7588. 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, Dr. Bhaven Sampat for data on university patenting, and Rich Beaudoin and Christine Rohde for excellent research assistance.

1. Introduction

In this paper we explore statistically the roles that domestic and foreign graduate students play in developing knowledge and innovation in science and engineering (S&E) at U.S. universities. Knowledge is measured by scientific publications and innovation by patent applications, all 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. Surprisingly, however, this basic proposition has not been examined empirically at the detailed level of specific student, discipline, and university.

This issue has taken on considerable importance in recent years. 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 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, outgoing 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 has reached the top levels of policy debate and the Bush administration recently indicated that it would consider relaxing visa limits and changing

[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.

immigration rules for highly skilled post-graduate students. Concerns about the risk of a declining U.S. advantage in developing and deploying new technologies clearly underlie these debates.[4]

There are well-known deficiencies in its secondary education system. Indeed, among the major developed countries and the newly industrialized countries, the United States ranks near the bottom in mathematics and science achievement among eighth graders.[5] 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 (586,000 in 2002 compared to 270,000 in Britain) may help explain this seeming inconsistency. Foreign students are disproportionately more likely to earn graduate degrees in S&E. Indeed, in recent years foreign graduate students studying engineering in the United States have outnumbered their American counterparts.[6]

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.[7] This reversed a 15-year trend in which foreign graduate students increased by four percent per year on average. Computer science (which experienced a

[4] A good 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 February 2005.
[5] For comparison with other countries see the results of the Trends in International Mathematics and Science Study (TIMMS) at http://timss.bc.edu/timss2003.html.
[6] Data on graduate students by citizenship and field come from the Council of Graduate Schools' *Survey of Graduate Enrollment*, various years. Data on the total number of graduate students in the U.S. and Britain reported in "U.S. Slips in Attracting the World's Best Students," *New York Times*, December 21, 2004.
[7] The decline in foreign student applications was actually much larger at 28 percent, which raises the possibility that the students now enrolling in U.S. universities are of lower average ability than their

15 percent drop in foreign enrollment in 2002) and other S&E disciplines have 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. Moreover, the drop in foreign student enrollment is likely to continue, as the Education Testing Service recently reported that registration for the *Graduate Record Exam* at its international centers dropped considerably in 2004, including reductions of 37 percent in India, 43 percent in Taiwan and 50 percent in China.

Given the concerns of university officials and researchers, in conjunction with the reduction in foreign enrollments, it is important to study whether international graduate students are, in fact, significant contributors to the development of new technological knowledge. Our purpose in this paper is to perform a detailed micro-level econometric study. In particular, we assemble a database that has individual records on domestic and foreign doctoral graduates of U.S. departments of science and engineering over the period 1973-2004. We combine these records with publications in scientific journals, defined at the level of 23 specific fields in the major research universities of the United States. As noted, publications (and citation-weighted publications) form our measure of knowledge creation. In addition we combine the data on enrollments with the far-smaller number of patent applications (and citation-weighted applications) registered by American universities in specific areas of engineering and science over this period. Patent applications form our measure of innovation.

---

antecedent cohorts. Note also that foreign student enrollments began to decline noticeably in the late 1990s, as is evident from the data in the *Survey of Earned Doctorates*.

Our point of departure is that to identify the impacts of foreign doctoral students on knowledge creation, we need to exploit information arising from unanticipated shocks in the supply of foreign students to particular disciplines and universities. Otherwise, any correlations found between enrollments and university-field productivity may be associated with unobserved omitted variables. Thus, we employ an instrumental variables approach to estimating enrollments of foreign students, broken out by region of origin, as those enrollments are affected by macroeconomic shocks and policy changes. The basic idea, for example, is to see whether macroeconomic crises and such factors as the collapse of the Soviet Union, or the opening up of China to trade and investment, tended exogenously to expand student supplies and affect knowledge production and innovation. 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 Russian or Chinese students. This approach should be considerably more informative about the true contributions of graduate students than analysis performed to date.

At this stage the econometric results remain preliminary and additional analysis is in order. However, our initial analysis suggests the following observations.

The paper proceeds as follows. In the next section a brief literature review is offered. In Section 3 we develop the methodology for instrumentation and identifying shocks to enrollments, noting the performance of the instrumental variables. In Section 4 we work through the econometric results. Section 5 concludes.

2. Prior Literature

The policy debate reviewed above rests on the presumption that graduate students and post-doctoral students in the natural sciences, engineering, and medicine are central inputs into the development of knowledge at universities and subsequent innovation through patent licensing. While this presumption is intuitive and sufficiently powerful for graduate departments to advocate policy changes, it has not been rigorously tested in statistical terms. Regarding knowledge generation, we are unaware of studies that have linked the presence of graduate students to the number of publications by university and field.

Regarding innovation, 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, Ding and Stuart, 2005) and, in reverse, the impact of faculty patenting on scientific productivity measured by publications and citations (Breschi, Lissoni and Montobbio, 2005).

However, these studies have not considered the role of graduate enrollments. The latter angle has been the subject of just two recent studies, which try to link statistically the presence of foreign graduate students to future patenting. Both of these papers suffer from specification problems that make it difficult to assign such causality. The first is a working paper by Chellaraj, Maskus and Mattoo (2006; henceforth CMM) and the second a follow-on comment by Stephan, Black and Gurmu (2005; henceforth SBG).[8]

[8] Although the SBG paper has an earlier date, our reference to CMM is to a later version, written earlier this year. In fact, the study by CMM stimulated the comment in SBG.

The CMM paper supports the basic proposition that foreign students contribute to innovation by documenting a strong positive correlation between the presence of such students and patenting activity in the United States.[9] This was the first academic paper to address statistically the foreign students - innovation issue. CMM use annual aggregate U.S. data and show that in regressions of the total number of patent applications and patents awarded, the coefficients on both the number of lagged foreign graduate students in the United States (actually the proportion of foreign students, controlling for total number of students) and the lagged share of skilled foreign workers in the labor force are significantly positive. According to their estimates, a 10% rise in foreign students 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 cointegration relationships among students, immigrants, patents, and other variables in the specification. Nevertheless, 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. For example, an increase in the quality of American educational institutions and faculty – presumably correlated with patent success – could attract more foreign graduate students to the United States.

The SBG study 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

---

[9] The CMM paper is available at http://spot.colorado.edu/~maskus/. In contrast, Borjas (2002, 2004) points out some potential costs of the U.S. student visa program, including the crowding-out effect on American

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. However, the SBG paper suffers as well from its inability to isolate causation, leaving open the possibility of other factors driving the results.

Accordingly, this area is ripe for further research that attempts to identify the contributory role of foreign doctoral students to both publications and patenting by universities, with the analysis at the student-university-field level. In the next section we set out a methodology for this purpose.

3. Methodology

A. Basic Specifications

The empirical analysis we conduct examines the impacts of foreign students from different regions of origin enrolled at U.S. science and engineering Ph.D. programs on innovation and 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 a variety of outputs – including patents, publications, and citations to them – as a function of foreign and domestic student enrollment, as well as fixed effects for each field in each university, linear trends specific to each university and each field, and year dummies. 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

students and corruption in visa administration.

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 patenting at a particular university or field of inquiry. A linearized form of the basic specification we run is as follows:

$$Innovation_{f,u,t} = \alpha_{f,u} + \delta_t + \gamma_u (D_u * Trend) + \rho_f (D_f * Trend)$$

$$+ \Sigma_{\mathbf{r}} \beta_r * Foreign\_Students_{f,u,t-5,r} + \beta_8 * U.S.\_Students_{f,u,t-5} + \varepsilon_{f,u,t} \quad (1)$$

where *f, u, t, r* index the field of study (e.g. biochemistry), university (e.g. Yale), year (e.g. 1985) and students' region of origin (e.g. Western Europe) 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 ε is a mean-zero error term. *Foreign_Students* and *U.S._Students* measure enrollment, typically with a 5-year lag in the patent specifications and with a 1-year lag in the publications specifications. In actuality, since our dependent variables are counts of patents, publications or citations, we do not run linear regressions, but (non-linear) negative binomial fixed effects count-data models of the following form:

$$Innovation_{f,u,t} = e^{X_{f,u,t} \cdot \eta} \cdot \alpha_{fu} + \varepsilon_{f,u,t} \quad (2)$$

The vector $X_{f,u,t}$ in equation (2) encompasses all the variables in the summation between δ_t and *U.S._Students* (inclusive) in equation (1) above. $\alpha_{f,u}$ is a separate indicator for each field-university pair.

Given the fixed effects negative binomial specification with time dummies and field and university specific trends, the estimates for the coefficients of interest on the student enrollment variables (β_r) are consistent even in the presence of correlation between those variables and time invariant field-university specific unobservables or time-variant unobservables that are either constant across field-universities (such as any macro shocks, including changes in U.S. patent law), or follow a linear trend for each field or university (Hausman, Hall and Griliches, 1984). The objects of concern are time-variant unobservable characteristics of academic departments that affect both the innovation produced by those departments, and also their foreign student enrollments. For example, if the quality of a department improves over time (e.g. through greater funding and better faculty recruitment), this may attract greater numbers of foreign students, and may also have an independent effect on the department's output. This is likely to bias the β_r coefficients upward. 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 science and engineering fields, we may observe drops in foreign student enrollment 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 better language skills), foreign students may get crowded out in a department of limited size once high-quality American students start applying. This is likely to bias the β_r coefficients downward. The marketability or popularity of a particular field of study among students at a given point in time is another example of an omitted variable that may bias the impact of foreign students in either direction, depending on how U.S. students respond to such changes.

B. 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 β_r would no longer consistent, since that would imply $E(\eta|X) \neq E(\eta)$ in equation (2). In other words, foreign 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' countries of origin. We try to identify shocks that affect foreigners' decisions of whether to travel to the U.S. for graduate study, but that are plausibly uncorrelated with the innovation produced at specific academic departments in the United States. For example, an exchange rate shock in East Asia (e.g. a currency devaluation) would affect East Asians' ability to pay for a U.S. education, and can lead to fluctuations in East Asian enrollment at U.S. universities. Figures 1 and 2 plot enrollments of doctoral students from India and Nigeria at U.S. universities[10] against two relevant instruments (GDP growth and oil dependence). The co-movement of enrollment counts and each instrument as 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 economic and policy shocks in students' source countries for the purpose of identifying variation in innovation outputs from departments where those students enroll. Any shocks to the supply of foreign students that are uncorrelated with factors related to innovation in the United States allow us to identify the causal impact of changes in foreign students on department-specific innovation. Our estimation strategy uses a set of first-stage regressions where we instrument foreign-

[10] As reported in the National Science Foundation *Survey of Earned Doctorates*

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., oil price shocks linked to revenues generated in oil producing countries, or fluctuations in the U.S. dollar – 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, region of origin, and years of study. In order to exploit variation across all four dimensions in the data, we use the idea that the vulnerability to a student supply shock from a particular country will differ by field and by 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 and innovation at that institution. Similarly, if Indians are more likely to study engineering, then this shock would affect engineering departments more (and perhaps Purdue engineering 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 innovation at Purdue (an institution-specific effect) and at relatively strong engineering departments (a discipline-specific effect). The attached figures 3-5 demonstrate the empirical relevance of these ideas using aggregations from the National Science Foundation *Survey of Earned Doctorates* micro data on all Ph.D. recipients in the United States. 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 relative to the University of California – San Diego, and that electrical engineering departments benefited more than biochemistry departments. We will 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]

In our second-stage estimates of the determinants of publications, citations and patents, we examine the innovation impacts of only that portion of the variation of foreign-student supply that is attributable to 'shocks' unrelated to U.S. innovation and plausibly unanticipated by American universities. In practical terms, we carry out this two stage estimation procedure by first predicting foreign students using a linear (OLS) first-stage regression of $Foreign\_Students_r$ on the set of instruments, and using these predicted values of foreign students (rather than the actual values) in the second stage negative binomial regression (2). This two-stage procedure produces consistent estimates of β_r (Mullahy, 1997), although the covariance matrix estimates have to be adjusted to account for the sampling variation introduced by first-stage prediction of foreign students. We report bootstrapped standards errors for the second stage regression rather than analytically derived standard errors.

4. Data

A. Patent Regressions

We used university patent data compiled by Bhaven Sampat from the NBER patent database to create four indicators of field-university specific innovative activity: a count of patents granted to a university in a particular field, a citation-weighted patent count, a count of citations to university patents by patents granted to non-academic entities (to measure innovation spillovers outside of academia), and the ratio of non-academic citations to total citations. We used information on the university assignee, application year and the Jaffe and Trajtenberg (2002) two-digit patent category to assign individual patents to a university-field-year. We modified the Jaffe and Trajtenberg patent categories to create our own two-digit classification of 23 “fields” (i.e. patent categories) that better match the dissertation specialization of graduate students that is reported in the National Science Foundation *Survey of Earned Doctorates* data, which is our main data source for Ph.D. student enrollments. We generated counts of patents and citations for university-field-year cells for all 143 universities with any patenting activity, the 23 patent fields, annually for the period 1968-2003. Beginning with a comprehensive sample of 143 universities with at least one patent in the period, we limit out analysis to the top 80 patenting schools (defined as universities that have been granted at least 110 patents in total over the 36 year period), for the years 1980-2000.[11] We thus create a balanced panel dataset with 38.640 (80 universities x 23 fields x 21 years) university-field-year observations. The number of patents granted to the average university-field-year in this dataset is 0.48, while the citations average 3.1 (see Table 2, Summary Statistics). Both patents and citations increase over time.

B. Publications Regressions

We create counts of all science and engineering publications associated with the top 100 foreign Ph.D. granting universities in the United States for the period 1973-2001 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 fields of science and engineering (see Appendix Table 1 for a list of these 23 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 university-field-year (100x23x29) cells (please see the Data Appendix for details on the algorithm used for this assignment). Our final database is a count of publications and total citations in each university-field-year cell.

C. 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 data, which contains a record for each individual who received a Ph.D. in the United States between 1959 to present. Doctoral recipients fill out this survey when they receive the Ph.D. degree, so the yearly enrollment counts we create are based on the graduation date and the date of entry into the doctoral program reported by the students, and reflect only those students who have finished the degree. We infer enrollment counts

[11] We begin the patenting sample at 1980 to include solely the period after the Bayh-Dole Act was passed.

for the period 1960-1997 only, since there are likely to be many students who entered doctoral programs in 1998 or thereafter who still have not received their degree, 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 U.S. since 1960) studying in the 100 largest universities (those with at least 2100 doctoral students since 1960), in 23 Science and Engineering fields (as defined by Lach and Schankerman, 2003) during the period 1960-1997. This dataset therefore consists of just over 4.3 million (50 x 100 x 23 x 38) observations. Although we generally exploit country-level variation in the instrumental variables, most regressions use enrollment counts from aggregated "regions of origin" rather than the "country of origin". We define 8 regions on the basis of economic, geographic and cultural similarities between countries, and taking into account each country's relative importance as a supplier of students to American universities.

We also use information on each student's bachelor's degree granting institution coupled with a binary rating for "high quality" institutions to generate enrollment counts for the subset of "high quality students". Since creating the quality indicator forced us to rate all domestic and foreign universities as either high quality or not, the "quality" enrollment counts are measured with a lot of error.

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 Asia, and Australia/New Zealand, but excluding China) is the next largest supplier of students at 4.8, followed by China and then South Asia. The average number of "high quality" students enrolled at each university-field-year was 18, and 15 of these students were American.

D. Instrumental Variables

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

A. *Fluctuations in the Special Drawing Right – Foreign Currency Exchange Rate*: This measure exploits the idea that movements in, say, the Rupiah – SDR exchange rate during the Asian financial crisis altered Indonesians' ability to pay for a U.S. education. Currently, 70% of foreign students in the United States are from Asian and Latin American countries, which suggests that financial-crises related indicators could serve as appropriate instruments. The raw data are strongly suggestive that this instrument is likely to be powerful. Students from Korea, Thailand, Malaysia and Indonesia grew by 41% between 1992 and 1997, but dropped by 15.5% during the financial crisis years (1997-1999).

B. *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, the appended Figure 1 shows the pro-cyclical evolution of Indian GDP and Indian doctorate recipients from 1960-2004. 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.

C. *Oil Dependence*: 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 innovative activity in the U.S.

D. *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 Chinese non-language students' study abroad 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 includes Russia, Romania, Cuba, Poland, Hungary, Bulgaria, Czechoslovakia and (East) Germany. Please see the Data Appendix for details.

E. *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 U.K., Australia, Singapore etc. The idea here is that fluctuations in the number of South Asian students in the U.K. 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 U.S., the correlation is driven by the commonality between the two variables, which are the economic and policy conditions in South Asia.

5. Results

A. First-stage Instrumental Variables Regressions

Table 3 reports the first-stage instrumental variables regressions 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 a 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.

As a group, these instruments are quite powerful. The F-statistic on the set of instruments excluded from the second stage innovation regressions varies between 43.22 in the Western Hemisphere region (Latin America and Caribbean, Mexico, Canada) and 810.11 in the Eastern European region. These F-tests on excluded instrument all carry p-values of 0.000 up to at least 3 significant digits. Not surprisingly, the instruments produce a better fit for regions that are made up of more economically homogenous countries (e.g. China, Western Europe and Eastern Europe). The lowest F-statistics are attached to the regressions that predict total enrollments from East Asia/Australia/Pacific, and the Western Hemisphere.

In general, exchange rate devaluations are associated with reductions in student enrollments. Increases in GDP per capita generally expands enrollments from non-OECD countries (where changes in ability-to-pay might be key), but reduces it from OECD countries (where the opportunity cost of a domestic labor market may dominate decision-making). Positive oil shocks increase supply of students from oil-producing nations.

Given all the interaction terms included in table 3, it is very difficult to see the direction of effects for each instruments, since those effects are heterogenous across fields and universities, conditional on each field's and each university's historical dependence on students from a particular region. In table 4, we construct some examples of 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 only 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.

We turn next to econometric results from various specifications relating university publications and patents, broken down by field, to instrumented doctoral enrollments of

S&E students from eight regions, including the United States.[12] We proceed first through an analysis of publications and publication citations, then move on to patents and patent citations. We focus on results broken down by source region, though we also present regressions in which all foreign students are aggregated to a single group, after first-stage instrumentation. Keep in mind that an observation is a university/field/year combination. Second-stage regressions contain a comprehensive set of fixed effects for years and university-field pairs, along with time trends interacted with universities and fields.

B. Publications Regressions

To reiterate, 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 are still in graduate school and not counted in the SED database by 2004.

Our first set of results is presented in Table 5, which takes the number of scientific publications in each university-field pair as an observation. As a benchmark, the first column lists estimates of the negative binomial regression without fixed effects and no instrumentation of the enrollment data. From these results it seems an increase in U.S. doctoral enrollment of one student would increase university-field publications one year later by 1.13 percent and this impact is statistically highly significant. Other coefficients may be read similarly. An additional Chinese graduate student would raise publications by 1.98 percent, an Eastern European student by 4.53 percent, and a Western

[12] In fact, enrollments of American graduate students are not instrumented since macroeconomic fluctuations in the United States are not plausibly uncorrelated with university productivity.

Hemisphere student by 5.18 percent. The impacts of East and Southeast Asian students and Western European students are marginally significant. In this initial regression the coefficients for South Asian and Middle Eastern and African students are negative, suggesting that their enrollments tend to diminish publication counts on average. It is conceivable that, at the margin, enrollments of such students reduce the available spaces for others who might be more likely to contribute to publishable results. A more likely explanation, given that this regression has no university or field fixed effects, is that in the data these students are disproportionately located in lower-publication departments.

The second column turns to a more sensible specification, including fixed effects for universities, fields, and years, though we still employ raw (uninstrumented) enrollment data. Inclusion of these fixed effects significantly reduces the magnitudes of the coefficients, with the impact of another U.S. student falling to 0.25 percent additional publications. The impacts of students from South Asia, Middle East and Africa, and Western Europe are now positive and significant, with the greatest contribution from South Asia at approximately 4 times the U.S. coefficient. In contrast, the Chinese coefficient shifts to being significantly negative. The third column introduces university-field pair effects, with little impact on the coefficients except that of Western Europe.

In the next two columns we introduce field-specific and university-specific time trends, which are dummies for fields and universities multiplied by time trends, in order to capture any idiosyncratic behavior of scientific progress, but still do not employ the first-state instrumented enrollments. We present an OLS regression in the fourth column largely to demonstrate that results from simple least squares are radically different. As shown, the OLS specification generates large coefficients for nearly all classes of students, with the largest from Eastern Europe. That coefficient suggests that an additional student would increase publications by 1.83 articles and a further Chinese student would achieve a growth in publications of 0.36 articles.

In the fifth column we return to the negative binomial specification and discover that all country groups except East and Southeast Asia contribute positively to knowledge generation, though that coefficient is insignificant. The coefficients are small, however, suggesting an increase in publications ranging from 0.15 percent (U.S. students) to 1.48 percent (Eastern Europe). It is evident that foreign students in the aggregate contribute more than U.S. students in this regard.

A final specification is in the sixth column, where we have the full set of fixed effects and introduce instrumented enrollment predictions from the first stage. Again, the notion here is that these figures emerge from supply shocks that are uncorrelated with factors related to knowledge creation and innovation by university and field. Here the U.S. coefficient nearly doubles, with its marginal impact rising to 0.29 percent, while the Chinese coefficient becomes insignificant. Students from South Asia and the Middle East and Africa provide the largest contributions to lagged publications, with the former region offering a rise of 7.57 percent in publications from an additional student. In contrast to the earlier regressions, the Eastern European coefficient turns negative and significant, suggesting that students from that region may be the source of a "crowding out" effect in science and engineering.

It may be more interesting to consider the implied impacts of foreign students in terms of the number of publications, rather than percentage increase. Thus, in the final column we list the marginal effect of each student type, where this effect is defined as the change in the number of articles published after an increase of one student from the indicated region.[13] As noted, the data suggest that another South Asian student has a large positive impact (almost three publications), while the gain from American and Western European students is approximately the same (around 0.1 to 0.2 articles). Put differently, the marginal productivity of knowledge with respect to East and Southeast Asian, South Asian, and Middle Eastern and African students is high.

These results confirm that first-stage instrumental variables make a marked difference in the evident participation of foreign doctoral students in generating knowledge in American universities. The large increase in coefficients for East and Southeast Asia, South Asia and Middle East and Africa suggest that student supplies from those regions are particularly sensitive to macroeconomic shocks.

The results in Table 6 provide additional perspective. All regressions in this table incorporate the university-field and year fixed effects and the university and field trends, which form a comprehensive set of controls. The first two columns report negative binomial regressions of publication counts on U.S. student enrollments and aggregated foreign enrollments, without and with first-stage instrumentation. The first column suggests that both geographic sources provide students that contribute positively to research leading to publications, with the U.S. coefficient somewhat higher. Inclusion of the instrumented enrollments in the second column dramatically raises the foreign coefficient, by a factor of over six. Here, while an additional U.S. student raises publications by 0.32 percent, and additional foreign student raises them by 0.85 percent. This is the closest result in our paper to answering the question whether foreign graduate

[13] These figures are based on the mean number of publications (39.28) in a university-field combination over the period. Similar computations for other regressions are available on request.

students in the aggregate contribute as much as domestic graduate students. In terms of knowledge creation the answer seems to be that foreign students are more productive at the margin.

In the final six columns we report analogous regressions, where the dependent variable is the total number of publication citation counts.[14] The attempt here is to see if contributions to higher-quality publications (those with more citations) are similar to contributions to publications per se. Columns 3 and 4 replicate the experiment in columns 1 and 2. Without instrumenting, the results are quite similar to those for publications (compare columns 1 and 3). However, when instrumented foreign student data are used in the second stage estimation, the coefficient on international students increases by a factor of ten and is far larger than that for U.S. students. In short, exogenous shifts in foreign student supply are a highly significant determinant of the creation of knowledge in U.S. universities.

With the aggregate results in place, the fifth through eighth columns repeat the analysis with the regional breakdown for international students. Columns five and six remain in the realm of patent citations, with the difference being that foreign enrollments are instrumented in the latter regression. Here, first-stage instrumentation again has a substantial impact, with coefficients on regional student supplies increasing generally by a factor of ten. In this case, China's coefficient becomes positive and over twice that of the U.S. variable. Again, the only negative coefficient is for Eastern Europe, while South Asian students are the strongest determinant of publications.

[14] 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.

The final two columns of Table 6 attempt a somewhat different exercise. It is possible that macroeconomic and policy shocks affect students coming from various universities differently. For example, suppose that under normal circumstances there are relatively large enrollments of students from high-quality universities. A negative macroeconomic shock could therefore disproportionately reduce the supply of such students, tending to reduce publications in fields and universities that rely on them. To test such possibilities, we developed a crude binary (high or low) measure of the quality of the undergraduate institution from which each student graduated, both in the United States and abroad.[15] This binary variable is multiplied by enrollments, resulting in each observation retaining only students from high-quality universities and colleges.

Considering the final column, the results seem anomalous, though not impossible to explain. South Asia and Middle East and Africa retain strong impacts, suggesting that students from higher-quality institutions in those regions contribute strongly to higher-quality publications. In contrast, significantly negative and large impacts arise for students from Eastern Europe and the Western Hemisphere. The former result is consistent throughout our instrumented regressions, but the latter is a large shift in inference. It may be that the frequent macroeconomic shocks in Latin America do disproportionately reduce the ability of students from higher quality schools to attend graduate school in the United States or pushes those students into other, non-S&E fields of study. At the same time, enrollments from Canada (the largest source of students in this group) may be less sensitive to economic fluctuations.[16] At a minimum, the

[15] See the data appendix. We believe there is considerable noise in this ranking.
[16] It should be noted that the "regional" supply shocks are averages of country shocks, so that our inclusion of Canada in Western Hemisphere may be inappropriate, as Latin American economic events would dominate the instrumentation.

difference between Western Hemisphere and South Asia supports the idea that stochastic shocks in student enrollments may vary considerably across regions in their impacts on knowledge development.

C. Patents Regressions

Various results for specifications of the patent regressions are in Tables 7 and 8. The techniques used are OLS and instrumental-variables (two stage least squares).[17] In every case the regressions contain the full set of university-field and year fixed effects, plus the university and field time trends. Enrollment data are lagged five years in order to account for standard lags in preparing patent applications. In these tables the patents data extend from 1980 to 2000 and the students data from 1975 to 1995. Note that we are including only the period after implementation of the Bayh-Dole Act.

The first two columns report results for patent counts regressed on simple enrollment counts. Once again, instrumentation makes a substantive difference as the 2SLS results for several regions differ considerably from the OLS results, to which we pay little attention. Considering the 2SLS equations, U.S. students again have a significantly positive but relatively small impact on patent applications by universities, measured at the university-field level. Here, students from China are revealed to contribute considerably to patent applications, with a one-unit rise in Chinese enrollment

[17] At the time of this writing our attempts to compute negative binomial count-model regressions have not converged, presumably due to the large number of zero observations for patents at the university-field level, especially early in the period.

stimulating a rise in patenting by 0.11 applications.[18] The impacts of students from other regions are insignificant, except for those from Middle East and Africa.

Equations 3 and 4 regress patent counts on our measure of higher-quality enrollments (students from higher-quality undergraduate institutions). Here the U.S. coefficient is somewhat larger than for the full sample of students and the Chinese coefficient is similar. The impact of Middle East and Africa becomes even stronger positively and that for Eastern European students becomes significantly negative. Again, instrumentation matters for taking account of exogenous shifts in student supplies. It is noteworthy that students from higher-quality Western European students contribute strongly to the registration of university patents, in essence offsetting the negative impact of Eastern European enrollments.

Equations 5 through 8 repeat this analysis but use patent citation counts as the dependent variable. Again, the notion here is to explain variations in higher-visibility patent applications through differences in all enrollments and in high-quality enrollments. In truth, there is little consistency in these results and it is difficult to reach any conclusions, particularly as regards the last column with quality students.

Table 8 reports some supplementary regressions that are of some interest. The first four columns again perform the econometric analysis for U.S. students and aggregated foreign students. It seems from column 2 that both doctoral students from both areas contribute significantly and positively to university patent applications. The coefficient on foreign applications is significantly larger than that on domestic students.[19]

---

[18] These are linear regressions and therefore coefficients have a different interpretation from the negative binomial cases.

[19] This result comes closest to the question posed by Chellaraj, Maskus and Mattoo (2006) and is reasonably consistent with their finding.

In contrast, when patent citation counts are analyzed, the foreign student contribution becomes significantly negative. Taken at face value, this result would suggest that foreign student supply shocks are more important for developing low-quality patents than for high-quality patents. Additional work will be undertaken to try to sort this out.

Finally, we are interested in studying whether the data are informative about spillovers of innovation arising from the enrollment of graduate students. To assess this question, columns 5 and 6 perform regressions of non-academic patent citations to university patents on the regional breakdown of students.[20] Columns 7 and 8 supplement these by adding regressions in which the dependent variable is the share of non-academic patent citations in total citations. Comparing the results in columns 6 and 8 the sign patterns are consistent between the absolute and relative shares. There is a marked difference in signs and significance across regions in column 6, however. Again, Chinese students seem to detract from the ability to write high-quality patents, while Western European and Western Hemisphere students contribute positively.

6. Concluding Remarks

In this paper we report our initial findings regarding the contributions of domestic and foreign graduate students in S&E programs to knowledge creation (articles published) and innovation (patent applications) in the United States. To identify these impacts we undertook a first-stage instrumental variables approach to explaining shifts in foreign enrollments by region. The instruments plausibly are uncorrelated with domestic U.S. factors that influence university publishing and patent applications.

[20] It is widely recognized that non-academic patent citations may not be particularly informative about spillovers for they may be inserted by patent examiners rather than listed by patent applicants.

Interest arises in this question for a number of reasons. If foreign students are strong contributors to knowledge and university-level innovation, U.S. visa restrictions could have significantly negative impacts on research in the future. 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 and negatively affect prospects for research.

Our findings are fairly clear at the most basic levels. First, it is important to employ the first-stage instrumental variables for second-stage results, both in the linear (patents) and non-linear (publications) cases, differ markedly from the non-instrumented outcomes. Second, using aggregate enrollments, the impact of foreign students is clearly positive on both publications and patenting. Indeed, in both cases the contribution of international students is stronger than that of U.S. students. In this context, worries about aggregate visa restrictions or growth in opportunities abroad that reduce the overall supply of foreign students may be warranted.

Third, there are differences across regions in the estimated effects of international students on U.S. university productivity by field. South Asia, East and Southeast Asia, and Middle East and Africa seem most consistently to have positive impacts, while students from Eastern Europe register negative coefficients. The effects of students from China seem to be positive, particularly as regards patent counts but not patent citations. There is considerable variability in the findings regarding Western Europe and Western Hemisphere. Further study is important to determine the source of these differences, which presumably emanate from differential responses to exogenous shocks by region, before reaching confident conclusions about the relative productivity gains from admitting students from various locations.

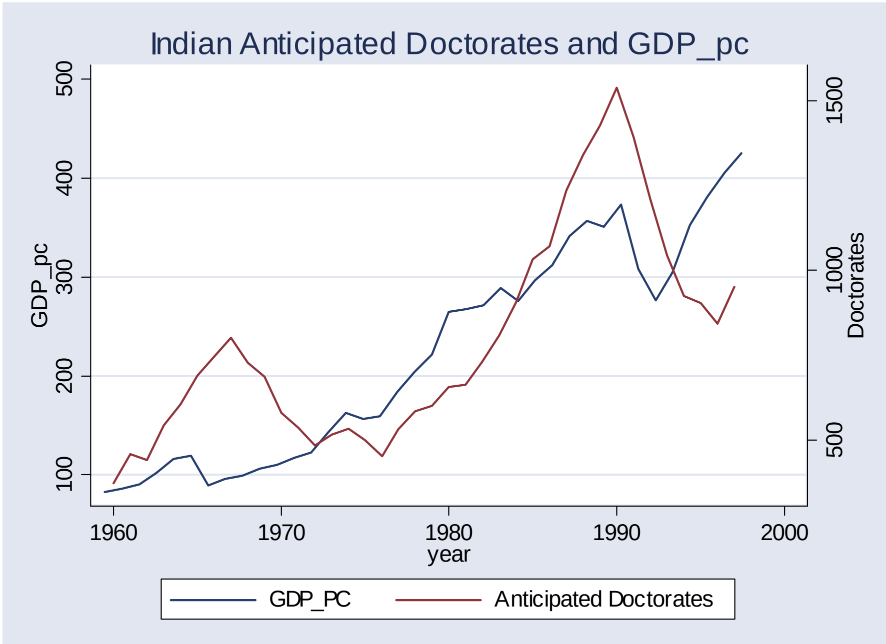

Figure 1

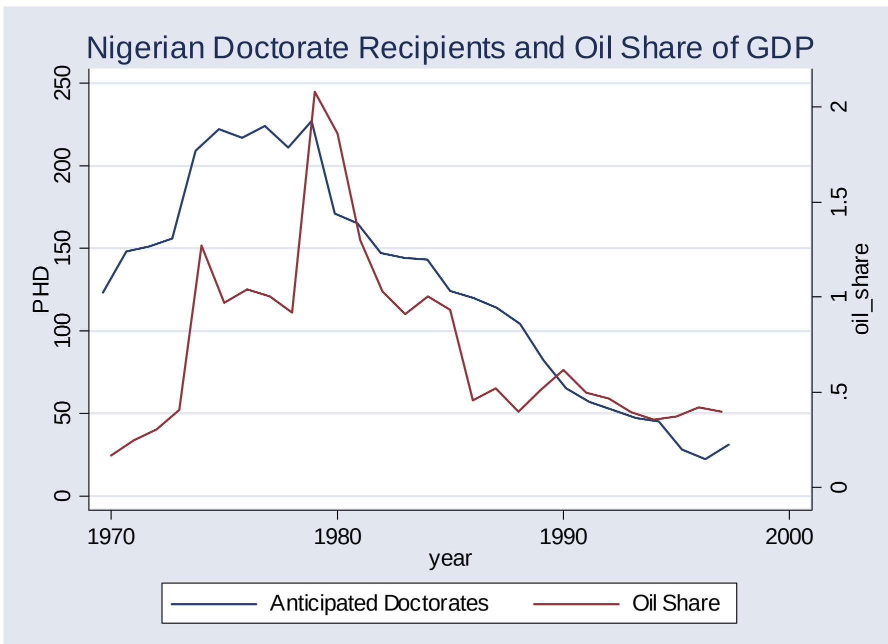

Figure 2

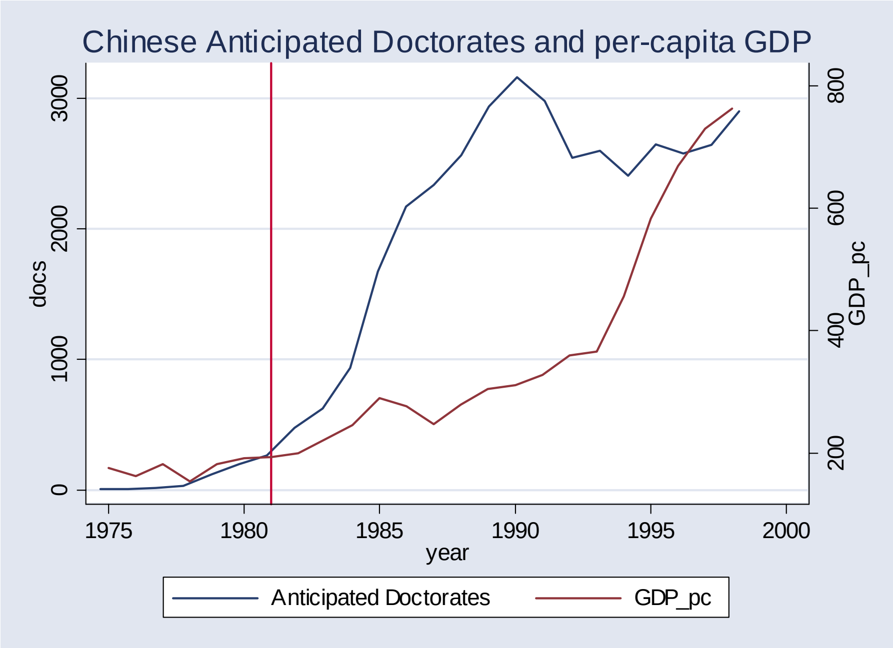

Figure 3

An exogenous 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 eighties may explain some of the magnitude in this spike of students.

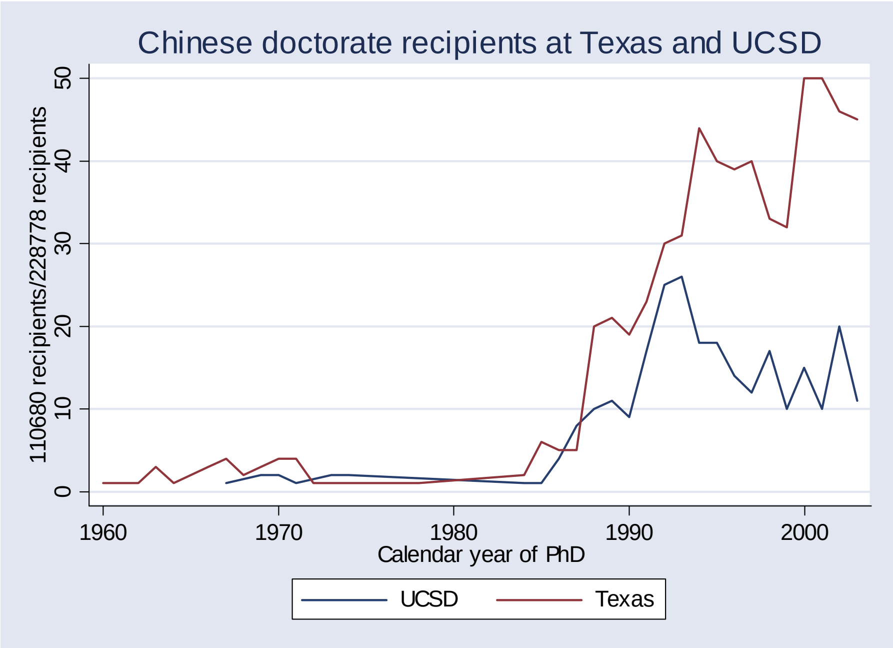

Figure 4

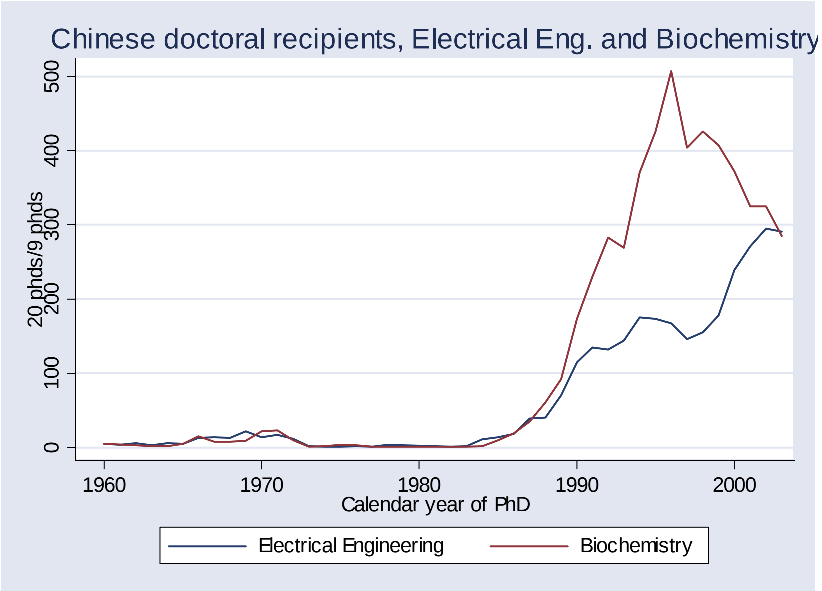
Chinese doctoral recipients, Electrical Eng. and Biochemistry
20 phds/9 phds
0
100
200
300
400
500
1960
1970
1980
1990
2000
Calendar year of PhD
Electrical Engineering
Biochemistry

Figure 5

References

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

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. (2006). “The Contribution of Skilled Immigration and International Graduate Students to U.S. Innovation.” University of Colorado Working Paper, March.

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.

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

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.

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

Stephan, P., Black, A. and Gurmu, S. (2005). "The Role of Foreign-Born Graduate Students in University Patenting." manuscript, Georgia State University.

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.

Table 1. Summary Statistics for Patent Data Set

	Variable	Mean	Std. Dev.	Min	Max
Patent Counts		0.480	2.156	0	134
	Patent Counts - 1970's	0.079	0.460	0	17
	Patent Counts - 1980's	0.318	1.069	0	26
	Patent Counts - 1990's	1.057	3.454	0	134
Patent Citations		3.097	15.574	0	497
	Patent Citations - 1970's	1.020	8.669	0	429
	Patent Citations - 1980's	4.395	19.133	0	497
	Patent Citations - 1990's	4.796	18.642	0	388
Enrollment - Total		36.326	77.930	0	1974
	Enrollment - US	23.548	55.825	0	1571
	Enrollment - Foreign	11.369	25.781	0	455
	Enrollment - China	2.356	7.034	0	150
	Enrollment - E., SE. Asia, Aus., NZ.	4.125	10.146	0	208
	Enrollment - South Asia	1.908	4.246	0	69
	Enrollment - Mid. East & Africa	1.701	4.368	0	76
	Enrollment - Eastern Europe	0.208	0.777	0	17
	Enrollment - Western Europe	1.147	3.057	0	56
	Enrollment - Western Hemisphere	1.363	4.116	0	76
Enrollment of High Quality Students - Total		15.937	40.389	0	1197
	High Quality Students - US	13.286	35.437	0	1098
	High Quality Students - Foreign	2.651	8.058	0	254
	High Quality Students - China	0.684	2.664	0	85
	High Quality Students - E., SE. Asia, Aus. & NZ.	1.303	4.359	0	101
	High Quality Students - South Asia	0.234	1.118	0	39
	High Quality Students - Mid. East & Africa	0.382	1.213	0	22
	High Quality Students - Eastern Europe	0.031	0.220	0	6
	High Quality Students - Western Europe	0.332	1.186	0	33
	High Quality Students - Western Hemisphere	0.432	1.631	0	43
Enrollment by Decade					
1970's	China	0.077	0.375	0	11
	E., SE. Asia, Aus. & NZ.	2.331	5.359	0	93
	South Asia	1.303	2.950	0	53
	Mid. East & Africa	1.213	3.696	0	75
	Eastern Europe	0.042	0.240	0	5
	Western Europe	0.819	2.519	0	48
	Western Hemisphere	0.994	3.466	0	69
1980's	China	1.680	4.698	0	129
	E., SE. Asia, Aus. & NZ.	4.802	10.965	0	200
	South Asia	1.822	3.779	0	56
	Mid. East & Africa	2.167	5.263	0	76
	Eastern Europe	0.088	0.348	0	5
	Western Europe	1.144	3.002	0	46
	Western Hemisphere	1.375	4.312	0	76
1990's	China	5.557	10.783	0	150
	E., SE. Asia, Aus. & NZ.	5.761	13.135	0	208
	South Asia	2.743	5.651	0	69
	Mid. East & Africa	1.956	4.341	0	55
	Eastern Europe	0.493	1.221	0	17
	Western Europe	1.546	3.602	0	56
	Western Hemisphere	1.803	4.669	0	74

Table 2. Summary Statistics for Publications Dataset

Variable		Mean	Std. Dev.	Min.	Max.
Publication Counts		38.305	59.179	0	1159
	Pub. Counts - 1970's	21.763	36.221	0	526
	Pub. Counts - 1980's	38.987	56.512	0	807
	Pub. Counts - 1990's	53.996	74.552	0	1159
Publication Citation Counts		1222.743	2753.796	0	71051
	Pub. Citations - 1970's	728.073	1719.857	0	37860
	Pub. Citations - 1980's	1273.471	2630.170	0	53257
	Pub. Citations - 1990's	1654.005	3550.745	0	71051
Total Enrollment		41.671	53.246	0	498
	Enrollment - US	26.970	37.102	0	380
	Enrollment - Foreign	14.701	21.985	0	317
	Enrollment - China	2.720	5.908	0	81
	Enrollment - E., SE. Asia, Aus. & NZ.	4.751	8.883	0	182
	Enrollment - South Asia	2.071	3.772	0	72
	Enrollment - Mid. East & Africa	1.879	3.636	0	56
	Enrollment - Eastern Europe	0.298	1.084	0	27
	Enrollment - Western Europe	1.434	2.728	0	35
	Enrollment - Western Hemisphere	1.548	3.088	0	36
Enrollment of High Quality Students - Total		18.180	26.657	0	345
	High Quality Students - US	15.031	22.440	0	254
	High Quality Students - Foreign	3.149	7.366	0	161
	High Quality Students - China	0.657	2.090	0	46
	High Quality Students - East&SE Asia, Aus,NZ.	1.170	3.452	0	104
	High Quality Students - South Asia	0.208	0.892	0	20
	High Quality Students - Mid. East & Africa	0.359	1.003	0	20
	High Quality Students - Eastern Europe	0.037	0.252	0	6
	High Quality Students - Western Europe	0.333	0.946	0	19
	High Quality Students - Western Hemisphere	0.386	1.121	0	25
Enrollment by Decade					
1970s	China	0.069	0.304	0	6
	E., SE. Asia, Aus. & NZ.	2.631	4.481	0	60
	South Asia	1.398	2.637	0	39
	Mid. East & Africa	1.352	2.890	0	40
	Eastern Europe	0.068	0.345	0	11
	Western Europe	0.955	2.102	0	34
	Western Hemisphere	1.126	2.586	0	27
1980's	China	2.125	4.436	0	58
	E., SE. Asia, Aus. & NZ.	5.131	9.064	0	175
	South Asia	1.947	3.398	0	41
	Mid. East & Africa	2.208	4.043	0	56
	Eastern Europe	0.147	0.563	0	14
	Western Europe	1.402	2.671	0	35
	Western Hemisphere	1.552	3.124	0	35
1990's	China	6.116	8.351	0	81
	E., SE. Asia, Aus. & NZ.	6.397	11.238	0	182
	South Asia	2.898	4.869	0	72
	Mid. East & Africa	1.995	3.709	0	55
	Eastern Europe	0.717	1.747	0	27
	Western Europe	1.952	3.218	0	33
	Western Hemisphere	1.965	3.429	0	36

Table 3. First Stage Instrumental Variables Regressions to Predict Enrollment Counts by Region of Origin

Region	China	E. Asia	S. Asia	ME & Africa	E. Europe	W. Europe	W. Hemisphere
	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 on	-0.0083	-0.0045	-0.0461	-0.0120	-0.0005	-0.0010	0.0041
Region	(0.72)	(12.98)***	(6.39)***	(11.70)***	(0.43)	(6.34)***	(3.12)***
Interacted with Field's Dependence on	-0.0829	-0.0017	-0.0207	-0.0041	-0.0020	0.0000	0.0018
Region	(10.42)***	(6.94)***	(4.89)***	(5.75)***	(2.21)**	(0.12)	(1.72)*
Triple Interaction: Both University's and	-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 on		0.0028		0.0036	-0.0262	0.0004	0.0003
Region		(2.46)**		(1.10)	(5.82)***	(2.10)**	(0.16)
Interacted with Field's Dependence on		0.0019		0.0000	0.0020	0.0001	0.0012
Region		(1.91)*		(0.01)	(1.13)	(0.34)	(0.63)
Triple Interaction: Both University's and		-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 U.S. Dollar Exchange Rate					-0.0415		-0.0414
(Region Median)					(1.86)*		(0.33)
Interacted with University's Dependence on	81.4627	-5.0376	6.5642	16.9729	-2.8568	18.3986	2.6629
Region	(5.73)***	(0.18)	(0.50)	(1.87)*	(0.82)	(1.68)*	(0.69)
Interacted with Field's Dependence on	53.2826	22.4271	14.4241	23.0192	1.2057	10.6179	0.7015
Region	(5.62)***	(1.34)	(1.92)*	(3.75)***	(0.60)	(1.31)	v
Triple Interaction: Both University's and	-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 Non-	0.0000	-0.0000	0.0000	-0.0000	-0.0000	0.0000	0.0000
U.S. Hosts	(3.19)***	(3.07)***	(3.21)***	(3.39)***	(1.49)	(3.93)***	(2.86)***
Interacted with University's Dependence on	-0.0004	0.0000	-0.0007	0.0001	-0.0001	-0.0001	-0.0006
Region	(2.25)**	(0.73)	(2.97)***	(3.75)***	(2.17)**	(3.81)***	(3.44)***
Interacted with Field's Dependence on	0.0000	0.0000	-0.0004	0.0000	-0.0001	-0.0000	-0.0003
Region	(0.35)	(0.17)	(3.08)***	(2.79)***	(3.91)***	(2.82)***	(1.87)*
Triple Interaction: Both University's and	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)***
	0.2616				0.8704	-0.8062	3.6078
Policy Restrictions on Study Abroad	(4.89)***				(44.14)***	(5.30)***	(1.55)
Interacted with University's Dependence on	-0.0512				60.5353	21.5546	3.0403
Region	(0.02)				(16.21)***	(4.47)***	(0.04)
Interacted with Field's Dependence on	15.1254				27.2640	15.8089	-130.5287
Region	(9.13)***				(13.00)***	(4.14)***	(2.15)**
Triple Interaction: Both University's and	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 U.S. Dollar 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 on		-32.7045	1.7015	-0.4159	0.5261	1.8033	0.0866
Region		(4.34)***	(0.39)	(3.51)***	(0.87)	(2.37)**	(1.96)*
Interacted with Field's Dependence on		-16.4295	-0.9475	-0.3903	0.9535	1.7816	0.0351
Region		(3.72)***	(0.40)	(4.65)***	(2.67)***	(3.22)***	(1.02)
Triple Interaction: Both University's and		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 on				24.5665	-192.9500		-125.8367
Region				(7.60)***	(4.44)***		(3.77)***
Interacted with Field's Dependence on				4.7928	-65.7356		-66.8567
Region				(2.53)**	(2.47)**		(2.52)**
Triple Interaction: Both University's and				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 on		-27.8052		-8.5849		-13.7152	17.2458
Region		(3.65)***		(1.51)		(4.51)***	(2.47)**
Interacted with Field's Dependence on		-25.1727		-7.5112		-7.7804	10.1142
Region		(3.84)***		(2.48)**		(3.25)***	(1.68)*
Triple Interaction: Both University's and		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 trends

Table 4. Some Examples of the Effects of the Instrumental Variables on Enrollment Counts

University, Field and Decade	Region of Origin	Instrument	Marginal Effect	Elasticity
U. of Maryland - Industrial Engineering - 1980's	China	GDP per Capita	0.027	7.01
U. of Virginia - Statistics - 1990's	East and South-East Asia	Percentage Change in Exchange Rate (max. in region)	-1.596	-0.57
Texas A&M U. - Electrical Engineering - 1980's	South Asia	GDP per Capita	0.0064	0.162
U. of Oklahoma - Cellular & Dev. Biology - 1980's	Middle East & Africa	Oil Share of GDP	3.82	4.22
Columbia University - Physiology - 1990's	Western Europe	GDP per Capita	-0.00023	-2.8
Cornell University - Ecology & Evolution - 1990's	Western Hemisphere	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 models 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 5. Regressions of Publication Counts

	(1)	(2)	(3)	(4)	(5)	(6)	
	Negative Binomial Count Data Models			Linear Regression	Negative Binomial Count Data Models		Marginal Effect
Enrollment from Region of Origin			FE	FE	FE	FE-IV	
United States	0.0113 (52.47)***	0.0025 (24.37)***	0.0029 (27.86)***	0.1801 (21.64)***	0.0015 (15.17)***	0.0029 (27.91)***	0.11
China	0.0198 (16.45)***	-0.0021 (6.23)***	-0.0015 (4.58)***	0.3643 (14.08)***	0.0016 (4.52)***	0.0009 (0.46)	0.04
East & South-East Asia, Oceania	0.0017 (1.83)*	0.0019 (5.98)***	0.0016 (4.96)***	-0.0478 (2.36)**	-0.0003 (0.96)	0.0200 (9.66)***	0.79
South Asia	-0.0232 (13.83)***	0.0111 (16.60)***	0.0121 (18.90)***	0.2031 (5.00)***	0.0061 (9.78)***	0.0757 (9.33)***	2.97
Middle East and Africa	-0.0339 (19.59)***	0.0053 (7.51)***	0.0040 (5.58)***	0.0227 (0.52)	0.0023 (3.51)***	0.0219 (5.10)***	0.86
Eastern Europe	0.0453 (9.53)***	0.0021 (1.62)	0.0017 (1.30)	1.8294 (18.83)***	0.0148 (11.94)***	-0.0172 (5.71)***	-0.68
Western Europe	-0.0047 (1.72)*	0.0025 (2.77)***	-0.0009 (0.95)	0.3512 (5.89)***	0.0045 (5.64)***	0.0047 (1.71)*	0.18
Western Hemisphere	0.0518 (22.12)***	0.0036 (4.46)***	0.0033 (3.82)***	0.4035 (7.06)***	0.0033 (4.49)***	0.0098 (1.62)	0.38
Constant	3.1819 (470.48)***	2.0431 (8.66)***	2.2789 (167.73)***	-3,154.1406 (71.23)***	-1,080.7575 (.)	-1,142.3124 (.)	
Observations	62100	62073	62073	59800	62073	59774	
R-Squared				0.51			
Year Fixed Effects:	No	Yes	Yes	Yes	Yes	Yes	
University Fixed Effects (100)	No	Yes	No	No	No	No	
Field Fixed Effects (23)	No	Yes	No	No	No	No	
University-Field Pair Fixed Effects (2300)	No	No	Yes	Yes	Yes	Yes	
University and Field Specific Trends (123)	No	No	No	Yes	Yes	Yes	
Enrollment Instrumented:	No	No	No	No	No	Yes	

Absolute value of z- or t-stats in parentheses

Table 6. Negative Binomial Fixed Effects Regressions of Publication and Citation Counts

	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Dependent Variable:	Publication Counts		Counts of Citations to All Publications					
Independent Variable:	Total Enrollment of All Students						Students from High Quality Institutions	
Enrollment from Region of Origin	FE	FE-IV	FE	FE-IV	FE	FE-IV	FE	FE-IV
United States	0.0028	0.0032	0.0033	0.0036	0.0030	0.0035	0.0062	0.0061
	(26.77)***	(28.61)***	(22.92)***	(26.33)***	(20.87)***	(25.15)***	(28.68)***	(27.70)***
Total Foreign	0.0018	0.0085	0.0024	0.0217				
	(12.40)***	(4.83)***	(11.79)***	(13.70)***				
China					-0.0049	0.0083	-0.0015	-0.0282
					(9.41)***	(2.61)***	(1.29)	(6.40)***
East & South-East Asia, Oceania					0.0019	0.0217	-0.0027	0.0010
					(4.00)***	(6.52)***	(2.94)***	(0.29)
South Asia					0.0162	0.1692	0.0222	0.4444
					(16.74)***	(13.50)***	(7.35)***	(16.06)***
Middle East and Africa					0.0104	0.0547	0.0274	0.1913
					(10.02)***	(7.95)***	(9.19)***	(6.75)***
Eastern Europe					-0.0065	-0.0472	-0.0586	-0.1475
					(3.00)***	(9.88)***	(7.60)***	(4.93)***
Western Europe					0.0054	0.0477	-0.0182	-0.0015
					(3.86)***	(10.54)***	(6.30)***	(0.13)
Western Hemisphere					0.0045	0.0367	-0.0110	-0.1237
					(3.59)***	(3.56)***	(5.07)***	(8.39)***
Constant	-605.9296	-473.4852	-1,087.5361	-1,093.8830	-1,068.9889	-1,045.7346	-976.1827	-1,008.1062
	(41.33)***	(.)	(.)	(.)	(.)	(.)	(.)	(.)
No. of Observations	59774	59774	59774	59774	62073	59774	62073	59774
No. of Field-University Pairs	2299	2299	2299	2299	2299	2299	2299	2299
Enrollments Instrumented:	No	Yes	No	Yes	No	Yes	No	Yes

All regressions have group and year fixed effects, and university and field trends.
Absolute value of z- stats in parentheses

Table 7. Fixed Effects and Instrumental Variables Linear Regressions of Patents and Patent Citation Counts

	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
	Patent Counts				Patent Citation Counts			
	All Students		Students from High Quality Institutions		All Students		Students from High Quality Institutions	
	FE	FE-IV	FE	FE-IV	FE	FE-IV	FE	FE-IV
Enrollment By region								
United States	0.0309	0.0245	0.0481	0.0363	-0.0019	0.0172	0.0437	0.1140
	(32.91)***	(12.34)***	(35.33)***	(12.82)***	(0.22)	(0.91)	(3.48)***	(4.47)***
China	0.0686	0.1114	0.1647	0.0990	-0.1582	-0.6959	-0.8369	-2.0355
	(23.91)***	(6.71)***	(22.51)***	(3.06)***	(5.98)***	(4.40)***	(12.41)***	(6.98)***
East & South-East Asia,	-0.0114	0.0084	-0.0403	-0.0519	0.1624	0.5353	-0.0415	-0.3836
Oceania	(4.79)***	(0.67)	(8.31)***	(1.55)	(7.40)***	(4.47)***	(0.93)	(1.27)
South Asia	-0.0460	-0.0339	-0.1803	-0.1056	-0.0367	-1.3977	-0.0745	2.1295
	(9.37)***	(0.97)	(12.92)***	(1.12)	(0.81)	(4.22)***	(0.58)	(2.52)**
Middle East and Africa	0.0067	0.0892	0.0359	0.3649	-0.1120	-1.7005	0.1274	1.3470
	(1.28)	(2.56)**	(2.63)***	(4.27)***	(2.32)**	(5.13)***	(1.01)	(1.75)*
Eastern Europe	0.0652	-0.0093	-0.0655	-0.8431	-1.5089	-0.4085	-3.5813	-1.1162
	(3.46)***	(0.24)	(1.06)	(2.79)***	(8.68)***	(1.09)	(6.30)***	(0.41)
Western Europe	-0.0477	0.0091	0.0859	0.8745	-0.1679	0.6501	-0.9215	-5.5481
	(6.11)***	(0.29)	(5.11)***	(8.77)***	(2.33)**	(2.19)**	(5.94)***	(6.17)***
Western Hemisphere	0.0245	-0.0572	0.1164	-0.2455	0.0396	1.8872	-0.0564	2.1210
	(3.51)***	(1.20)	(9.65)***	(3.33)***	(0.61)	(4.16)***	(0.51)	(3.19)***
Constant	-69.4881	-38.6966	-70.9654	-64.5891	157.0476	-72.2426	-3.1802	-229.5297
	(11.27)***	(3.68)***	(11.56)***	(7.15)***	(2.76)***	(0.72)	(0.06)	(2.82)***
Observations	34839	32706	34839	32706	34839	32706	34839	32706
No. of University-Field Pairs	1659	1580	1659	1580	1659	1580	1659	1580
R-Squared	0.35		0.36		0.05		0.06	

All regressions control for university-field pair and year fixed effects, and university and field specific trends.
Absolute value of t-stats in parentheses

Table 8. FE and IV Regressions of Patents, Citations, Non-Academic Citations, and Ratio of Non-Academic to Total Citations

	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
Total Enrollment of All Students from Region of Origin	Patent Counts		Patent Citation Counts		Non-Academic Patent Citation Counts		Fraction of Citations from Non-Academic Patentees	
	FE	FE-IV	FE	FE-IV	FE	FE-IV	FE	FE-IV
United States	0.0285	0.0208	0.0077	0.0531	-0.0071	0.0102	0.000	-0.000
	(30.33)***	(13.03)***	(0.89)	(3.65)***	(0.95)	(0.63)	(0.35)	(0.57)
Total Foreign	0.0115	0.0359	-0.0316	-0.1753				
	(10.97)***	(7.85)***	(3.28)***	(4.21)***				
China					-0.1397	-0.4995	-0.002	-0.015
					(6.08)***	(3.66)***	(3.82)**	(5.21)**
East & South-East Asia, Oceania					0.1341	0.4138	0.002	0.013
					(7.03)***	(4.01)***	(4.01)**	(5.79)**
South Asia					-0.0030	-1.1257	0.001	-0.005
					(0.08)	(3.94)***	(1.31)	(0.87)
Middle East and Africa					-0.0915	-1.3861	-0.000	-0.028
					(2.18)**	(4.85)***	(0.18)	(4.54)**
Eastern Europe					-1.1901	-0.3799	-0.019	-0.005
					(7.88)***	(1.18)	(5.81)**	(0.67)
Western Europe					-0.2005	0.5948	0.002	0.010
					(3.21)***	(2.33)**	(1.55)	(1.92)
Western Hemisphere					0.0152	1.3344	-0.000	0.015
					(0.27)	(3.42)***	(0.39)	(1.77)
Constant	-91.1086	-59.9896	290.5655	110.8683	137.0525	-15.7905	4.874	0.132
	(14.89)***	(6.58)***	(5.18)***	(1.33)	(2.78)***	(0.18)	(4.59)**	(0.07)
Observations	34839	32706	34839	32706	34839	32706	34839	32706
Groups	1659	1580	1659	1580	1659	1580	1659	1580
R-squared	0.34		0.04		0.05		0.07	

Absolute value of z- or t-stats in parentheses

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

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

Categories of Patents

Organic Compounds & Resins
Agriculture, Food, Textiles
Coating
Gas
Miscellaneous-chemical
Computer Technology
Communications
Drugs
Surgery & Medical Inst.
Biotechnology
Miscellaneous-Drgs & Medical
Electrical and Electronic Devices
Nuclear & X-rays
Semiconductor Devices
Mechanical Devices and Metal Working
Mat. Proc & Handling
Motors & Engines + Parts
Optics
Transportation
Animal Husbandry, Food
Apparel & Textile

1. ***Dependent Variables***

a) **Patent Counts:** University patent data were compiled by Bhaven Sampat from the NBER patent database. We used information on the university assignee, application year and two-digit patent category to assign patents to university/field/year count observations. The sample contains all 143 universities with patents, and spans 1968-2003. We modified the two-digit classification, designed by Jaffe and Trajetenberg (JT) as a broader classification than the USPTO's three-digit classification, to gain a better match of dissertation specializations of graduate students with patent groups. Table 1 includes a description of both the JT classifications and ours (SMM).

JT Subcategory	SMM Subcategory	Patent Count	
14 & 15	10	5018	Organic Compounds & Resins
11	11	220	Agriculture,Food,Textiles
12	12	584	Coating
13	13	122	Gas
19	19	3786	Miscellaneous-chemical
22, 23 & 24	20	1915	Computer Technology
21	21	1115	Communications
31	31	6723	Drugs
32	32	2400	Surgery & Med Inst.
33	33	6606	Biotechnology
39	39	376	Miscellaneous-Drgs&Med
41, 42, 43 & 49	40	3214	Electrical and Electronic Devices
44	44	1150	Nuclear & X-rays
45	45	895	Power
46	46	1299	Semiconductor Devices
52 & 59	50	744	Mechanical Devices and Metal Working
51	51	542	Mat. Proc & Handling
53	53	204	Motors & Engines + Parts
54	54	594	Optics
55	55	168	Transportation
61	61	564	Animal Husbandry, Food
63	63	43	Apparel & Textile
NA	70	0	Basic Scientific Research
62, 64, 65, 66, 67, 68, 69	0	dropped	Non-science

Dr. Sampat graciously also provided data on patents that cite patents in the university set, and marked these as academic or non-academic. From this, we created three measures of patent quality: **patent citation counts**, **non-academic citation counts** and the ratio **non-academic citations/total citations**.

b**)** **Publication Counts:** We chose 100 universities based on 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 catories and department affiliation(s). Using an algorithm (described below), we sorted the publication records into 23 fields of science and engineering. We then constructed count data based on 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 (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 1 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:
a) and it is a "unique" category, it is assigned directly to associated field;
b) and it is "non-unique", but the associated "typical department" is listed and matches a field, then it is assigned to the associated field;
c) and it is "non-unique", and the department does not match a field, it should be assigned to the "highest ranking" (see below) field that is associated with the subject category.

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

3. If there are no subject categories listed:
a) and the department listing matches a field, it should be assigned to that field;

b) and 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 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 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,* as well as a twenty-fourth field, "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[21], 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 enrollment" is that since the SED only goes to 2004, inferred enrollment counts for the most recent years underestimate total enrollment. Assuming an

[21] 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.

average time to degree of six years a student entering in 1999 would graduate in 2004. To be conservative, we use inferred enrollment counts only to 1997, although counts for 1995-1997 will have some slight truncation because students finishing in 2005 and 2006, who took 9 or 10 years to complete their Ph.D., 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 regarded "high-quality" or just "quality". We then followed a process similar to (a) to create a count of "quality" students per university/field/year observation.

3. Instrumental Variables

a) Per-capita GDP

GDP data were constructed from the World Bank's *World Development Indicators* 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 taken from ****Mad****. 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

"Oil Share of GDP" is the ratio of real oil revenues to real GDP. Real oil revenues were calculated as production quantity multiplied by real oil price. Production data for crude oil were taken from the US 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 comes from UNESCO's *Statistical Yearbooks* 1963-1998, and UNESCO's online database for post-1998. The data is 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-US 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 Unite 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 early seventies, 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 *glasnost*. 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.