06-022

Ethnic Scientific Communities and International Technology Diffusion

William R. Kerr

Ethnic Scientific Communities and International Technology Diffusion

William R. Kerr*
Harvard Business School and MIT

25 April 2007

Abstract

This study explores the importance of knowledge transfer for international technology diffusion by examining ethnic scientific and entrepreneurial communities in the US and their ties to their home countries. US ethnic research communities are quantified by applying an ethnic-name database to individual patent records. International patent citations confirm knowledge diffuses through ethnic networks, and manufacturing output in foreign countries increases with an elasticity of 0.1-0.3 to stronger scientific integration with the US frontier. To address reverse-causality concerns, reduced-form specifications exploit exogenous changes in US immigration quotas. Consistent with a model of sector reallocation, output growth in less developed economies is facilitated by employment gains, while more advanced economies experience sharper increases in labor productivity. The ethnic transfer mechanism is especially strong in high-tech industries and among Chinese economies. The findings suggest channels for transferring codified and tacit knowledge partly shape the effective technology frontiers of developing and emerging economies.

JEL Classification: F22, J44, J61, O31, O32, O33, O41, O57.

Key Words: Technology Transfer, Tacit Knowledge, Productivity, Patents, Innovation, Research and Development, Entrepreneurship, Immigration, Networks.

*Comments are appreciated and can be sent to wkerr@hbs.edu. This paper is forthcoming in the *Review of Economics and Statistics*. I am especially grateful to Daron Acemoglu for advice on this project; to Ajay Agrawal, George-Marios Angeletos, Josh Angrist, Pol Antras, David Autor, Nick Bloom, Ricardo Caballero, Richard Freeman, Xavier Gabaix, Francesco Giavazzi, Michael Greenstone, Tarun Khanna, Josh Lerner, Ashley Lester, Guido Lorenzoni, Byron Lutz, Megan MacGarvie, Ramana Nanda, Philip Oreopoulos, Sari Pekkala, Michael Piore, Julio Rotemberg, John Van Reenen, and Ivan Werning and four referees for insightful thoughts; to seminar participants at Boston University, Clemson, Federal Reserve Board of Governors, Goldman Sachs, Harvard, LSE, MIT, and the NBER for comments; and to Dan Johnson and Brian Silverman for data assistance. This research is supported by the National Science Foundation, the MIT George Schultz Fund, and HBS Research. This paper was first circulated in August 2004; a previous version was released as HBS Working Paper 06-022.

1 Introduction

The adoption of new technologies and innovations is a primary engine for economic growth, improving worker productivity and spurring higher standards of living. Invention, however, is concentrated in advanced economies. OECD countries account for 83% of the world's R&D expenditure and 98% of its patenting (OECD 2004). Even within the OECD, a disproportionate share of R&D is undertaken in the US. Diffusion of new innovations from technologically leading nations to following economies is thus necessary for the economic development of poorer regions and the achievement of global prosperity.

Economic models often describe a worldwide technology frontier, where new ideas and innovations travel quickly to all countries.[1] Rapid diffusion may be a good approximation for industrialized economies, but many advances are either not available or not adopted in poorer countries. Case studies in the business sociology and economic history literatures suggest this poor adoption may result from inadequate access to the informal or practical knowledge that complements the codified details of new innovations. Be it between two people or two countries, knowledge transfer is much more complicated than sharing blueprints, process designs, or journal articles. Intellectual spillovers are often thought to be important for the formation of cities and high-tech clusters, and perhaps heterogeneous access to the codified and tacit knowledge associated with new innovations shapes the effective technology sets of following countries.[2]

Recent research stresses the importance of ethnic scientific communities in frontier countries for conveying new technologies to their home countries. In surveys of Silicon Valley, 82% of Chinese and Indian immigrant scientists and engineers report exchanging technical information with their respective nations; 18% further invest in business partnerships (Saxenian 2002a,b). Studies of software off-shoring suggest 30% of India's systems workforce rotates through the US to obtain the tacit knowledge necessary for their work (Piore 2004). Moreover, some observers believe the success of India versus Mexico and other countries in this field derives in part from India's strong US entrepreneurial community. More generally, explorations of knowledge diffusion find countries with a common language have larger R&D spillovers and international patent citation rates (e.g., Keller 2002b, Jaffe and Trajtenberg 1999).

Ethnicity thus offers an observable channel for exploring the extent to which international networks transmit the codified and tacit knowledge of new inventions. This study examines whether a larger ethnic research community in the US improves technology diffusion to foreign

[1] For example, Mankiw et al. (1992) and Heckscher-Ohlin trade theory. Recent descriptions of multiple technology frontiers build on geographic distances to major R&D nations (e.g., Keller 2002b), the innovations of trading partners (e.g., Grossman and Helpman 1991, Coe and Helpman 1995, Coe et al. 1997), or international patenting decisions (e.g., Eaton and Kortum 1999). Keller (2004) reviews the technology transfer literature.

[2] Marshall (1890) and Jacobs (1970) describe the forces contributing to spatial agglomeration, while Rosenthal and Strange (2003) and Ellison et al. (2007) provide more recent empirical tests. Other country-specific differences that inhibit adoption include barriers to technological investment, capital-labor or human-capital disparities, differences in the organization of production, and the appropriateness of technology. Representative papers in this literature are Parente and Prescott (1994), Atkinson and Stiglitz (1969), Nelson and Phelps (1966), Banerjee and Newman (1993), and Acemoglu and Zilibotti (2001), respectively.

countries of the same ethnicity. US ethnic research communities are quantified by applying an ethnic-name database to individual US patent records (e.g., identifies inventors with Chinese versus Hispanic names). These matched records describe the ethnic composition of US scientists and engineers with previously unavailable detail. These trends are joined with industry-level manufacturing data for foreign countries (e.g., Chinese computer research in the US is paired with China's computer industry) in an econometric framework that isolates the role of scientific integration by exploiting within-industry variation.

To clarify this empirical methodology, the next section develops a theoretical model where a technology follower depends on the imitation of frontier innovations for technical progress in its manufacturing sector. In order to imitate these frontier technologies, however, scientists in the following country require codified and tacit knowledge with respect to the frontier inventions. This knowledge is acquired and transferred through the scientists of the following country's ethnicity who work in the frontier economy. The model thereby relates the technology follower's manufacturing output and productivity growth to its scientific integration with the leader. The primary estimating equations employed in this study are determined within this framework.

Section 3 then describes the ethnic patenting dataset constructed, and a first characterization of ethnicity's role in international knowledge transfer is undertaken through citation patterns. Foreign researchers are found to cite US researchers of their own ethnicity 30%-50% more frequently than researchers of other ethnicities, even after controlling for detailed technology classes. A further examination divides the sample into different time lags from the filing dates of the cited US patents to the dates of the citing foreign patents. This analysis reveals that the own-ethnicity effect is most important during the first five years of the diffusion process. After peaking in the fifth year, the higher ethnic citation rates decline to the tenth year.

While informative, citation patterns do not quantify the extent to which following countries realize economic benefits from better access to US innovations. To characterize foreign output and productivity realizations, the US ethnic patenting data are combined with industry-level manufacturing data for foreign countries in Section 4. Ethnic research communities are quantified at the industry-year level by aggregating individual patent records. Panel estimations then test whether output increases in foreign countries as their respective ethnic research communities in the US develop. The specifications only exploit within-industry variation. The results suggest growth in US ethnic scientific communities increases foreign output with elasticities of 0.1-0.3 depending upon how the data are weighted. These parameter estimates are economically and statistically significant, and the output expansion is decomposed into employment and labor productivity gains.

The inclusion of multiple countries and industries affords a more structured characterization of ethnicity's role for technology diffusion and economic growth than case-based or survey-based research. This platform also allows us to test the robustness of the results to other explanatory factors and to assess the extent to which the well-documented transfers of Asian high-tech enclaves in Silicon Valley generalize to other settings. The measured elasticities are moderately robust to further incorporating human-capital and physical-capital developments abroad, general country trends, and so on. Performance in these tests is weakened by the less variation in growth of US research communities that exists across industries within an ethnicity than across ethnicities. Sample decompositions further find that the outcomes are especially strong in high-tech sectors and within the Chinese ethnicity. While measurable growth effects are present in the broader sample, they are substantially weaker than the showcase examples often discussed.

Reverse causality is a prominent concern in these types of specifications, where human-capital developments in the foreign country could simultaneously result in higher output growth and more ethnic researchers emigrating to the US. Section 5 returns to the theoretical model to highlight how immigration quotas offer a foothold for addressing these issues. The resulting reduced-form strategy is applied in the context of the Immigration Act of 1990, a major revision of the US quotas system, that led to a surge in the immigration of scientists and engineers from previously constrained countries. The immigration quotas exercise suggests that growth in US ethnic research communities increases foreign output with elasticities of 0.3-0.4. While the immigration experiment cannot resolve omitted variable biases, the qualitative findings of this exercise support the results found with the ethnic patenting approach.

Finally, the diverse set of countries studied affords additional insights regarding how the benefits accruing to technology followers differ by development stage. An extension to the theoretical model allows sector reallocation from agriculture to manufacturing. After a transition point to full employment in the manufacturing sector, greater technology transfer raises labor productivity and output levels with constant employment. This is the steady-state description developed in Section 2. Prior to this transition, however, the following country responds with growth in manufacturing employment as well as labor productivity gains. Consistent with these predictions, interactions with development stage show labor productivity growth is mostly concentrated in economies that have transitioned to full manufacturing employment (e.g., the Asian tiger economies); countries with large agricultural sectors instead increase industry output through higher employment levels (e.g., Mainland China, India).

The results of this project suggest poor access to the codified and tacit knowledge regarding new innovations does contribute to slow technology diffusion. Ethnic scientific and entrepreneurial channels are important for the transfer of this practical or informal information, and thus differences in ethnic research communities in frontier economies are partly responsible for the heterogeneous technology opportunities of developing or emerging countries. In addition to characterizing technology diffusion, a better understanding of these ethnic linkages provides an important contribution to the "brain drain" versus "brain circulation" debate. While a full cost-benefit analysis is beyond this paper's scope (e.g., Kapur and McHale 2005), the technology transfer results highlight a potential benefit from high-skilled immigration to advanced countries.

2 Theoretical Framework

This section outlines a simple leader-follower model of technology transfer. Both economies feature a manufacturing sector characterized by an expanding-product-variety production function where technological progress occurs through the adoption of new intermediate products used in production of final goods. Entrepreneurial scientists living in each country supply these new technologies for profit, and they can either invent the intermediate products themselves or imitate foreign innovations. Spillovers from past innovations increase the research productivity of current scientists for invention and generate endogenous growth.[3] Knowledge is local, however, in that a country's researcher productivity for invention builds only on its own past research. That is, the capabilities of the two nations to invent evolve separately.

Researchers can alternatively imitate foreign inventions for use in their own country. Their effectiveness in doing so, however, depends upon their codified and tacit knowledge with respect to the foreign country's innovations. In preparation for the empirical analysis, ethnicity is incorporated into the framework to model this knowledge network. Specifically, the following country is of homogeneous ethnicity; the frontier country is primarily of another ethnicity but is home to some researchers of the following country's ethnicity. These frontier expatriates acquire and transmit the knowledge necessary for effective imitation in the following country.

Variables for the leader's economy are denoted by a tilda (e.g., $\tilde{Y}$), while the follower's variables are in plain font (e.g., Y). Superscripts and subscripts further distinguish ethnicity and sector as required. The first section outlines the core elements of follower's economy, followed by differences in the leader's economy. The steady-state outcome is then characterized.[4]

2.1 Follower's Economy

The technology follower's economy contains L workers of homogeneous ethnicity F employed in manufacturing and research. Its labor market is competitive, such that workers are free to move between the two sectors and are paid their marginal products of labor in each. Denote the workers employed in manufacturing and research by L_M and L_R, respectively. The behavior of the manufacturing sector is first described, followed by the research sector and consumers.

The competitive manufacturing sector produces final goods Y_M that can be consumed or used to make intermediate manufacturing goods. The price of final goods is normalized to one. Production for a representative firm i that employs labor L_{M_i} and non-durable intermediates X_{ij} of type j takes the form

$$Y_{M_i} = AL_{M_i}^{1-\alpha} \sum_{j=1}^{N} (X_{ij})^{\alpha}. \tag{1}$$

[3]For example, Romer (1990), Rivera-Batiz and Romer (1991), and Barro and Sala-i-Martin (1995).

[4]Section 4 discusses transitional dynamics to this steady state when labor reallocation from an agricultural sector is introduced. Technology flows are the only interactions between the two countries. The model abstracts from trade, and immigration is restricted in the base scenario.

α is the elasticity of output with respect to intermediate inputs ($0 < \alpha < 1$), A is a common manufacturing productivity parameter, and N is the number of intermediate product varieties currently available in the follower. In equilibrium firms employ equal amounts of all intermediate inputs ($X_{ij} = X_i \ \forall j$) and (1) can be simplified to $Y_{M_i} = AL_{M_i}^{1-\alpha} X_i^{\alpha} N = AL_{M_i}^{1-\alpha}(NX_i)^{\alpha} N^{1-\alpha}$. Thus, the production function exhibits constant returns to scale in labor and total intermediate inputs NX_i, but a larger number N of intermediate goods increases output by distributing the total intermediate inputs over more goods and thereby raising the marginal product of each.

Technical progress takes the form of increases in N, either through inventions I or imitations M of the leader's inventions ($N = I + M$). Entrepreneurial research firms choose between invention and imitation by comparing the productivity of the two techniques. The research productivity for invention in the follower is determined by the existing stock of the follower's inventions, or $\partial I/\partial t = I \cdot L_R$. There are no international knowledge spillovers in the sense that researchers in the follower cannot build on the the leader's stock of inventions directly in innovation. The follower's researchers can alternatively imitate the leader's inventions at a rate

$$\frac{\partial M}{\partial t} = \left(\tilde{I} \Psi \left[\frac{M}{\tilde{I}} \right] (\tilde{H}^F)^{\beta} \right) \cdot L_R, \tag{2}$$

where $\tilde{I}$ is the leader's invention stock and $\tilde{H}^F$ is the follower's human-capital stock with respect to the leader's inventions. A larger stock of frontier inventions affords a larger pool of technologies that can be imitated, thus raising the imitation productivity for the follower's researchers. The imitation of products exhausts the available pool, however, and the function Ψ decreases with the ratio of imitated products to the available frontier stock, $\Psi' < 0$. $\Psi[1] = 0$ when all available products have been imitated, and $\Psi[0]$ is sufficiently large to ensure some imitation occurs with human capital for foreign technologies. The $(\tilde{H}^F)^{\beta}$ specification models that tacit knowledge of frontier inventions is necessary for successfully adopting them in the follower. This human-capital stock depreciates at a rate δ, and the population of follower's ethnic researchers in the leader undertaking inventive activity adds to it: $\partial \tilde{H}^F/\partial t = -\delta \tilde{H}^F + \tilde{L}_R^F$.

Regardless of how new products are acquired, the entrepreneurial research firms gain perpetual monopoly rights over the production and sale of new intermediate goods in the follower. The present discounted value of these rents for a good j at time t is $V(t) = \int_t^{\infty} (P_j - C_j) X_j e^{-\bar{r}(s,t)\cdot(s-t)} ds$, where P_j is the selling price and C_j is the cost of producing the intermediate good. $\bar{r}(s,t)$ is the average interest rate between times t and s, which is constant in equilibrium. $C_j = 1$ for research firms as one unit of Y_M is required to produce one unit of intermediate input.

Monopoly rights afford research firms the power to set P_j in each period to maximize $(P_j - 1)X_j$. As price takers, the manufacturing firms equate the marginal product of an intermediate good, $\partial Y_{M_i}/\partial X_{ij}$ in (1), with its price P_j for a demand of $X_{ij} = (A\alpha/P_j)^{1/(1-\alpha)} L_{M_i}$. Substituting this demand function into the research firm's maximization problem, summing across final-goods producers, and taking the derivative with respect to P_j yields the monopoly price $P_j = \alpha^{-1}$. Thus, research firms charge the same price ($P_j = P$) and face similar aggregate demands of

$X = A^{1/(1-\alpha)}\alpha^{2/(1-\alpha)}L_M$. The constant interest rate, price, and aggregate demand relationships simplify the value of inventing or imitating a new technology $V(t)$ to

$$V = \left(\frac{1-\alpha}{\alpha}\right) A^{1/(1-\alpha)}\alpha^{2/(1-\alpha)}\frac{1}{r}L_M. \tag{3}$$

Constant intermediate demand functions also simplify the follower's aggregate output,

$$Y_M = A^{1/(1-\alpha)}\alpha^{2\alpha/(1-\alpha)}L_M N. \tag{4}$$

On the consumer side, households maximize a linear lifetime utility function $U = \int_0^\infty c(t) \cdot e^{-\rho t}dt$, where ρ is the rate of time preference. Consumers earn wage w and receive the interest rate r on savings. In equilibrium, $\rho = r$.

2.2 Leader's Economy

Before the equilibrium for the follower's economy can be determined, the frontier economy must be described. The leader's economy is identical to the follower's except in its ethnically heterogeneous labor force and in its invention of new intermediate goods. Workers of both the leader's and follower's ethnicity live in the leader. Workers of the leader's ethnicity move between the manufacturing and research sectors, but the follower's expatriates work only in the research sector ($\tilde{L}_M = \tilde{L}_M^L$, $\tilde{L}_R = \tilde{L}_R^L + \tilde{L}_R^F$). The follower's ethnic population in the leader is small enough to ensure some scientists of the leader's ethnicity are always required. The aggregate populations of the two countries are equal ($L = \tilde{L}$).

Researchers of both ethnicities contribute to and utilize the existing frontier invention stock $\tilde{I}$ in developing new intermediate products: $\partial\tilde{I}^F/\partial t = \tilde{I} \cdot \tilde{L}_R^F$ and $\partial\tilde{I}^L/\partial t = \tilde{I} \cdot \tilde{L}_R^L$, where $\tilde{I} = \tilde{I}^F + \tilde{I}^L$. This research specification again highlights the role of past inventions $\tilde{I}$ in making current researchers more productive, and assumes inventions made in the follower do not contribute to the leader's researcher productivity for invention. More subtly, ethnicity does not matter for invention in the leader — both types of scientists are symmetric with respect to the frontier invention stock. Finally, frontier researchers of the follower's ethnicity can imitate products made in the follower with a productivity analogous to (2).[5]

2.3 Steady-State Description: Leader Invents, Follower Imitates

This case determines the core estimating equation for this study. Without invention in the follower, the frontier economy operates in isolation, and imitation does not occur ($\tilde{N} = \tilde{I}$). The leader's research sector is competitive with respect to labor markets, and scientists earn the marginal product of their innovative efforts. Denote by $\tilde{V}$ the present discounted value of

[5]The potential crowding out of US workers and students from science and engineering fields by immigrants is often debated (e.g., Borjas 2005, Freeman 2005). This model incorporates a crowding-out effect for analytical convenience only.

making a new invention in the leader. As researchers invent $\tilde{I}$ new products each period (i.e., $(\partial \tilde{I}/\partial t)/\tilde{L}_R = \tilde{I}$), the wage paid to scientists is $\tilde{V} \cdot \tilde{I}$. Likewise, wages in the manufacturing sector are equal to the marginal product of labor $(1-\alpha)\tilde{Y}_M/\tilde{L}_M$. Labor mobility between sectors requires that these wages be equal, $\tilde{V} \cdot \tilde{I} = (1-\alpha)\tilde{Y}_M/\tilde{L}_M$. Substituting into this free-entry condition the leader's versions of the value of innovations (3) and aggregate output (4), and noting $r = \rho$, the steady-state allocation of labor in the frontier economy is found to be $\tilde{L}_M = \rho/\alpha$ and $\tilde{L}_R = L - \rho/\alpha$. Thus, the growth rate of both the stock of frontier intermediate technologies and manufacturing output is $L - \rho/\alpha$.

Returning to the follower's economy, all intermediate products come through imitation of the leader ($N = M$). Labor mobility again requires wage equality for the follower, $V \cdot (\tilde{I}\Psi[M/\tilde{I}](\tilde{H}^F)^\beta) = (1-\alpha)Y_M/L_M$. Substituting in the value of new intermediates V from (3) and aggregate output Y_M from (4),

$$r = \frac{\tilde{I}}{M}\Psi\left[\frac{M}{\tilde{I}}\right](\tilde{H}^F)^\beta \alpha L_M. \tag{5}$$

With identical preferences and aggregate populations, the follower's interest rate and allocations of labor to manufacturing and research are the same as the leader.[6] Equation (5) further shows the steady-state ratio of the follower's imitated products to available frontier products $M/\tilde{I}$ is constant and increases with the follower's human-capital stock with respect to leader's technologies ($\Psi' < 0$). Stronger knowledge transfer improves researcher productivity for imitation in the follower and closes the steady-state gap to the frontier.

Simplifying (5) for economies of equal size relates the follower's imitated technology stock to the technology frontier and the follower's knowledge for frontier innovations, $M = \tilde{I}\Psi[M/\tilde{I}](\tilde{H}^F)^\beta$. Substituting this relationship into the follower's manufacturing output (4), taking logs, and collapsing time-invariant terms into a constant ϕ, the follower's manufacturing output depends upon its human-capital stock with respect to frontier research with elasticity β: $\ln(Y_M) = \phi + \ln(\tilde{I}) + \beta \ln(\tilde{H}^F)$. The human-capital stock is $\delta^{-1}\tilde{L}_R^F$ in steady-state, so that

$$\ln(Y_M) = \phi + \ln(\tilde{I}) + \beta \ln(\tilde{L}_R^F), \tag{6}$$

where δ^{-1} is absorbed into the constant. Equation (6) is the basis for the estimating equations employed in Sections 4 and 5. The statistical framework will return to the intricacies of empirically estimating this relationship, but the outlook is promising that the relationship will be directly identified if this scenario holds.

The follower's imitation-versus-invention decision determines the condition required for this steady-state description. Specifically, the productivity of the follower's researchers for invention must be less than the researcher productivity for imitating frontier innovations in equilibrium,

$$I < \tilde{I}\Psi\left[\frac{M}{\tilde{I}}\right](\tilde{H}^F)^\beta. \tag{7}$$

[6] These conditions hold for more general utility functions. As Barro and Sala-i-Martin (1995) note, technological diffusion can equalize rates of return without other interactions between economies.

The assumption $I = 0$ requires (7) hold forever; without a knowledge stock on which to build, a first invention is impossible. While this may describe extremely poor regions, the more interesting implication for developing or emerging countries is that, even with a small invention stock, the comparative benefit to imitation can be sustained so long as access to the codified and tacit knowledge for a growing stock of frontier innovations is maintained. Section 5 discusses the case where (7) no longer holds.[7]

3 Ethnic Patenting and International Citations Analysis

The above model is applied to technology transfer from the US through ethnic networks. Estimation of the β parameter requires quantifying each ethnicity's human-capital stock with respect to US research. This section outlines the dataset built for this exercise, and presents an analysis of knowledge flows using international patent citation records. The ethnic patenting data are then joined with foreign output metrics in Section 4 to evaluate (6) directly.

3.1 Ethnic Patenting Records

Ethnic technology development in the US is quantified through the NBER Patent Data File originally compiled by Hall et al. (2001). This dataset offers detailed records for all patents granted by the United States Patent and Trademark Office (USPTO) from January 1975 to December 1999. Each patent record provides information about the invention (e.g., technology classification, citations of prior art) and the inventors submitting the application (e.g., name, city). To estimate ethnicities, a commercial database of ethnic first names and surnames is mapped into the inventor records. Kerr (2007c) documents the name-matching algorithms, lists frequent ethnic names, and provides extensive descriptive statistics. The match rate is 98% for US patent records, and the process affords the distinction of nine ethnicities: Chinese, English, European, Hispanic, Indian, Japanese, Korean, Russian, and Vietnamese.

Table 1 describes the 1985-1997 US sample, while Figure 1 illustrates the evolving ethnic contribution to US technology development as a percentage of patents granted by the USPTO. The trends demonstrate a growing ethnic contribution to US technological development, especially among Chinese and Indian scientists. Ethnic inventors are more concentrated in high-tech industries like computers and pharmaceuticals and in gateway cities relatively closer to their home countries (e.g., Chinese in San Francisco, European in New York, and Hispanic in Miami). The final three rows demonstrate a close correspondence of the estimated ethnic composition to the country-of-birth composition of the US science and engineering workforce in the 1990 Census.

[7]Immigration is restricted in this framework. Moreover, the follower's workers would prefer to emigrate to the leader as the frontier wage rate is higher *ceteris paribus* due to the larger stock of intermediate goods.

3.2 International Patent Citation Analysis

The ethnic-name database is also applied to foreign patent records registered in the US. Inventions originating outside the US account for just under half of USPTO patents, with applications from Japan comprising 45% of this foreign total. Kerr (2007c) presents the matched characteristics for countries grouped to the ethnicities identifiable with the database. From a quality-assurance perspective, the results are very encouraging. First, the ethnic-name database assigns ethnicities to 98% of foreign records. Second, the estimated inventor compositions are quite reasonable, with the own-ethnicity contributions in all but three regions being greater than 80% (e.g., 89% of inventors filing from Chinese countries and regions are classified as ethnically Chinese). Like the US, own-ethnicity shares should be less than 100% due to foreign researchers.

In addition to serving as a quality-assurance check, patents registered with the USPTO by foreign inventors afford an initial characterization of international knowledge flows through ethnic scientific networks. Each patent record includes citations of prior inventions on which the current patent builds, and the pattern of these citations can be informative about communication channels between researchers.[8] This first exercise simply compares the ethnic composition of cited US inventors across different foreign inventor ethnicities. That is, do Chinese inventors living outside of the US tend to cite more Chinese inventors living in the US than their technology field would suggest?

Inventor names are only included with patents granted from 1975-1999, and the data are cut in two ways to form a uniform sample. First, only the citations of foreign patent applications to the USPTO from 1985-1997 are considered. Second, the application year of the cited US patent must be within ten years of the application date of the citing foreign patent. That is, citations of 1975-1984 US domestic patents are considered for foreign patents applied for in 1985, while 1976-1985 is the appropriate ten-year window for 1986 patents. In addition, all within-company citations and patents with inventors in multiple countries are excluded.[9]

From this sample, citation counts are developed by cells that contain four dimensions: 1) the ethnicity of the citing foreign inventor, 2) the ethnicity of the cited US inventor, 3) the technology class of the citing foreign inventor, and 4) the technology class of the cited US inventor. The latter two dimensions are necessary for isolating ethnicity's role since patents cite other patents within their technology field far more frequently than those outside of their field. If ethnicities concentrate in different industries in the US and abroad, measured ethnic flows could be merely capturing that technologies build upon prior art in their own discipline.

[8] Jaffe et al. (2000) and Duguet and MacGarvie (2005) discuss using patent citations to study knowledge transfer. Jaffe et al. (1993), Peri (2005), Hu and Jaffe (2004), Agrawal et al. (2006), and MacGarvie (2006) are examples of applications in an international distance context.

[9] Patents may have multiple inventors with different ethnicities. The reported regressions only consider citations for which a dominate ethnicity can be assigned to both patents (i.e., a single ethnicity accounts for strictly more than 50% of multiple inventors). English-ethnicity inventors abroad are excluded. These restrictions are required for the cells constructed for the citations estimations and are not carried forward into the output and productivity analyses. The results are robust to alternative techniques like Thompson (2006) below.

Almost 100,000 cells are formed with this organization, and many cells contain zero values. The zero values are due to both the small sizes of some ethnicities (e.g., Vietnamese inventors outside of the US) and that researchers in a given field simply do not cite the universe of technologies in their work. Count data containing zero values can be appropriately handled with a Negative Binomial model. The counts are regressed on an indicator variable for whether the citing foreign ethnicity and cited US ethnicity are the same, as well as vectors of fixed effects for each of the four dimensions on which cells are formed. These fixed effects remove basic levels differences between the series (e.g., English in the US receiving uniformly more citations, Vietnamese researchers abroad making uniformly fewer inventions and citations). An indicator variable is also included for whether the cited and citing technology categories are the same.

The coefficient on the indicator variable for same-ethnicity is transformed into an incidence rate ratio that gives the higher rate of citations within an ethnic group. The incidence rate ratio for all citations is 1.496 with a standard error of 0.052. This coefficient is statistically different from one, the level where own-ethnicity citations have the same frequency as citations of other ethnicities, and suggests a moderate effect that own-ethnicity citations are 50% higher than citations to other ethnicities once the basic levels and industry effects are removed. This ethnic differential is a couple of orders of magnitude less than the within-technology field effect, and Kerr (2007a) shows that tighter technology controls by disaggregating the sample can weaken the own-ethnicity differential to 20%-30%. The tighter specifications, however, remain economically and statistically important. To further study the time path of these knowledge flows, the Negative Binomial regressions are performed separately for each citation lag of one to ten years, rather than collapsing the data into a single regression. The coefficients from these regressions and their confidence bands (two standard deviations) are presented graphically in Figure 2. Common ethnicity appears most important for international technology diffusion in the first few years after an invention, peaking in a citation lag of four to five years.[10]

3.3 Codified and Tacit Knowledge Transfer

The international patent citation exercises confirm that knowledge diffusion occurs at an uneven rate across countries and further suggest that ethnic scientific networks are important for short-run technology transfer from the US. The declining importance to common ethnic ties over time in many respects resembles the declining importance of geographic distance in knowledge diffusion over time (e.g., Keller 2002b). While the citation regressions explicitly measure inventor-to-inventor knowledge flows, the short-term differentials are more generally representative of the transfer of codified and tacit knowledge.

The heightened transfer of codified knowledge can arise from several factors. Most sim-

[10] Kerr (2007a) tabulates these Negative Binomial regression results. This unpublished appendix also contains results using the dataset and techniques developed by Thompson (2006). After assigning ethnicities to inventors in Thompson's dataset, estimations using Thompson's technique yield a quantitatively similar role for own-ethnicity in international citations of 40%-60% depending upon the specification.

ply, ethnic networks aid awareness of new technologies that are developed. Even with modern communications advances, information continues to diffuse through professional networks and word-of-mouth. Second, ethnic business networks can aid trust and informal contracts where traditional legal enforcement is uncertain. Ethnic diasporas have facilitated trade flows for centuries (e.g., Rauch 2001, Rauch and Trindade 2002), and frontier ethnic expatriates can serve as reputation intermediaries for the transfer of new technologies, too. These transfers are understandably cautious given weaker international intellectual property protections. Kapur (2001) notes that US ethnic scientists and entrepreneurs are likely to play a greater role as reputation intermediaries in industries where tacit knowledge is important with respect to quality. US Indian entrepreneurs have substantially enhanced the brand reputation of India's programmers.

The transfer of the practical knowledge necessary for using or adapting new innovations is also aided by frontier expatriates. This tacit knowledge applies to both the specific technologies developed and the broader context of innovation. Often times, the technology diffusion encouraged by cross-border ethnic transfers encourages the formation of new firms seeking to integrate into industries characterized by decentralized production and cross-firm collaborations. In these environments, informal knowledge regarding component integration and the industry's future direction are critical; these insights can moreover illuminate pitfalls to avoid. The importance of this tacit knowledge cannot be overestimated. Lester and Piore (2004) describe how a Japanese communications equipment manufacturer withdrew from the US market after being excluded from standards hearings held by the Federal Communications Commission (FCC), despite the fact that the FCC published the transcripts of its sessions! The Japanese vendor felt it would not understand adequately the unspoken or implicit decisions being made.[11]

Finally, this study has interesting parallels to two recent papers regarding knowledge diffusion through Indian entrepreneurial and scientific networks. In a study of India's software industry, Nanda and Khanna (2006) find that entrepreneurs outside of software hubs rely more on the Indian diaspora than those working within centers like Bangalore. These findings suggest that diaspora networks may serve as substitutes for local institutions and technology opportunities. Looking within a single economy, Agrawal et al. (2007) jointly examine knowledge diffusion through co-location and co-ethnicity using domestic patent citations made by Indian inventors living in the US. While being in the same city or the same ethnicity both encourage knowledge diffusion, their estimations suggest that the marginal benefit of co-location is four times larger for inventors of different ethnicities. This substitutability between social and geographic proximity can create differences between a social planner's optimal distribution of ethnic members, and what the inventors themselves would choose.[12]

[11] A second intuitive example is the construction of an atomic bomb. While the basic designs are available on the internet, efforts to stem nuclear weapons proliferation focus extensively on the scientists with the tacit knowledge necessary for implementation. Other examples are drawn from Amsden (2001), Feinstein and Howe (1997), Kim (1997), Lim (1999), and Saxenian (2006). Polanyi (1958, 1966) introduces tacit knowledge; Granovetter (1973) highlights the strength of weak ties.

[12] Technology diffusion is also facilitated by foreign direct investment and multinational enterprises (e.g.,

4 Output and Productivity Analysis

This section turns to the next question of whether this greater transfer of knowledge for US innovations through ethnic networks produces measurable economic improvements for foreign countries. The US ethnic patenting trends are joined with data on foreign manufacturing industries, and an empirical extension of specification (6) is developed and estimated.

4.1 Foreign Manufacturing Data

The benefit of knowledge integration for foreign development is evaluated through the Industrial Statistics Database of the United Nations Industrial Development Organization (UNIDO). The UNIDO collects industry-level manufacturing statistics for *The International Yearbook of Industrial Statistics* and specialized publications on topics like development and competition. Researchers at the UNIDO supplement the data resources of the OECD with national records for non-OECD members, creating a unique global resource. The UNIDO's stated objective is the compilation of internationally comparable and internally consistent series (e.g., variable definitions, accounting units, collection procedures).

Table 2 describes the sample and lists the three-digit ISIC industries. The panels include all country-industry observations surveyed at least four times from 1985-1997 that correspond to non-English ethnicities identifiable with the ethnic-name database (e.g., Canada, the United Kingdom, Africa, and the Middle East are excluded). Three industry characteristics are considered: output, employment, and labor productivity measured as output per employee. Table 2 aggregates the annual industry-level data to describe the country-level manufacturing sectors. While direct comparisons across countries are limited with an unbalanced panel, the output and labor productivity differences between industrialized countries (e.g., Japan) and developing nations are clearly evident. The underlying industry-level metrics also agree with published UNIDO and World Bank statistics.[13]

4.2 Output and Productivity Estimation Framework

The combined dataset affords an industry-level analysis of technology transfer with multiple countries and ethnicities. Extending (6) to industry i and country c of ethnicity e,

$$\ln(Y_{ci}) = \phi_{ci} + \ln(\tilde{I}_i) + \beta \ln(\tilde{L}_{R,ei}), \tag{8}$$

where $\tilde{L}_{R,ei}$ is the size of the US research community of ethnicity e in industry i. While analytically convenient, this steady-state description must be adapted for the empirical exercises.

Branstetter 2006, Singh 2004). Foley and Kerr (2007) find growth in the US-based ethnic researchers within US multinational firms is correlated with larger FDI into countries of the researchers' ethnicity. Moreover, the organizational form of the FDI shifts towards more direct entry versus joint ventures.

[13] Kerr (2007a) documents additional descriptive statistics for this sample, the dataset development process, and alternative UNIDO panels considered. The appendix also describes the mapping of USPTO technology classifications to ISIC industries, building on Johnson (1999) and Silverman (1999).

The ethnic human-capital stocks for US technologies change over the 1985-1997 period — the source of identification for the β parameter. The citation regressions in Figure 2 highlight that ethnic ties have an important lag structure, especially for the first five years of knowledge dissemination. Rewriting (8) in discrete time to model this five-year dependency,

$$\ln(Y_{cit}) = \phi_{ci} + \ln(\tilde{I}_{it}) + \beta \ln\left(\sum_{s=1}^{5} \tilde{L}_{R,ei,t-s}\right). \tag{9}$$

Ethnic patenting data provide an empirical foothold for estimating these US ethnic scientific research communities. Rewriting the US researcher productivity function into a discrete-time form for industry i and ethnicity e, $\tilde{I}_{eit}^{Flow} = \tilde{I}_{it} \cdot \tilde{L}_{R,eit}$. The measured patenting of ethnicity e in year t again depends upon the overall stock of US knowledge and the size of the ethnic research group in the US (measured at the beginning of the year). By abstracting from the endogenous growth stimulus, the researcher productivity becomes time-invariant: $\tilde{I}_{it} = \tilde{I}_{it_0}$. Thus, the US ethnic research community can be inferred from the patent flow divided by the constant researcher productivity ($\tilde{L}_{R,eit} = \tilde{I}_{it_0}^{-1} \cdot \tilde{I}_{eit}^{Flow}$). Substituting this simplified form into (9), the time-invariant researcher productivity $\tilde{I}_{it_0}^{-1}$ is separated from the patent sum and incorporated with $\ln(\tilde{I}_{it})$ into an industry-year fixed effect η_{it}. Likewise, the base productivity constants ϕ_{ci} are extended into country-industry fixed effects.

To keep the exposition simple, define PAT_{eit}^{US} to be the five-year sum of recent US ethnic patenting in an industry. The core estimating equation becomes

$$\ln(Y_{cit}) = \alpha + \beta \ln(PAT_{eit}^{US}) + \phi_{ci} + \eta_{it} + \epsilon_{cit}, \tag{10}$$

where ϕ_{ci} and η_{it} are the vectors of country-industry and industry-year fixed effects, respectively. These fixed effects warrant careful discussion. First, the country-industry effects ϕ_{ci} remove levels differences between series. Without ϕ_{ci}, a positive β would be found if output in China's computer industry and US Chinese research in the computer industry are higher than average. Incorporating ϕ_{ci} instead requires the output growth in China's computer industry be above average if the US Chinese computer research growth is above average. Focusing on relative growth rates removes time-invariant factors that potentially confound the analysis (e.g., the productivity parameters A, ethnicity size).

The derivation of (10) highlights two important roles for the industry-year fixed effects η_{it}. First, η_{it} extract the overall growth in the US knowledge stock for an industry (e.g., the strong increase in computer and pharmaceutical research vis-à-vis mechanical research). Second, η_{it} control for the invention productivity of researchers, so that ethnic patenting flows are viable proxies for ethnic research in the US. More generally, the industry-year effects remove all industry-level trends common to the countries (e.g., demand shifts, price changes) and fluctuations in patent statistics due to changes in USPTO resources (e.g., Griliches 1990).

These fixed effects are crucial for the interpretation of the β parameter. This project does *not* estimate the effect of US patenting on foreign output and productivity; indeed, isolating that specific channel from other knowledge flows between countries is not feasible with industry-level outcomes. Moreover, the substantial increase in the number of patents granted by the USPTO over the last two decades is difficult to interpret. Instead, (10) forces variation to be within industries, isolating the size of ethnic communities from aggregate industry trends. A positive β coefficient requires that higher relative growth of Chinese computer research compared to Indian computer research in the US correlate with higher relative output growth in China's computer industry compared to India's computer industry.

4.3 Ethnic Patenting Estimator

The five-year patent sums PAT_{eit}^{US} are developed for each ethnicity-industry from the patent database. The matched USPTO records describe the ethnic composition of US scientists and engineers with previously unavailable detail: incorporating the major ethnicities working in the US scientific community; separating out detailed technologies and manufacturing industries; and providing annual metrics. The panel econometrics (10) require this level of cross-sectional and longitudinal variation to estimate general elasticities. The procedure does, however, have three potential limitations that should be discussed before presenting the results.

First, the approach does not distinguish foreign-born ethnic researchers in the US from later generations working as scientists and engineers, especially for the European contribution. While research on social and business networks finds the strength of ties to home countries declines for later generations, the ethnic patenting approach can only estimate total ethnic scientific populations. The panel econometrics employed for the output and productivity analyses, however, identify off of relative changes in these community sizes. Census and INS records confirm these changes are primarily due to new immigration for the period studied, substantially weakening this overall concern. Moreover, the immigration reform exercises in Section 5 yield similar results when focusing specifically on new arrivals through US quotas changes.

On a related topic, recent surveys of ethnic technology transfer from the US to China and India suggests technical exchanges are particularly aided by the circular labor movements of US-trained researchers and entrepreneurs (e.g., Saxenian 2006, Nanda and Khanna 2006). The ethnic patenting technique cannot quantify the magnitudes of reverse migration and circular migration flows, instead being restricted to net growth in US ethnic researcher populations. In this metric's defense, the scientific integration it captures likely embodies circular flows too, and the 1985-1997 period pre-dates most large-scale return migration decisions. If anything, the extent to which return migrations are important should lead to finding a negative β coefficient in the estimations. Return migration and circular movements are rapidly growing in importance, however, and it is hoped that future research will illuminate these issues further.

Finally, the name-matching technique does not distinguish finer ethnic and linguistic divisions within the nine major ethnic groupings. It would be advantageous to separate Mexican from Chilean scientists within the Hispanic ethnicity, to distinguish Chinese engineers with ties to Taipei versus Beijing versus Shanghai, and so on. These distinctions are not possible for this study's large scale analysis, and several countries will map into the Chinese, European, and Hispanic ethnicities for the output and productivity analyses below. The empirical analysis accounts for this multiplicity by conservatively clustering standard errors at the ethnicity-industry level; this cross-sectional clustering further addresses the serial-correlation concerns of Bertrand et al. (2004). Despite the clustering, measurement error from the broader ethnic divisions may still bias the estimated coefficients downward. The positive elasticities evident will nevertheless support the conclusion that technology following countries experience economic growth due to stronger technology transfer from the US.

4.4 Basic Output and Productivity Regressions

As a final preparation step, the levels specification (10) is first differenced for estimation,

$$\Delta \ln(Y_{cit}) = \alpha + \beta \Delta \ln(PAT_{eit}^{US}) + \eta_{it} + \hat{\epsilon}_{cit}, \tag{11}$$

where $\hat{\epsilon}_{cit} = \epsilon_{cit} - \epsilon_{cit-1}$.[14] Table 3 reports the primary results. The top row finds that output rises with strong scientific integration to the US. As both variables are in logs, the 0.091 coefficient in the upper-left corner finds a 10% increase in US ethnic research is associated with a 1% increase in foreign output. Industry output expansion can come through both labor productivity gains and expansion in employment. Disaggregating the output regression, Panels B and C find labor productivity growth facilitates most of the manufacturing development captured in this sample.

Three weighting schemes are tested: no weights, weighted by the 1985-1987 industry-level patenting in the US, and weighted by the 1985-1987 size of the foreign manufacturing industry. The β coefficients in the weighted regressions are larger than the unweighted specification, measuring an output elasticity of approximately 0.3. The patent weights emphasizes high-tech industries and the strong interactions of the Chinese and Indian research communities with their home countries. The output weights instead focus on the largest industries and offer a sense of the average treatment effect for industries. Coefficient estimates tend to be marginally smaller with the output weights than the patent weights due to the output weights' greater emphasis on traditional economic sectors (e.g., food products, textiles). Both approaches, however, yield more consistent results than the unweighted regressions by focusing attention on larger countries and industries and reducing measurement error in the ethnic patenting estimator. The weighted estimations are the preferred specifications of this study.[15]

[14] The efficiency of this first-differences form versus the levels specification turns on whether the error term ϵ_{cit} is autoregressive. If autoregressive deviations are substantial, the first-differences form is preferred; a unit-root error is fully corrected. If there is no serial correlation, however, first differencing introduces a moving-average error component. Estimations of the autoregressive parameter in the levels specification for this study find serial correlations of 0.5-0.6, while -0.1 is evident in the first-differences form.

[15] The elasticities are larger and more uniform in the levels estimation (10). The unweighted output elasticity is 0.241 (0.126), while the patent and output weighted elasticities are 0.420 (0.228) and 0.400 (0.147), respectively. Kerr (2007a) documents equivalent results using the levels specifications for all of the tables presented below.

The basic estimations reported in Table 3 are consistent with technology following countries realizing economic gains from stronger scientific integration with the US. These benefits appear to extend beyond the inventor-to-inventor flows evident with the citations analysis, as these US ethnic research communities facilitate broader manufacturing output growth through superior access to the US technology set. The remainder of this section further tests this finding by incorporating country-level controls, examining sector reallocations, and so on.[16]

4.5 Foreign Country Development Controls

The industry-year fixed effects create an empirical environment where US ethnic patenting serves as a viable metric for the strength of ethnic research communities. Moreover, the focus on within-industry variation circumvents many problems in interpretation that could arise from different industry trends (e.g., rapid high-tech growth). As the constructed panel includes multiple industries within a country, additional tests can be performed that further control for country-wide development. Table 4 undertakes four such tests, with Panel A simply replicating the base first-differences regressions for foreign output from Table 3.

An immediate concern is whether the results are capturing only foreign human-capital development, which could reasonably lead to an expansion in foreign manufacturing and the emigration of researchers to the US. The NSF collects annual data on the US Ph.D. science and engineering graduates by country-of-birth. As an initial robustness check on the general human-capital development story, Panel B adds the log trend in these graduates as an additional covariate. The role of the US ethnic scientific community remains strong and significant. (Section 5's immigration analysis returns to these Ph.D. trends and the reverse causality concern.)

Panel C next explores the role of physical-capital development in explaining the output growth. Section 2's theory only models non-durable intermediate inputs, a simplification that removes the need to track two state variables. Labor productivity and output growth occurs with capital deepening as well as technology adoption, however. This investment in physical machinery and structures is clearly endogenous to technology transfer from the frontier economy, due to both the larger available technology set and the general equilibrium economic development experienced. Nevertheless, additional confidence for the role of frontier scientific communities can be established through joint tests with this factor input. The output-weighted coefficient retains most of its economic magnitude and statistical strength; the patent-weighted coefficient retains 80% of its original economic importance but is no longer statistically significant.[17]

[16] The differential technology transfer explains 1%-2% of the sample's output and productivity growth variation after removing aggregate industry trends. The percentage accounted for rises to 3%-5% with the sector reallocation specifications studied below. These percentages provide order-of-magnitude estimates for the total growth accounted for, although calculations after removing industry-year effects likely understate the total impact due to technology transfer. These technology gains in turn produce comparative advantages for trade (Kerr 2007b).

[17] The UNIDO data unfortunately lack capital records for almost half of the sample. Moreover, the available capital stocks are measured with substantial error, downward biasing the capital coefficients. Kerr (2007a) details the construction of the capital stocks and provides additional tests. These results can be extended to include labor in a production function estimation, although the employment response is even more endogenous

More generally, Panels D and E incorporate into (11) linear country time trends and non-parametric country-year fixed effects, respectively. These additional controls remove trends common to the industries within a country, including the overall growth in each ethnicity's US research community (e.g., the strong increases in Chinese and Indian patenting in the US). For foreign output, the country effects extract national business cycles, trend manufacturing gains, trade agreements, and so on. A positive β coefficient in these estimations requires higher relative growth of Chinese computer research to Chinese pharmaceutical research in the US be partially correlated with higher relative output growth in China's computer industry to its pharmaceutical industry (after worldwide industry trends are removed).

The inclusion of both country-year and industry-year fixed effects in a first-differenced specification is a very stringent test, and much of the variation is removed from the sample. While the positive correlations are preserved in three of the four weighted regressions, only one coefficient is statistically significant. Moreover, the correlations are zero or negative in the unweighted specifications. These declines in coefficient magnitudes are partly explained by the relatively uniform growth (versus levels) in each ethnicity's US research communities across industries in a log expansion. That is, much greater variation exists across ethnicities than across industries within an ethnicity (Kerr 2006c). To the extent that this uniform growth is what is being captured by the country-year fixed effects, the core estimations correctly measure the general elasticity. This study cannot reject, however, that the base elasticities are upward biased due to presence of an omitted variable operating at the country-year level too.

4.6 Sample Decompositions

The core objective of these empirical exercises is quantifying the mean output gains from US technology transfer through ethnic networks across a diverse group of countries and industries. It is informative, however, to identify which observations are most responsible for the aggregate findings. Table 5 investigates this question through several sample decompositions.

Case studies of successful technology diffusion often focus on the computer and pharmaceutical industries, and the exceptional outcomes of Asian scientific communities in Silicon Valley are widely noted. While the industry-year effects control for the overall growth in each industry's research and output (e.g., Griliches 1994), ethnic differences in high-tech industries alone could still be responsible for the positive correlations. To some extent, the stronger coefficients in the patent-weighted regressions suggest this is true, and Panel B begins by directly excluding the computer and pharmaceutical industries from the sample. The results are mixed. On one hand, both the unweighted and patent-weighted coefficients decline substantially in economic magnitude. On the other hand, the patent-weighted coefficient does remain statistically significant and the output-weighted elasticity is broadly robust. These mixed results suggest the gains are concentrated in high-tech sectors, but that they are not entirely exclusive to them.

to technology transfer as discussed below.

Chinese economies, more often than not, are also the centerpieces of technology transfer stories. The US Chinese research community experiences strong growth during the sample period, and Mainland China has exceptional manufacturing gains too. When excluding Mainland China in Panel C, the unweighted elasticity loses a third of its magnitude and its statistical significance, but the weighted regressions deliver fairly similar results. Unreported regressions further find that the weighted parameter estimates do not depend significantly on the inclusion of any one country in the sample. Panel D of Table 5 demonstrates, however, that excluding the full Chinese ethnicity can be important even for the weighted estimations. Given that the Chinese grouping includes three of the four Asian "tiger" economies (i.e., Hong Kong, Singapore, and Taiwan) and Mainland China, it is not too surprising that the effect is sensitive to their inclusion. Further tests find that the decline in the coefficient size is mostly linked to dropping the computer and drug industries for the Chinese economies. This cautions that the well-documented outcomes for Silicon Valley are in some sense special even for the Chinese, with the benefits of scientific collaboration for manufacturing being weaker in most other contexts.[18]

The UNIDO sample also includes several industrialized economies that are undertaking extensive R&D themselves. For example, Japanese inventors living in the US, who are well identified with the ethnic-name database, patented less than 10,000 inventions from 1985-1997; almost 300,000 USPTO patents were awarded to Japanese inventors living outside of the US during this period.[19] Positive correlations of foreign country growth to US ethnic research may simply be capturing reverse technology flows, intra-company patenting, or defensive patenting from these advanced economies. Exploring this issue, Panel E excludes Japan, European countries, and Russia and finds similar results to the full sample. Likewise, the last row drops the large bloc of Hispanic countries and finds similar coefficients in the weighted regressions.

To summarize, the unweighted elasticities are clearly sensitive to the sample composition, while the weighted elasticities are more robust across sample compositions. The ethnic technology transfer mechanism is especially strong for high-tech and Chinese outcomes, reflective of the disproportionate number of case studies written. The weighted specifications suggest, however, that some transfer benefits extend beyond these special outcomes to other ethnicities and more traditional industries. The next section refines the main effects to characterize further differences in outcomes by development stage.

4.7 Sector Reallocation

Section 2's theoretical framework builds on the assumption of full employment in the technology follower's manufacturing and research sectors. While the estimating equation (6) relates the follower's output to its research presence in the leader, the same elasticity β would hold for

[18] Dropping only the computer and drug industries for Chinese economies yields coefficients slightly larger than those in Panel D that are statistically significant in the two weighted regressions.

[19] The estimates are sums over inventor ethnicity percentages at the patent level. Japanese inventors are associated with more patents due to multiple inventors.

labor productivity specifications. With full employment, output gains can only come through labor productivity enhancements. Many developing economies have large agricultural sectors, however, and the migration from agriculture to manufacturing is important for characterizing economic development (e.g., Harris and Todaro 1970).

Kerr (2007a) incorporates into the basic model an agricultural sector in the follower. In this extension, technology transfer from the leader to the follower induces sector reallocation, with labor shifting from agriculture to the manufacturing and research sectors. Thus, output growth occurs through both labor productivity gains, as in the steady-state scenario, and through employment growth along the transition path. After a sufficient number of frontier innovations are imitated, the follower's economy transitions to full employment in the manufacturing and research sectors. Thus, the steady-state of the expanded economy is the same as the basic framework described in Section 2.[20]

To test these transition path predictions for developing economies, Table 2 lists the 1980 share of national employment in agriculture for each economy. The three smallest agricultural sectors are found in Hong Kong (1%), Singapore (2%), and Belgium (3%), while the three largest sectors are India (70%), Vietnam (73%), and Mainland China (74%). A modified form of (11) interacts the ethnic scientific community regressor with this pre-period agricultural share,

$$\Delta \ln(Y_{cit}) = \alpha + \beta\Delta \ln(PAT_{eit}^{US}) + \gamma\Delta \ln(PAT_{eit}^{US}) \cdot AGR\%_{c,1980} + \eta_{it} + \hat{\epsilon}_{cit},$$

where the main effect for the agricultural share is absorbed into the first differencing. The main effects are demeaned prior to the interaction to restore the β coefficient to close to its base level. A positive γ coefficient indicates output growth due to scientific integration is stronger in countries with larger agricultural workforces in 1980.

Table 6 reports the results from these interacted regressions. Foreign country output growth due to stronger US ethnic research integration is higher in economies with large agricultural shares in 1980. Panels B and C again disaggregate the output regression into labor productivity and employment shifts, respectively. Labor productivity gains are weaker in the less developed economies, though the difference is usually not statistically significant, while substantial sector reallocation through employment growth is clearly evident in Panel C. The interacted regressions thus support the model's predictions regarding the stage of development being important for how technology transfer gains are realized. Economies with large agricultural sectors facilitate employment reallocation across sectors that aid manufacturing output expansion.[21]

[20] The output gains through labor productivity and employment growth are of similar magnitude in the numerical simulations Kerr (2007a) models. In alternative models, output growth would come only through labor reallocation (e.g., fixed physical capital stocks and constant outside wages).

[21] These sector reallocation findings are robust to the earlier sample decompositions. Notably, the interactions are more robust than the main effects to dropping high-tech industries and the Chinese ethnicity (Kerr 2007a).

5 Exogenous Changes from US Immigration Reforms

While OLS regressions establish partial correlations present in the data, they frequently fail to identify causal relationships due to the endogenous relationships between outcomes or due to omitted variable biases. Domestic human-capital developments in Chinese economies, for example, could lead to both higher productivity and output growth at home and the export of scientists to the US. Alternatively, R&D in Japan might be responsible for the growth of its Asian neighbors and feed into higher US research output. Despite the strong fixed-effect specifications employed, further exercises can aid in the interpretation of the positive outcomes evident in patent-based regressions.

The earlier model helps understand and address these concerns. Consider the initial transition from the equilibrium described in Section 2 following an industrialization push in the follower. The follower's government temporarily subsidizes invention until condition (7) no longer holds. As $I > \tilde{I}\Psi[M/\tilde{I}](\tilde{H}^F)^\beta$, it is more profitable for researchers in the follower to invent rather than imitate; the follower's output growth and sector reallocation are now driven solely by domestic innovations. In the leader, researchers of the follower's ethnicity switch from inventing to imitating, as the latter is initially very easy (i.e., $\Psi[0]$ is high). If international property rights are weak, so that ethnic researchers in the leader can register their imitations with the leader's patent office, a positive β coefficient will be found in the core estimating equations even though the follower's manufacturing gains no longer depend on its frontier research community. In fact, data trends will show contemporaneous accelerations in the growth of foreign output and the leader's ethnic patenting.[22]

The leader's population of the follower's ethnic researchers is a foothold for establishing greater confidence in the direction of technology flows as the expatriates only influence the follower's development through their transmission of knowledge regarding frontier innovations. If the size of this research population is exogenously determined by immigration restrictions, a reduced-form strategy for the size of the ethnic research community can be developed within the quotas system. In this paper's context, US immigration law does not control the population size of foreigners in the US, but it does control the inflow of new immigrants. Define the quota on follower's inflows of researchers to the US to be $QUOTA_{R_F,t}$. Assuming that only the previous three years of immigration matter for a research stock[23], a reduced-form immigration estimator for ethnic scientific integration to the US is modelled as

$$\ln(IMM^{RF}_{R_F,t}) = \ln\left[\sum_{s=1}^{5}(QUOTA_{R_F,t-s} + QUOTA_{R_F,t-s-1} + QUOTA_{R_F,t-s-2})\right]. \qquad (12)$$

The summation over the previous five years maintains the human-capital stock modelling technique employed with the ethnic patenting dataset. This section designs and implements an

[22] The follower's economy still depends on previously imitated products, as well as new inventions. Kerr (2007a) further discusses the transitions following this disturbance.

[23] The reform below produced a very sharp immigration surge that makes this assumption more reasonable.

empirical version of (12) using exogenous changes in US immigration quotas.

5.1 The Immigration Act of 1990

The disproportionate influence of immigrant scientists and engineers (ISEs) in the US is staggering: while immigrants account for 10% of the US working population, they represent 25% of the US science and engineering workforce and 50% of those with doctorates. Even looking within the Ph.D. level, immigrant researchers have an exceptional contribution to science as measured by Nobel Prizes, election to the National Academy of Sciences, patent citation counts, and so on.[24] Yet, the US immigration system significantly restricted the inflow of ISEs from certain nations prior to its reform with the Immigration Act of 1990 (1990 Act).

Immigrants can obtain permanent residency in the US through numerically unrestricted categories (e.g., immediate family members) or numerically restricted categories (e.g., extended family members, employment based applications). The immigration exercises focus on the numerically restricted categories that admit 75% of ISEs, versus 43% of all immigrants. US immigration law applies two distinct quotas within these restricted categories. Both of these quotas were increased by the 1990 Act, and their combined change dramatically released pent-up immigration demand from researchers in constrained countries.

The first quota governs the annual number of immigrants admitted per country. This quota is uniform across nations, and the 1990 Act increased the limit from 20,000 to approximately 25,620. Larger nations are more constrained by country quotas than smaller nations and benefited most from these higher admission rates. Second, separately applied quotas govern the relative admissions of family-based versus employment-based immigrants. Prior to the 1990 Act, the quotas substantially favored family-reunification applications (216,000) to employment applications (54,000). The 1990 Act shifted this priority structure by raising employment-based immigration to 120,120 (20% to 36% of the total) and reducing family-based admissions to 196,000. Moreover, the relative admissions of high-skilled professionals to low-skilled workers significantly increased within the employment-based admissions.[25]

The uniform country quotas and weak employment preferences constrained high-skilled immigration from large nations, and long waiting lists for Chinese, Indian, and Filipino applicants formed in the 1980s. When the 1990 Act simultaneously raised both of these quotas, the number of ISEs entering the US dramatically increased. Figure 3 uses records from the Immigration and Naturalization Service (INS) to detail the response. This graph plots the number of ISEs granted permanent residency in the US from 1983-1997 for selected ethnicities (summed over

[24] For example, Stephan and Levin (2001), Burton and Wang (1999), Johnson (1998, 2001), and Streeter (1997).

[25] Kerr (2007a) describes the 1990 Act in greater detail and discusses ISE immigration through temporary visas and numerically unrestricted categories. This supplement further catalogues the construction of the INS data employed in this section. The worldwide ceiling for numerically restricted immigration now fluctuates slightly year-to-year based on past levels; maximum immigration from a single country is limited to 7% of the worldwide ceiling. The employment limit increased to 140,000, but 120,120 corresponds to the previously restricted categories. Jasso et al. (2000) also discuss behavioral responses to the 1990 Act.

countries within each ethnicity). Prior to the 1990 Act, no trends are evident in ISE immigration. The 1990 Act took effect in October 1991, and a small increase occurred in the final three months of 1991 for Chinese and Indian ISEs. Immigration further surged in 1992-1995 as the pent-up demand was released. Low-skilled immigration did not respond to the 1990 Act.

The extremely large Chinese response and sharp decline is partly due to a second law that slightly modified the timing of the 1990 Act's reforms. Following the Tiananmen Square crisis in June 1989, Chinese students present in the US from the time of the crisis until May 1990 were permitted to remain in the US until at least 1994 if they so desired. The Chinese Student Protection Act (CSPA), signed in 1992, further granted this cohort the option to change from temporary to permanent status during a one-year period lasting from July 1993 to July 1994. The CSPA stipulated, however, that excess immigration from the CSPA over Mainland China's numerical limit be deducted from later admissions. The timing of the CSPA partly explains the 1993 spike, and the ability of graduating Chinese science and engineering students to remain in the US in 1990 should factor into the timing of the reduced-form estimator.

Finally, NSF surveys of graduating science and engineering doctoral students — the group most important for developing human capital with respect to US innovations — confirm the strong responses evident in the INS data. The questionnaires ask foreign-born Ph.D. students in their final year of US study about their plans after graduation. Figure 4 exhibits the percentage intending to remain in the US after graduation for available countries. The 60% to 90% jump for Mainland China from 1990 to 1992 is striking. Substantial increases are also apparent for India and Western Europe.

5.2 Immigration Responses

The reduced-form strategy exploits differences in the extent to which countries were affected by the 1990 reform. It is inappropriate, however, to use the outcomes exhibited in Figures 3-4 to determine treatment and control groups. A proper designation of the affected countries requires a more formal analysis of researcher immigration responses to the legislation change. Kerr (2007a) undertakes such an analysis and further characterizes immigration waiting lists around the time of the reform. From this analysis, the treated groups are determined to be India, Mainland China, the Philippines, and Taiwan. The reduced-form immigration estimator (12) then takes the form

$$\ln(IMM^{RF}_{cit}) = \ln\left[\sum_{s=1}^{5}(QUOTA^{Eff}_{c,t-s} + QUOTA^{Eff}_{c,t-s-1} + QUOTA^{Eff}_{c,t-s-2})\right], \qquad (13)$$

where $QUOTA^{Eff}_{ct}$ is the effective quota for country c in year t. Raising the numerical ceilings did not change the effective quotas for nations unconstrained by the former immigration regime, and their effective quotas are held constant at the pre-reform theoretical limit. For constrained countries, the effective quota increases to reflect both the higher country limit of 25,600 and the larger employment preference allocation of 36% (i.e., 120,120/336,000). This quota increase occurs in 1991 and is moved forward to 1990 for Mainland China to account for the CSPA.

This simple reduced-form approach abstracts from several issues: return migration (e.g., Taiwanese scientists in the mid 1990s), occupational or industry changes by ISEs, second-generation immigrant demographics, shifts in researcher productivity, and others. If these types of concerns are overwhelming, panel regressions of US ethnic patenting on the reduced-form estimator will yield weak coefficients. Unreported regressions find this relationship is quite strong, however, despite the design's simplicity. However, two more serious reservations regarding the estimator should be addressed before viewing the results.

First, the quota change affected all skilled workers seeking admission into the US, not just researchers, and the impact of other occupations should be considered. The reduced-form estimator should only influence foreign manufacturing output and productivity through the development of human capital with respect to US technologies. Most skilled occupations can be dismissed immediately, yet immigration of business executives and lawyers also increased after the 1990 Act. It is possible this business group might influence foreign output growth through better sales contacts or higher foreign investment independent of technology transfer. The relative volumes argue against this concern, as the size of the influx relative to the existing base for advanced-degree researchers dwarfs other occupations. The planned inflow of Chinese science and engineering Ph.D.s for 1991-1995, as measured by the NSF surveys, would have doubled the existing Chinese-born Ph.D. stock in the 1990 Census. The business inflow over this period is only about 20% of the 1990 stock.

A second liability is that the reduced-form estimator may be correlated with other factors. Here, the simplicity of its design is a concern. While determined by the data, the quotas technique only distinguishes between the treatment group (i.e., India, Mainland China, the Philippines, and Taiwan) and the remainder of the sample. Other changes occurring around 1991 that affect the output growth of the treatment group differentially from the control group could confound the analysis. As with any country-level change, possible confounding factors can be hypothesized for each treatment member. While the results are robust to excluding any one country from the treatment group, it is of course not possible to drop them all.

These concerns are why the US quotas are employed for a reduced-form estimator rather than in an instrumental-variables specification. Immigration quotas directly influence the size of ethnic research communities in the US, and thus the unobserved human-capital stocks. As such, these quotas offer a nice complement to and check on the earlier metrics derived from ethnic patents; moreover the quotas-based metric is more robust to reverse causality criticisms. The estimator does not resolve omitted variable concerns, however, and lacks industry-level variation that can be exploited. As the exclusion restriction for two-stage least squares would not hold, this study concentrates on the reduced-form outcomes.

5.3 Reduced-Form Results

The reduced-form regressions for 1985-1997 mirror the patent-based approach,

$$\Delta \ln(Y_{cit}) = \alpha + \beta\Delta \ln(IMM_{ct}^{RF}) + \eta_{it} + \hat{\epsilon}_{cit}, \tag{14}$$

with $\ln(IMM_{ct}^{RF})$ defined by (13). Table 7 exhibits the main results in a format similar to that of Table 3. The reduced-form estimator suggests foreign output increases with an elasticity of about 0.3 to higher ethnic research in the US. While the β coefficients should not be directly compared to the patent-based approach, the interpretation that greater scientific integration with the US boosts foreign manufacturing development is supported. The lower variance in Table 7's estimates across weights reflects the country-level design of the immigration estimator.

In contrast to the patent-based results, Panels B and C find output growth comes mainly through higher employment levels rather than labor productivity gains. This difference is easily explained with the sector reallocation model. Three of the four treated economies had large agricultural sectors in 1980 that supported significant expansions in employment; Taiwan is the one exception at 8%. The immigration estimator contrasts the outcomes in these economies with the control sample and thus emphasizes the sector reallocation process. The patent-based regressions, on the other hand, paid greater attention to the outcomes of Hong Kong, Macao, and Singapore through the application of the US Chinese ethnic patenting series to all economies within the Chinese ethnicity. Without an agricultural sector from which to draw labor, these economies experienced sharper labor productivity gains.

Table 8 next turns to robustness checks on the output growth finding. As a test of the foreign human-capital development story, Panel B again incorporates the log trends in foreign graduates from US science and engineering Ph.D. programs. The technology transfer coefficients hold up well in the augmented specification. Panel C finds that Mainland China can again be excluded from the sample with only minor shifts in the outcomes. The results are also robust to dropping any other country, the computer and drug industries, the full Chinese ethnicity, and the other sample decompositions studied above.[26] Panel D incorporates a linear ethnic time trend that removes the trend growth in both the foreign country output and the US immigration estimator. By doing so, the framework emphasizes the discontinuity of the 1990 Act for identification of the β parameter. The coefficients remain economically and statistically significant in this stringent specification, providing confidence against the estimator reflecting a spurious correlation.

The last two rows incorporate into (14) two counterfactual estimators that move the 1991 effective date of the immigration reform earlier to 1987 or later to 1995. The results with the 1987 counterfactual are mixed. Encouragingly, the coefficients on the true estimator retain most of their value and are still statistically different from zero. Moreover, the standard errors for the placebo estimators are 400% larger than those of the true estimator, and the placebo estimators

[26] These decompositions are more stable than those with the patent-based metric due to the country-level design of the estimator.

are not statistically significant. The coefficient estimates on the 1987 estimator, however, are of similar magnitude to the true reform, and it cannot be rejected that the coefficients are the same. Panel F, on the other hand, shows better performance with the 1995 counterfactual. These results support the conclusion of stronger scientific integration leading to foreign output growth, but also highlight that the estimated elasticity with the immigration estimator may be partly capturing an earlier differential change for the treatment group.

Establishing the causal direction of international technology flows is a very daunting task. The reduced-form quotas estimator offers more confidence than the patent-based approach that coefficient estimates are not determined by reverse causality (especially foreign human-capital developments). The price for this exogenous determinant, however, is the loss of industry variation that can be exploited. This reduced variation may leave the quotas estimator exposed to omitted variable biases contemporaneous to or slightly preceding the reform, although the multiple robustness checks suggest spurious correlations are not solely responsible for the outcomes measured. Overall, the reduced-form regressions support Section 4's conclusion that foreign manufacturing output increases with stronger ethnic scientific integration to the US frontier.

6 Conclusions

The international diffusion of new innovations from frontier countries is necessary for broad economic growth. Successful transfer, however, is complicated by the difficult dissemination of the codified and tacit knowledge necessary for adoption. This project considers the role and importance of knowledge networks for exchanging this information through the observable channel of ethnicity, examining the ties between US ethnic research and entrepreneurial communities and their home countries. The findings suggest that these frontier expatriates do play an important role in technology transfer, and more generally that inadequate access to the codified and tacit knowledge complementing new frontier innovations can slow development in following regions.

This study concentrates on estimating the general elasticities for technology transfer across multiple ethnicities and manufacturing industries. The platform, however, identifies particular strength for high-tech industries and Chinese communities. While still measurable, the responses are weaker elsewhere. Future work should investigate whether these patterns hold in other samples too. Chinese economies experienced exceptional manufacturing development during the 1985-1997 period. The extent to which these results extend to non-manufacturing sectors will shed light on whether the strong Chinese outcomes are due to the manufacturing focus, as all datasets have above-average outcomes, or due to unique qualities of this ethnicity's network (e.g., size and network effects). Likewise, characterizing portions of non-manufacturing sectors like financial and business services that are conducive to technology transfer will refine our understanding of the traits of industries (e.g., vertical integration, product cycles) where US ethnic scientists and entrepreneurs can most aid their home countries.

References

[1] Acemoglu, Daron, and Fabrizio Zilibotti, "Productivity Differences", *The Quarterly Journal of Economics,* 116:2 (2001), 563-606.

[2] Agrawal, Ajay, Iain Cockburne, and John McHale, "Gone But Not Forgotten: Knowledge Flows, Labor Mobility, and Enduring Social Relationships", *Journal of Economic Geography* 6 (2006), 571-591.

[3] Agrawal, Ajay, Devesh Kapur, and John McHale, "Birds of a Feather – Better Together? Exploring the Optimal Spatial Distribution of Ethnic Inventors", NBER Working Paper 12823 (2007).

[4] Amsden, Alice, *The Rise of "The Rest"* (Oxford, UK: Oxford University Press, 2001).

[5] Atkinson, Anthony, and Joseph Stiglitz, "A New View of Technological Change", *The Economic Journal* 79:315 (1969), 573-578.

[6] Banerjee, Abhijit, and Andrew Newman, "Occupational Choice and the Process of Development", *The Journal of Political Economy* 101:2 (1993), 274-298.

[7] Barro, Robert, and Xavier Sala-i-Martin, *Economic Growth* (Cambridge: MIT Press, 1995).

[8] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Difference in Differences Estimates?", *The Quarterly Journal of Economics* 119:1 (2004), 249-275.

[9] Borjas, George, "Do Foreign Students Crowd Out Native Students from Graduate Programs?", in Ehrenberg, Ronald, and Paula Stephan (ed.), *Science and the University* (Madison, WI: University of Wisconsin Press, 2005).

[10] Branstetter, Lee, "Is Foreign Direct Investment a Channel of Knowledge Spillovers? Evidence from Japan's FDI in the United States", *Journal of International Economics* 68 (2006), 325-344.

[11] Burton, Lawrence, and Jack Wang, "How Much Does the U.S. Rely on Immigrant Engineers?", NSF SRS Issue Brief (1999).

[12] Coe, David, and Elhanan Helpman, "International R&D Spillovers", *European Economic Review* 39:5 (1995), 859-887.

[13] Coe, David, Elhanan Helpman, and Alexander Hoffmaister, "North-South R&D Spillovers", *The Economic Journal* 107:440 (1997), 134-149.

[14] Duguet, Emmanuel, and Megan MacGarvie, "How Well Do Patent Citations Measure Flows of Technology? Evidence from French Innovation Surveys", *Economics of Innovation and New Technology* 14:5 (2005), 375-393.

[15] Eaton, Jonathan, and Samuel Kortum, "International Technology Diffusion: Theory and Measurement", *International Economic Review* 40:3 (1999), 537-570.

[16] Ellison, Glenn, Edward Glaeser, and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns", NBER Working Paper 13068 (2007).

[17] Feinstein, Charles, and Christopher Howe, ed., *Chinese Technology Transfer in the 1990s* (Cheltenham, UK: Edward Elgar, 1997).

[18] Foley, C. Fritz and William Kerr, "U.S. Ethnic Scientists and Foreign Direct Investment Placement", Working Paper (2007).

[19] Freeman, Richard, "Does Globalization of the Scientific/Engineering Workforce Threaten U.S. Economic Leadership?", NBER Working Paper 11457 (2005).

[20] Granovetter, Mark, "The Strength of Weak Ties", *American Journal of Sociology* 78:6 (1973), 1360-1380.

[21] Griliches, Zvi, "Patent Statistics as Economic Indicators: A Survey", *Journal of Economic Literature* 28:4 (1990), 1661-1707.

[22] Griliches, Zvi, "Productivity, R&D, and the Data Constraint", *The American Economic Review* 84:1 (1994), 1-23.

[23] Grossman, Gene, and Elhanan Helpman, *Innovation and Growth in the Global Economy* (Cambridge, MA: MIT Press, 1991).

[24] Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools", NBER Working Paper 8498 (2001).

[25] Harris, John, and Michael Todaro, "Migration, Unemployment, and Development: A Two-Sector Analysis", *The American Economic Review* 60:1 (1970), 126-142.

[26] Helpman, Elhanan, "R&D and Productivity: The International Connection", in Razin, Assaf, and Efraim Sadka, ed., *The Economics of Globalization* (Cambridge, UK: Cambridge University Press, 1999).

[27] Hu, Albert, and Adam Jaffe, "Patent Citations and International Knowledge Flow: The Cases of Korea and Taiwan", *International Journal of Industrial Organization* 21:6 (2004) 849-880.

[28] Jacobs, Jane, *The Economy of Cities* (New York, NY: Vintage Books, 1970).

[29] Jaffe, Adam, and Manuel Trajtenberg, "International Knowledge Flows: Evidence from Patent Citations", *Economics of Innovation and New Technology* 8 (1999), 105-136.

[30] Jaffe, Adam, Manuel Trajtenberg, and Michael Fogarty, "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors", *The American Economic Review* 90:2 (2000), 215-218.

[31] Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations", *The Quarterly Journal of Economics* 108:3 (1993), 577-598.

[32] Jasso, Guillermina, Mark Rosenzweig, and James Smith, "The Changing Skill of New Immigrants to the United States: Recent Trends and Their Determinants", in Borjas, George (ed.), *Issues in the Economics of Immigration* (Chicago, IL: University of Chicago, 2000).

[33] Johnson, Daniel, "150 Years of American Invention: Methodology and a First Geographic Application", Wellesley College Economics Working Paper 99-01 (1999).

[34] Johnson, Jean, "Statistical Profiles of Foreign Doctoral Recipients in Science and Engineering: Plans to Stay in the United States", NSF SRS Report (1998).

[35] Johnson, Jean, "Human Resource Contribution to U.S. Science and Engineering From China", NSF SRS Issue Brief (2001).

[36] Kapur, Devesh, "Diasporas and Technology Transfer", *Journal of Human Development* 2:2 (2001), 265-286.

[37] Kapur, Devesh, and John McHale, "Sojourns and Software: Internationally Mobile Human Capital and High-Tech Industry Development in India, Ireland, and Israel", in Arora, Ashish, and Alfonso Gambardella, *From Underdogs to Tigers: The Rise and Growth of the Software Industry in Some Emerging Economies* (Oxford, UK: Oxford University Press, 2005).

[38] Keller, Wolfgang, "Trade and the Transmission of Technology", *Journal of Economic Growth* 7:1 (2002a), 5-24.

[39] Keller, Wolfgang, "Geographic Localization of International Technology Diffusion", *The American Economic Review* 92:1 (2002b), 120-142.

[40] Keller, Wolfgang, "International Technology Diffusion", *Journal of Economic Literature* 42:3 (2004), 752-782.

[41] Kerr, William, "Appendix to Ethnic Scientific Communities and International Technology Diffusion", Working Paper (2007a).

[42] Kerr, William, "Heterogeneous Technology Diffusion and Ricardian Trade Patterns", Working Paper (2007b).

[43] Kerr, William, "The Ethnic Composition of US Inventors", Working Paper (2007c).

[44] Kim, Linsu, *Imitation to Innovation* (Boston, MA: Harvard Business School Press, 1997).

[45] Lester, Richard, and Michael Piore, *Innovation: The Missing Dimension* (Cambridge, MA: Harvard University Press, 2004).

[46] Lim, Youngil, *Technology and Productivity* (Cambridge, MA: MIT Press, 1999).

[47] MacGarvie, Megan, "Foreign Students and the Diffusion of Scientific and Technological Knowledge to and from American Universities", Working Paper (2006).

[48] Mankiw, N. Gregory, David Romer, and David Weil, "A Contribution to the Empirics of Economic Growth", *The Quarterly Journal of Economics*, 107:2 (1992), 407-437.

[49] Marshall, Alfred, *Principles of Economics* (London, UK: MacMillan and Co., 1890).

[50] Nanda, Ramana, and Tarun Khanna, "Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software Industry", Working Paper (2006).

[51] Nelson, Richard, and Edmund Phelps, "Investment in Humans, Technological Diffusion, and Economic Growth", *The American Economic Review* 56:1 (1966), 69-75.

[52] OECD, *Science and Technology Indicator Scoreboard* (2004).

[53] Parente, Stephen, and Edward Prescott, "Barriers to Technology Adoption and Development", *The Journal of Political Economy* 102:2 (1994), 298-321.

[54] Peri, Giovanni, "Determinants of Knowledge Flows and their Effect on Innovation", *Review of Economics and Statistics* 87:2 (2005), 308-322.

[55] Piore, Michael, "The Limits of the Division of Labor in Design and the Prospects for Off-Shore Software Development in Mexico", Working Paper (2004).

[56] Polanyi, Michael, *Personal Knowledge: Towards a Post-Critical Philosophy*, (Chicago, IL: University of Chicago Press, 1958).

[57] Polanyi, Michael, *The Tacit Dimension* (Garden City, NY: Doubleday & Company, 1966).

[58] Rauch, James, "Business and Social Networks in International Trade", *Journal of Economic Literature* 39:4 (2001), 1177-1203.

[59] Rauch, James, and Vitor Trindade, "Ethnic Chinese Networks in International Trade", *Review of Economics and Statistics* 84:1 (2002), 116-130.

[60] Riveria-Baitz, Luis, and Paul Romer, "Economic Integration and Endogenous Growth", *The Quarterly Journal of Economics* 106:2 (1991), 531–555.

[61] Romer, Paul, "Endogenous Technological Change", *The Journal of Political Economy* 98:5 (1990), S71-S102.

[62] Rosenthal, Stuart, and William Strange, "Geography, Industrial Organization, and Agglomeration", *Review of Economics and Statistics* 85:2 (2003), 377-393.

[63] Saxenian, AnnaLee, with Yasuyuki Motoyama and Xiaohong Quan, *Local and Global Networks of Immigrant Professionals in Silicon Valley* (San Francisco, CA: Public Policy Institute of California, 2002a).

[64] Saxenian, AnnaLee, "Silicon Valley's New Immigrant High-Growth Entrepreneurs", *Economic Development Quarterly* 16:1 (2002b), 20-31.

[65] Saxenian, AnnaLee, *The New Argonauts* (Cambridge, MA: Harvard University Press, 2006).

[66] Silverman, Brian, "Technological Resources and the Direction of Corporate Diversification: Toward an Integration of the Resource-Based View and Transaction Cost Economics", *Management Science* 45:8 (1999), 1109-1124.

[67] Singh, Jasjit, "Multinational Firms and International Knowledge Diffusion: Evidence from Patent Citation Data", *Best Paper Proceedings of 2004 Meeting of the Academy of Management* (2004).

[68] Stephan, Paula, and Sharon Levin, "Exceptional Contributions to US Science by the Foreign-Born and Foreign-Educated", *Population Research and Policy Review* 20:1 (2001), 59-79.

[69] Streeter, Joanne, "Major Declines in Admissions of Immigrant Scientists and Engineers in Fiscal Year 1994", NSF SRS Issue Brief (1997).

[70] Thompson, Peter, "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-Added Citations", *Review of Economics and Statistics* 88:2 (2006), 383-388.

Figure 1: US Ethnic Patenting

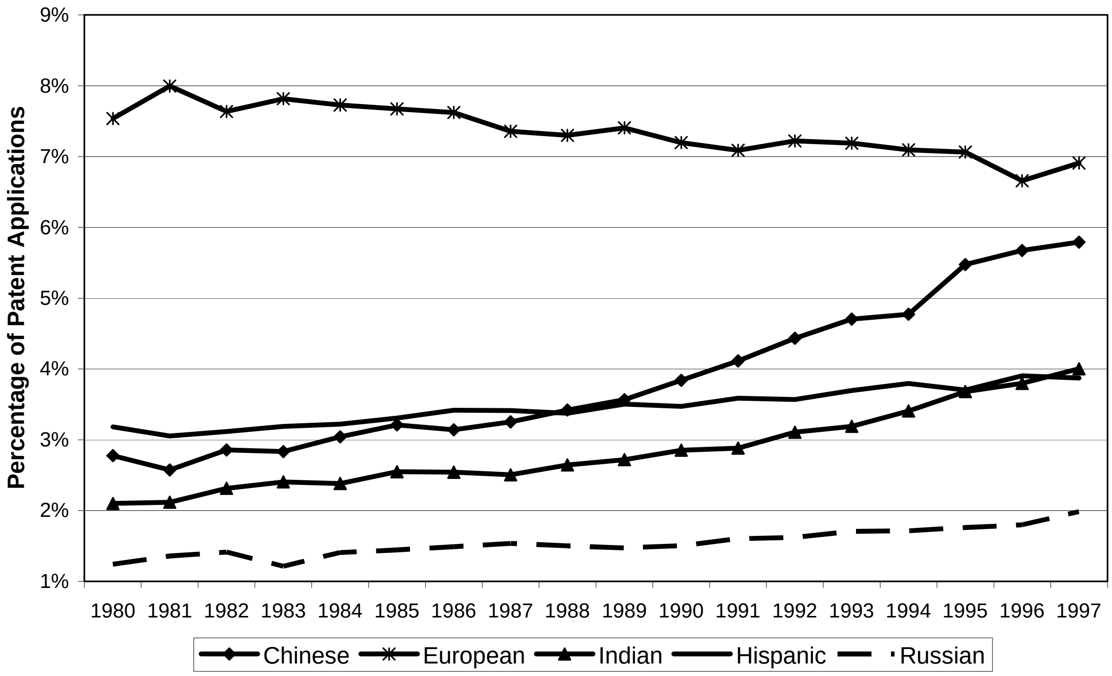

Figure 2: Own-Ethnicity Citation Rate

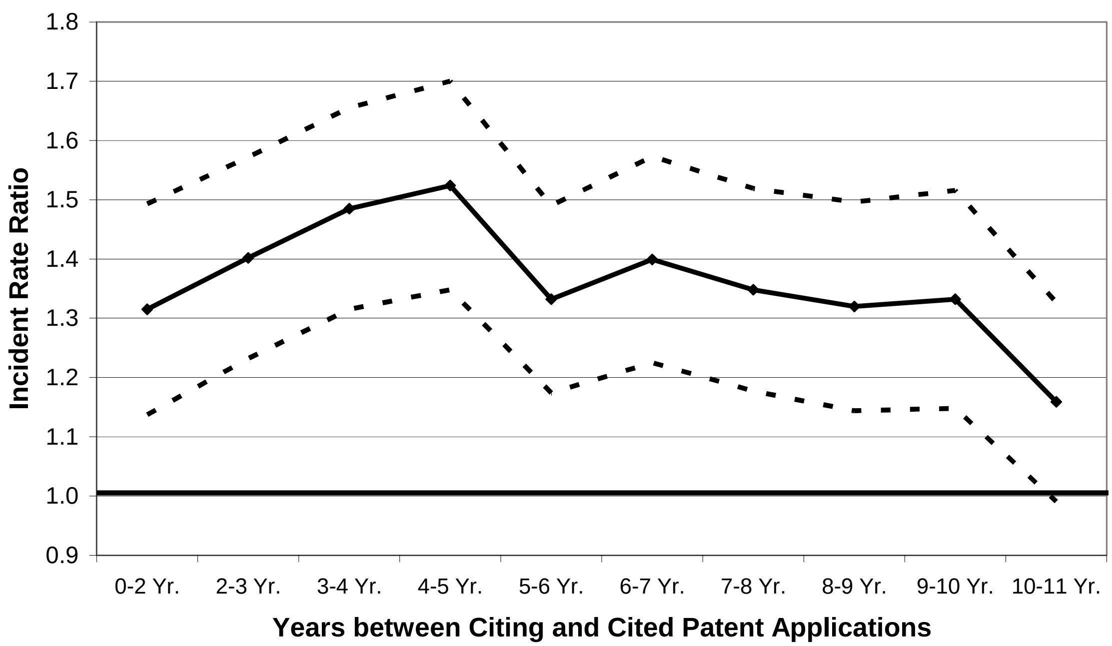

Figure 3: Science & Engineering Immigration

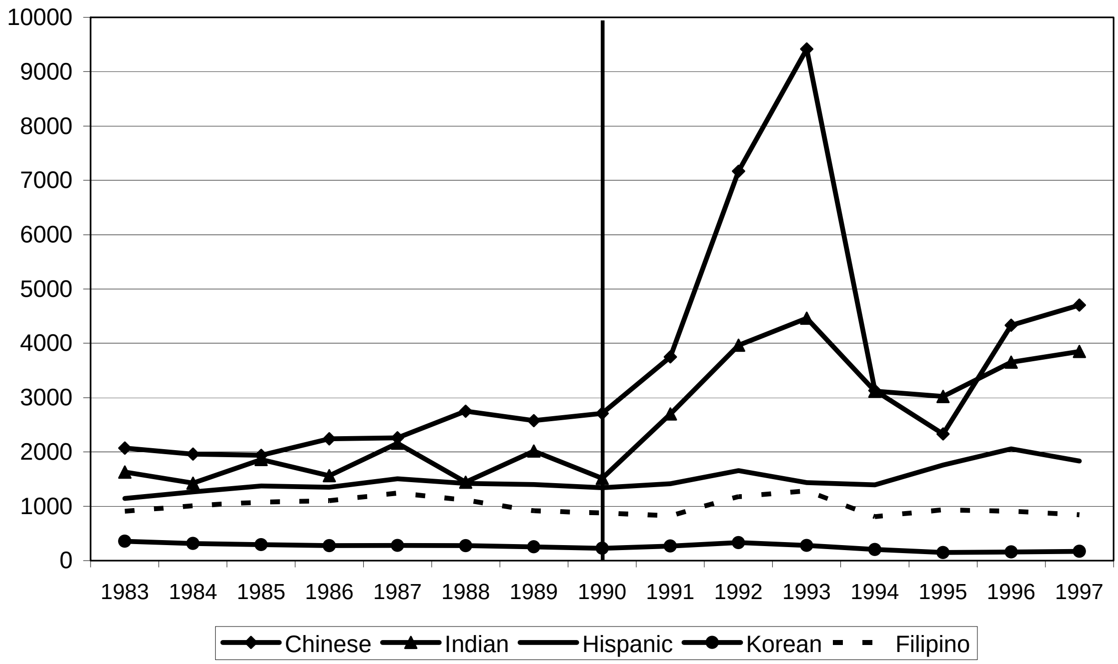

Figure 4: US SE Ph.D. Graduates Staying

Percentage of Graduates from Country Expecting to Stay in US

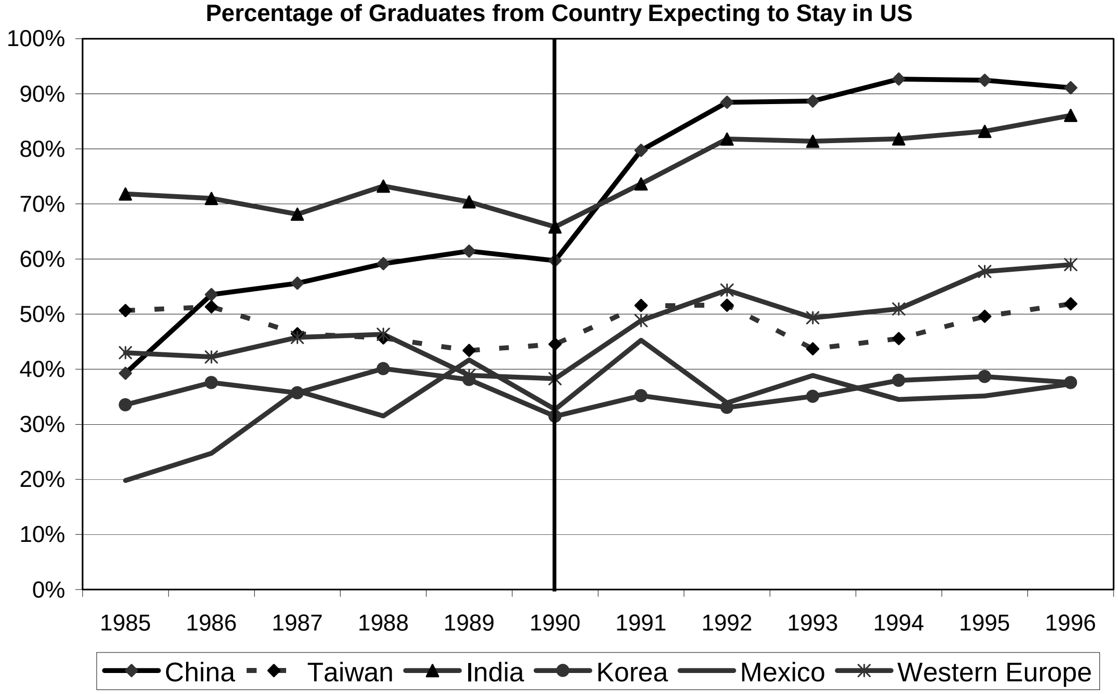

Table 1: Descriptive Statistics for Inventors Residing In US

	Ethnicity of Inventor (Percent Distribution)								
	English	Chinese	European	Hispanic	Indian	Japanese	Korean	Russian	Vietnamese
	A. Ethnic Inventor Shares Estimated from US Inventor Records								
1985-1990 Share	79.7	3.7	7.3	3.3	2.9	0.8	0.7	1.5	0.2
1990-1997 Share	76.4	5.4	6.9	3.7	3.7	0.9	0.8	1.7	0.4
Chemicals	74.4	6.5	7.5	3.6	4.3	0.9	0.9	1.6	0.3
Computers	75.2	6.4	6.2	3.5	4.7	0.9	0.8	1.7	0.7
Pharmaceuticals	75.5	5.2	7.5	4.1	3.8	1.1	1.0	1.6	0.3
Electrical	75.0	6.3	7.0	3.6	3.7	1.0	0.9	1.9	0.5
Mechanical	81.9	2.5	7.2	3.2	2.4	0.6	0.5	1.5	0.2
Miscellaneous	82.6	2.4	7.0	3.5	2.0	0.5	0.5	1.3	0.2
Top MSAs as a	KC (89)	SF (12)	NYC (11)	MIA (17)	NYC (6)	LA (2)	BAL (3)	BOS (3)	AUS (2)
Percentage of MSA's	WS (89)	LA (7)	NOR (11)	SD (8)	BUF (6)	SD (2)	COL (2)	NYC (3)	LA (1)
Patents	MEM (86)	NYC (7)	STL (11)	WPB (6)	AUS (6)	SF (2)	SF (2)	PRO (3)	SF (1)
	B. Ethnic Scientist and Engineer Shares Estimated from 1990 US Census Records								
Bachelors Share	87.6	2.7	2.3	2.4	2.3	0.6	0.5	0.4	1.2
Masters Share	78.9	6.7	3.4	2.2	5.4	0.9	0.7	0.8	1.0
Doctorate Share	71.2	13.2	4.0	1.7	6.5	0.9	1.5	0.5	0.4

Notes: MSAs - AUS (Austin), BAL (Baltimore), BOS (Boston), BUF (Buffalo), COL (Columbus), HRT (Hartford), KC (Kansas City), LA (Los Angeles), MEM (Memphis), MIA (Miami), NOR (New Orleans), NYC (New York City), PRO (Providence), SA (San Antonio), SD (San Diego), SF (San Francisco), STL (St. Louis), WPB (West Palm Beach), and WS (Winston-Salem). MSAs are identified from inventors' city names using city lists collected from the Office of Social and Economic Data Analysis at the University of Missouri, with a matching rate of 98%. Manual coding further ensures all patents with more than 100 citations and all city names with more than 100 patents are identified. 1990 Census statistics are calculated by country-of-birth using the country-ethnicity groupings listed in Table 2; English provides a residual in the Census statistics.

Table 2: UNIDO Industry Sample

	1980	UNIDO3	Output			1980	UNIDO3	Output	
Country	Agr. Share	Panel	Level	Growth	Country	Agr. Share	Panel	Level	Growth
Single Ethnic Mappings:					*Chinese Economies:*				
India	70%	85-97	117,950	6%	China, Mainland	74%	85-97	327,173	11%
Japan	11%	85-97	2,053,048	7%	Hong Kong	1%	85-97	30,520	3%
South Korea	37%	85-97	230,942	14%	Macao	6%	85-97	1,209	8%
Russia	16%	93-97	109,729	12%	Singapore	2%	85-97	37,830	16%
Soviet Union	16%	85-89	1,087,914	7%	Taiwan	8%	85-96	145,055	11%
European Economies:					*Hispanic Economies:*				
Austria	10%	85-97	73,524	5%	Argentina	13%	85-90, 93-96	66,160	11%
Belgium	3%	85-92, 95-97	31,958	5%	Bolivia	53%	85-97	1,474	7%
Denmark	7%	85-91	38,198	9%	Brazil	37%	90, 92-95	127,807	11%
Finland	12%	85-97	52,510	4%	Chile	21%	85-97	20,604	10%
France	8%	85-96	517,276	8%	Columbia	40%	85-97	20,099	5%
Germany	7%	91-97	870,625	7%	Costa Rica	35%	85-97	3,264	5%
Germany, East		85-92	233,905	12%	Cuba	24%	85-89	10,531	-1%
Germany, West		85-89	734,523	12%	Ecuador	40%	85-97	4,372	3%
Italy	13%	85-94, 96-97	390,266	7%	Honduras	57%	90-95	989	8%
Luxembourg	5%	85-97	2,952	3%	Mexico	36%	85-97	61,612	4%
Netherlands	6%	85-97	117,868	6%	Panama	29%	85-94, 96-97	1,468	4%
Norway	8%	85-97	37,467	4%	Peru	40%	85-92, 94-96	13,944	8%
Poland	30%	90-97	54,895	6%	Philippines	52%	85-97	23,238	11%
Sweden	6%	85-97	93,727	6%	Portugal	26%	85-97	36,365	8%
Switzerland	6%	86-96	37,827	7%	Spain	18%	85-97	201,951	8%
					Uruguay	17%	85-97	4,648	6%
					Venezuela	15%	85-97	24,174	1%

Notes: Output values are expressed in millions of 1987 US dollars. Levels and growth rates are unweighted averages of yearly country-level aggregates derived from the industry data used in the UNIDO3 panel. See Kerr (2007a) for additional descriptive statistics and data preparation steps. ISIC Rev. 2 Industries: Food products (311), Beverages (313), Tobacco (314), Textiles (321), Wearing apparel, except footwear (322), Leather products (323), Footwear, except rubber or plastic (324), Wood products, except furniture (331), Furniture, except metal (332), Paper and products (341), Printing and publishing (342), Industrial chemicals (351), Other chemicals (352), Petroleum refineries (353), Misc. petroleum and coal products (354), Rubber products (355), Plastic products (356), Pottery, china, earthenware (361), Glass and products (362), Other non-metallic mineral products (369), Iron and steel (371), Non-ferrous metals (372), Fabricated metal products (381), Machinery, except electrical (382), Machinery, electric (383), Transport equipment (384), Professional & scientific equipment (385), and Other manufactured products (390). Industry 390 is excluded.

Table 3: UNIDO First-Differences Specifications

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Δ Log Foreign Output		
Δ Log US Ethnic Research Community	0.091	0.340	0.285
	(0.056)	(0.133)	(0.074)
Observations	8736	8736	8736
	B. Δ Log Foreign Labor Productivity		
Δ Log US Ethnic Research Community	0.087	0.214	0.217
	(0.049)	(0.114)	(0.072)
Observations	8736	8736	8736
	C. Δ Log Foreign Employment		
Δ Log US Ethnic Research Community	0.003	0.127	0.068
	(0.036)	(0.084)	(0.047)
Observations	8736	8736	8736

Notes: Row titles document the dependent variable studied; column titles document the weighting scheme employed. Panel estimations consider country-industry-year observations taken from the 1985-1997 UNIDO manufacturing database. Log US Ethnic Research Communities are estimated at the ethnicity-industry-year level from the US ethnic patenting dataset. Regressions include industry-year fixed effects. Standard errors are conservatively clustered at the ethnicity-industry level.

Table 4: UNIDO Country Controls Specifications

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Base Foreign Output Regression		
Δ Log US Ethnic Research Community	0.091 (0.056)	0.340 (0.133)	0.285 (0.074)
Observations	8736	8736	8736
	B. Including Foreign Ph.D. Students in US		
Δ Log US Ethnic Research Community	0.061 (0.035)	0.313 (0.073)	0.210 (0.065)
Δ Log Foreign Ph.D. Students in US	0.038 (0.068)	0.050 (0.081)	0.053 (0.073)
Observations	7780	7780	7780
	C. Including Foreign Physical-Capital Stocks		
Δ Log US Ethnic Research Community	0.026 (0.069)	0.275 (0.173)	0.209 (0.091)
Δ Log Foreign Capital Stock	0.069 (0.030)	0.112 (0.047)	0.059 (0.034)
Observations	4866	4866	4866
	D. Including Country Time Trends		
Δ Log US Ethnic Research Community	0.000 (0.061)	0.130 (0.102)	0.153 (0.068)
Observations	8736	8736	8736
	E. Including Country-Year Effects		
Δ Log US Ethnic Research Community	-0.092 (0.048)	0.149 (0.107)	-0.022 (0.059)
Observations	8736	8736	8736

Notes: See Table 3. Panel A replicates the foreign country-industry output regressions from Table 3. Panels B through E incorporate the country controls indicated by the row titles. All regressions maintain industry-year fixed effects and the clustering of standard errors.

Table 5: UNIDO Sample Decompositions

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Base Foreign Output Regression		
Δ Log US Ethnic Research Community	0.091	0.340	0.285
	(0.056)	(0.133)	(0.074)
Observations	8736	8736	8736
	B. Excluding Computers and Drugs		
Δ Log US Ethnic Research Community	0.058	0.126	0.207
	(0.054)	(0.076)	(0.063)
Observations	7991	7991	7991
	C. Excluding Mainland China		
Δ Log US Ethnic Research Community	0.059	0.308	0.258
	(0.061)	(0.166)	(0.086)
Observations	8518	8518	8518
	D. Excluding All Chinese Economies		
Δ Log US Ethnic Research Community	0.059	0.195	0.238
	(0.058)	(0.131)	(0.082)
Observations	7616	7616	7616
	E. Excluding All Advanced Economies		
Δ Log US Ethnic Research Community	0.117	0.386	0.255
	(0.080)	(0.116)	(0.079)
Observations	5549	5549	5549
	F. Excluding All Hispanic Economies		
Δ Log US Ethnic Research Community	0.055	0.334	0.243
	(0.071)	(0.162)	(0.091)
Observations	4821	4821	4821

Notes: See Table 3. Panel A replicates the foreign country-industry output regressions from Table 3. Panels B through F exclude the observations indicated by the row titles. All regressions maintain industry-year fixed effects and the clustering of standard errors.

Table 6: UNIDO Sector Reallocation Specifications

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Δ Log Foreign Output		
Δ Log US Ethnic Research Community	0.043	0.315	0.252
	(0.062)	(0.153)	(0.086)
Δ Log US Ethnic Comm. x 1980 Agriculture Share	0.765	0.442	0.647
	(0.185)	(0.353)	(0.242)
Observations	8736	8736	8736
	B. Δ Log Foreign Labor Productivity		
Δ Log US Ethnic Research Community	0.105	0.225	0.228
	(0.047)	(0.106)	(0.068)
Δ Log US Ethnic Comm. x 1980 Agriculture Share	-0.284	-0.191	-0.216
	(0.097)	(0.162)	(0.120)
Observations	8736	8736	8736
	C. Δ Log Foreign Employment		
Δ Log US Ethnic Research Community	-0.062	0.091	0.024
	(0.037)	(0.084)	(0.047)
Δ Log US Ethnic Comm. x 1980 Agriculture Share	1.049	0.633	0.863
	(0.146)	(0.266)	(0.198)
Observations	8736	8736	8736

Notes: Row titles document the dependent variable studied; column titles document the weighting scheme employed. Panel estimations consider country-industry-year observations taken from the 1985-1997 UNIDO manufacturing database. 1980 Agriculture Shares for foreign countries are listed in Table 2. Log US Ethnic Research Communities are estimated at the ethnicity-industry-year level from the US ethnic patenting dataset. Main effects are demeaned prior to interactions. Regressions include industry-year fixed effects. Standard errors are conservatively clustered at the ethnicity-industry level.

Table 7: Immigration Quotas Specifications

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Δ Log Foreign Output		
Δ Log US Immigration Quotas Estimator	0.294	0.370	0.320
	(0.044)	(0.086)	(0.057)
Observations	8736	8736	8736
	B. Δ Log Foreign Labor Productivity		
Δ Log US Immigration Quotas Estimator	0.054	0.135	0.086
	(0.069)	(0.080)	(0.073)
Observations	8736	8736	8736
	C. Δ Log Foreign Employment		
Δ Log US Immigration Quotas Estimator	0.240	0.236	0.234
	(0.045)	(0.052)	(0.037)
Observations	8736	8736	8736

Notes: Row titles document the dependent variable studied; column titles document the weighting scheme employed. Panel estimations consider country-industry-year observations taken from the 1985-1997 UNIDO manufacturing database. Log US Immigration Quotas Estimators are developed from quotas changes due to the 1990 Act. Regressions include industry-year fixed effects. Standard errors are conservatively clustered at the ethnicity level.

Table 8: Imm. Quotas Country Controls Specifications

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Base Foreign Output Regression		
Δ Log US Immigration Quotas Estimator	0.294	0.370	0.320
	(0.044)	(0.086)	(0.057)
Observations	8736	8736	8736
	B. Including Foreign Ph.D.s in US		
Δ Log US Immigration Quotas Estimator	0.280	0.350	0.297
	(0.057)	(0.110)	(0.078)
Δ Log Foreign Ph.D. Students in US	0.033	0.054	0.054
	(0.068)	(0.088)	(0.076)
Observations	7780	7780	7780
	C. Excluding Mainland China		
Δ Log US Immigration Quotas Estimator	0.210	0.306	0.252
	(0.094)	(0.152)	(0.111)
Observations	8518	8518	8518
	D. Including Ethnic Time Trend		
Δ Log US Immigration Quotas Estimator	0.317	0.385	0.327
	(0.077)	(0.135)	(0.101)
Observations	8736	8736	8736
	E. Including 1987 Counterfactual		
Δ Log US Immigration Quotas Estimator	0.262	0.335	0.287
	(0.047)	(0.061)	(0.050)
Δ 1987 Counterfactual Quotas Estimator	0.296	0.326	0.311
	(0.199)	(0.262)	(0.232)
Observations	8736	8736	8736
	F. Including 1995 Counterfactual		
Δ Log US Immigration Quotas Estimator	0.266	0.439	0.345
	(0.081)	(0.096)	(0.067)
Δ 1995 Counterfactual Quotas Estimator	0.069	-0.170	-0.063
	(0.152)	(0.099)	(0.110)
Observations	8736	8736	8736

Notes: See Table 7. Panel A replicates the foreign country-industry output regressions from Table 7. Panels B through F incorporate the country controls indicated by the row titles. All regressions maintain industry-year fixed effects and the clustering of standard errors.