The Role of Immigrant Scientists and Entrepreneurs in International Technology Transfer

by

William Robert Kerr

B.S., Systems Engineering and Economics
University of Virginia 1996

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2005

Signature of Author ....
Department of Economics
10 August 2005

Certified by ....
K. Daron Acemgolu
Charles P. Kindleberger Professor of Applied Economics
Thesis Supervisor

Certified by ....
David H. Autor
Pentti J.K. Kouri Career Development Associate Professor
Thesis Supervisor

Certified by ....
~~Ricardo J. Caballero~~
Ford International Professor of Economics
Thesis Supervisor

Accepted by ....
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee on Graduate Studies

The Role of Immigrant Scientists and Entrepreneurs in International Technology Transfer

by

William Robert Kerr

Submitted to the Department of Economics
on 10 August 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis characterizes the important role of US ethnic scientists and entrepreneurs for international technology diffusion. Chapter 1 studies the transfer of tacit knowledge regarding new innovations through ethnic scientific 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 approximately 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 findings suggest tacit knowledge channels partly shape the effective technology frontiers of developing economies.

Chapter 2 further exploits this heterogeneous technology diffusion through ethnic networks to test the importance of Ricardian technology differences for international trade. Panel regressions find technology growth increases manufacturing exports. To establish a causal relationship between technology and trade, instrumental-variables specifications exploit uneven technology diffusion from the US through ethnic scientific networks. The instrumented elasticity of export growth to the exporter's technology development is 0.9 in the preferred specification. Supplemental specifications show this elasticity is robust to controlling for the importer's technology development and to Rybczynski effect due to factor accumulation. Exogenous reforms of US immigration law again test for reverse causality. The findings suggest technology differences are an important determinant of trade patterns.

As a supplement to these first two studies, Chapter 3 provides detailed documentation on the ethnic-name strategy employed with US patent records. The growing contribution of Chinese and Indian scientists to US technology formation, especially in high-tech industries, is described. The institutional and geographic dimensions of US ethnic innovation are further characterized. Finally, Chapter 4 concludes with an independent study of income inequality and social norms for compensation differentials and government-led redistribution. This work demonstrates that short-run responses in social norms do not amplify income inequality shocks (e.g., due to skill-biased technical change).

Thesis Supervisor: K. Daron Acemgolu
Title: Charles P. Kindleberger Professor of Applied Economics

Thesis Supervisor: David H. Autor
Title: Pentti J.K. Kouri Career Development Associate Professor

Thesis Supervisor: Ricardo J. Caballero
Title: Ford International Professor of Economics

Acknowledgements

This thesis is only possible due to the exceptional contributions and support of Daron Acemoglu, my steadfast mentor and primary advisor. Over the past five years, Daron consistently taught me how to establish and attain intellectual standards I once thought unachievable. His tremendous breadth and creativity for economic research further facilitated cross-disciplinary inquiries normally not possible. Most importantly, his abundant patience and hearty laugh made the strenuous work and heavy sacrifices bearable. His devotion to intellectual excellence and rigorous scholarly research will inspire me throughout my career as an economist.

I am further grateful to my secondary advisors, David Autor and Ricardo Caballero, for their strong commitment to my academic development. I also thank Josh Angrist, Olivier Blanchard, Glenn Ellison, Mike Piore, and the faculty members of MIT Economics for their persistent and exceptional support of graduate students through classroom instruction, thesis advising, and job market preparation. John Arditi and Gary King brought a welcomed cheerfulness to each day. Jim Davis, Simon Gilchrist, Robert King, Josh Lerner, Steve Machin, and Ken Rogoff provided extensive support outside of MIT. I hope I can pass on to future students the same kindness and wisdom shown to me.

Exceptional classmates are another major benefit of studying at MIT. Ashley Lester and Byron Lutz, with whom I have shared an office for the last three years, have been particularly influential in my intellectual development. As we move in our separate directions, I know they will remain close research advisors and great friends. An incomplete list of my other classmates I want to thank are Elizabeth Oltmans-Annat, Angus Armstrong, Emek Basker, Melissa Boyle, Shawn Cole, Rema Hanna, Christian Hansen, Youngjin Hwang, David Lyle, Nancy Qian, and Petia Topalova.

Outside of MIT, several communities in Boston have supported me over the past five years. In particular, I thank my small group members at Park Street Church for the spiritual growth they fostered during our time together. Their patience and prayers during my extended absences are especially appreciated. I also thank Community Running and my Virginia friends living in Boston for welcomed breaks from my labors. Sari Pekkala provided wonderful encouragement and support during the completion of this work in the summer of 2005.

Finally, I wish to dedicate this thesis first to God's glory. He has been supported me in ways I fail to understand; He has also used the trials and struggles of graduate school to bring me closer to Him than I ever have been before. This work is also dedicated to my loving parents, Joe and Carole Kerr, who have showered me with unconditional support and affection throughout my life. They have encouraged me to achieve the most with my talents and weaknesses, but they have also taught me through their own lives and careers that priorities should be carefully chosen. It is not possible to express completely my love and gratitude to them.

Contents

Chapter 1

Ethnic Scientific Communities and International Technology Diffusion

Summary 1 *This study explores the importance of tacit knowledge transfer for international technology diffusion by examining ethnic scientific 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 approximately 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 findings suggest tacit knowledge channels partly shape the effective technology frontiers of developing economies.*

1.1 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.[2] 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 tacit knowledge associated with new innovations shapes the effective technology sets of following countries.[3,4]

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, 2002b). 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

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

[2] An 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 Lester and Piore (2004), Amsden (2001), Feinstein and Howe (1997), Kim (1997), and Lim (1999). Polanyi (1958, 1966) introduces tacit knowledge.

[3] Marshall (1890) and Jacobs (1970) describe the forces contributing to spatial agglomeration, while Dumais, Ellison, and Glaeser (1997) and Rosenthal and Strange (2003) provide more recent empirical tests.

[4] 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.

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 whether and how international networks transmit the tacit knowledge of new inventions. The primary question this project addresses is whether a larger ethnic research community in the US improves technology diffusion to foreign countries of the same ethnicity. US research communities are quantified for ethnicities 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 cross-sectional and longitudinal 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. This approach affords a more structured characterization of ethnicity's role in the diffusion process; it further allows the outcomes of different ethnicities and industries to be contrasted.

To clarify this empirical methodology, the next section develops a theoretical model focusing on tacit knowledge and technology transfer. The model considers a technology follower that 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 tacit knowledge with respect to the frontier inventions. This tacit 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 technology leader. The primary estimating equations employed in this study are determined within this framework, and the conditions under which ordinary least squares estimations capture causal relationships are identified.

Section 1.3 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 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 four to five years of the diffusion process.

While informative, citation patterns do not demonstrate that 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 1.4. Ethnic research communities are quantified at the industry-year level by aggregating individual patent records. Stringent fixed-effects 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, and robustness checks further consider human-capital and physical-capital developments abroad, general country trends, and so on. The results suggest growth in ethnic scientific communities in the US increases foreign output with elasticities of 0.2-0.4, with the elasticity estimate depending upon how the data are weighted. The parameter estimates are statistically significant and robust across a wide class of specifications. The positive benefits are evident throughout the manufacturing industries studied, and the output expansion is decomposed into employment and labor productivity gains.

Estimated technology transfers between countries may be capturing the true diffusion process, or they could simply be correlated with omitted factors. Reverse causality is also a prominent concern, where human-capital developments in the foreign country could simultaneously result in higher output growth and more ethnic researchers emigrating to the US. Section 1.5 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 coefficients of the two approaches cannot be directly compared, the qualitative directions support each other.

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 1.2. 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 tacit knowledge regarding new innovations does contribute to slow technology diffusion. Ethnic 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 countries. The chapter concludes in Section 1.6 with a discussion of related projects currently being pursued with the ethnicity approach that will further refine our understanding of how these tacit knowledge channels operate.[5]

1.2 Theoretical Framework

This section outlines a simple technology transfer model between a frontier country and a following nation. 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.

The invention process is characterized by a "standing on the shoulders of giants" framework, where spillovers from past innovations increase the research productivity of current scientists

[5]In addition to characterizing technology diffusion, quantifying these ethnic linkages is important in its own right. Saxenian's work is frequently discussed in the "brain drain" versus "brain circulation" debate regarding high-skilled immigration to advanced countries. The structured econometrics of this study are an important complement to case studies for evaluating these arguments.

and generate endogenous growth.[6] Knowledge is local, however, in that a country's research 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 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 tacit knowledge channel. 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 tacit knowledge necessary for effective imitation in the following country.

To greatly simplify the exposition, the frontier economy is labeled the US and the following economy is labeled China. Variables for the US economy are denoted by a tilda (e.g., $\tilde{Y}$), while China'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 China's economy, followed by differences in the US economy. The steady-state outcome is then characterized. The section closes with simulations of the transitional dynamics to this steady-state.[7,8]

1.2.1 China's Economy

China's economy contains L workers of homogeneous ethnicity 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 product 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 a brief description of the consumer side of the economy.

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.

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

[7] Technology flows are the only interactions between the two countries. The model abstracts from trade, and immigration is restricted in the base scenario.

[8] "China" is selected for the following-country label due to the prominence of the Chinese ethnicity in the empirical findings. This labeling brings to the forefront the contrast of Mainland China and the Chinese tiger economies of Hong Kong, Singapore, and Taiwan discussed below.

Production for a representative firm i that employs labor L_{M_i} and non-durable intermediates X_{ij} of type j takes the form

$$Y_{M_i} = AL_{M_i}^{1-\alpha} \sum_{j=1}^{N} (X_{ij})^{\alpha}. \tag{1.1}$$

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 China. In equilibrium firms employ equal amounts of all intermediate inputs ($X_{ij} = X_i \ \forall j$) and (1.1) 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 US 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 China is determined by the existing stock of China's inventions, or $\partial I / \partial t = I \cdot L_R$. There are no international knowledge spillovers in the sense that researchers in China cannot build on the US stock of inventions directly in innovation. China's researchers can alternatively imitate US inventions at a rate

$$\frac{\partial M}{\partial t} = \left(\tilde{I} \Psi \left[\frac{M}{\tilde{I}} \right] (\tilde{H}^C)^{\beta} \right) \cdot L_R, \tag{1.2}$$

where $\tilde{I}$ is the US invention stock and $\tilde{H}^C$ is the Chinese human-capital stock with respect to US inventions. A larger stock of US inventions affords a larger pool of technologies that can be imitated, thus raising the imitation productivity for a researcher in China. The imitation of products exhausts the available pool, however, and the function Ψ decreases with the ratio of imitated products to the available US 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 $(\tilde{H}^C)^{\beta}$ specification models that tacit knowledge of US inventions is necessary for successfully adopting them in China. This human-capital stock depreciates at a rate δ, and the population of Chinese researchers in the US undertaking inventive activity adds to it: $\partial \tilde{H}^C / \partial t = -\delta \tilde{H}^C + \tilde{L}_R^C$. If the number of US Chinese researchers is constant, the steady-state stock of Chinese human capital with respect to US inventions is $\delta^{-1}\tilde{L}_R^C$.

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 China. The present discounted value of these rents for a good j at time t is

$$V(t) = \int_t^{\infty} (P_j - C_j) X_j e^{-\bar{r}(s,t)\cdot(s-t)} ds, \tag{1.3}$$

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 all research firms as one unit of Y_M is required to produce one unit of an 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 (1.1), 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 (1.3) to

$$V = \left(\frac{1-\alpha}{\alpha}\right) A^{1/(1-\alpha)}\alpha^{2/(1-\alpha)} \frac{1}{r} L_M. \tag{1.4}$$

Constant intermediate demand functions also simplify China's aggregate output,

$$Y_M = A^{1/(1-\alpha)}\alpha^{2\alpha/(1-\alpha)} L_M N. \tag{1.5}$$

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, $\rho = r$.

1.2.2 US Economy

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

Researchers of both ethnicities contribute to and utilize the existing US invention stock $\tilde{I}$ in developing new intermediate products: $\partial\tilde{I}^C/\partial t = \tilde{I} \cdot \tilde{L}_R^C$ and $\partial\tilde{I}^E/\partial t = \tilde{I} \cdot \tilde{L}_R^E$, where $\tilde{I} = \tilde{I}^C + \tilde{I}^E$. This research specification again highlights the role of past inventions $\tilde{I}$ in making current researchers more productive, and assumes inventions made in China do not contribute to US researcher productivity for invention. More subtly, ethnicity does not matter for invention in the US — Chinese and English scientists are symmetric with respect to the US invention stock. Finally, US researchers of Chinese ethnicity can imitate products made in China with a productivity analogous to (1.2).[9]

1.2.3 Steady-State Description: US Invents, China Imitates

This case determines the core estimating equation for this study. Without invention in China, the US economy operates in isolation, and imitation does not occur ($\tilde{N} = \tilde{I}$). The US 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 making a new invention in the US. 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 US versions of the value of innovations (1.4) and aggregate output (1.5), and noting $r = \rho$, the steady-state

[9]The potential crowding out of US workers and students from science and engineering fields by immigrants is often debated (e.g., Borjas 2004). This model incorporates a crowding-out effect for analytical convenience only.

allocation of labor in the US 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 US intermediate technologies and manufacturing output is $L - \rho/\alpha$.

Returning to China's economy, all intermediate products come through imitation of US goods ($N = M$). Labor mobility again requires that wages be equal across China's research and manufacturing sectors,

$$V \cdot \left(\tilde{I}\Psi \left[\frac{M}{\tilde{I}} \right] (\tilde{H}^C)^\beta \right) = (1-\alpha)\frac{Y_M}{L_M}.$$

Substituting in the value of new intermediates V from (1.4) and aggregate output Y_M from (1.5),

$$r = \frac{\tilde{I}}{M}\Psi \left[\frac{M}{\tilde{I}} \right] (\tilde{H}^C)^\beta \alpha L_M. \tag{1.6}$$

With identical preferences and aggregate populations, China's interest rate and allocations of labor to manufacturing and research are the same as the US.[10] Equation (1.6) further shows the steady-state ratio of China's imitated products to available US products $M/\tilde{I}$ is constant and increases with the Chinese human-capital stock with respect to US technologies ($\Psi' < 0$). Stronger tacit knowledge improves researcher productivity for imitation in China and closes the steady-state gap to the US frontier.

Simplifying (1.6) for economies of equal size relates China's imitated technology stock to the US technology frontier and China's tacit knowledge for US innovations,

$$M = \tilde{I}\Psi \left[\frac{M}{\tilde{I}} \right] (\tilde{H}^C)^\beta. \tag{1.7}$$

Substituting this relationship into China's manufacturing output (1.5),

$$Y_M = A^{1/(1-\alpha)}\alpha^{2\alpha/(1-\alpha)} L_M \cdot \left(\tilde{I}\Psi \left[\frac{M}{\tilde{I}} \right] (\tilde{H}^C)^\beta \right).$$

Taking logs and collapsing time-invariant terms into a constant ϕ, China's manufacturing output depends upon its human-capital stock with respect to US research with elasticity β, $\ln(Y_M) =$

[10] 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.

$\phi + \ln(\tilde{I}) + \beta \ln(\tilde{H}^C)$. The human-capital stock is $\delta^{-1}\tilde{L}_R^C$ in steady-state, so that

$$\ln(Y_M) = \phi + \ln(\tilde{I}) + \beta \ln(\tilde{L}_R^C), \tag{1.8}$$

where δ^{-1} is absorbed into the constant. Equation (1.8) is the basis for the estimating equations employed in Sections 1.4 and 1.5. 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 imitation-versus-invention decision in China determines the condition required for this steady-state description. Specifically, the productivity of researchers for undertaking invention in China must be less than the researcher productivity for imitating US innovations in equilibrium,

$$I < \tilde{I}\Psi \left[\frac{M}{\tilde{I}}\right] (\tilde{H}^C)^{\beta}. \tag{1.9}$$

The assumption $I = 0$ requires (1.9) hold forever; without a knowledge stock on which to build, a first invention is impossible. While this is a valid description for 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 tacit knowledge for a growing stock of frontier innovations is maintained. Section 1.5 discusses the case where (1.9) no longer holds.[11]

1.2.4 Agriculture to Manufacturing Sector Reallocation

The steady-state characterization of China's economy builds on the assumption of full employment in the manufacturing and research sectors. While the estimating equation (1.8) relates China's output to its research presence in the US, the same elasticity β would hold for 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).

[11] Immigration is restricted in this framework. Moreover, Chinese workers would prefer to emigrate to the US as the US wage rate is higher *ceteris paribus* due to the larger stock of intermediate goods.

This section incorporates into the basic model an agricultural sector in China. The framework highlights how technology transfer induces a different response when sector reallocation is possible. Specifically, as more technologies are transferred from the US to China, labor shifts from agriculture to the manufacturing and research sectors. After a sufficient number of frontier innovations have been imitated, China'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 above; numerical simulations of the transition path, however, offer additional guidance for the empirical exercises this study undertakes.

The agricultural sector for China is characterized by a decreasing returns to scale technology that employs only labor L_A,

$$Y_A = BL_A - \frac{1}{2}L_A^2. \tag{1.10}$$

B is a common agricultural productivity parameter, and the final goods from agriculture and manufacturing are identical ($Y = Y_A + Y_M$). Labor is again free to move across sectors, and the proportion of the labor force allocated to each of the three sectors along the development path can be related to China's technology stock. China is assumed to possess only imitated technologies ($M = N$).

First, the marginal products of labor for agriculture and manufacturing are $B - L_A$ and $(1-\alpha)A^{1/(1-\alpha)}\alpha^{2\alpha/(1-\alpha)}M$, respectively. Wage equality between these two sectors relates the size of the agricultural workforce to the number of imitated technologies,

$$L_A = \max[B - (1-\alpha)A^{1/(1-\alpha)}\alpha^{2\alpha/(1-\alpha)}M, 0]. \tag{1.11}$$

Thus, growth in China's technology stock lowers agricultural employment until the economy reaches a transition point with full employment in the research and manufacturing sectors. This transition occurs when $M > (1-\alpha)^{-1}A^{-1/(1-\alpha)}\alpha^{-2\alpha/(1-\alpha)}B$. If this condition is satisfied in the current period, it will hold in all future periods as the wages of the manufacturing and research sectors continue to grow with further technological advancement.[12] Likewise, the size of the manufacturing labor force can be related to China's existing technology stock and the

[12] The agricultural production function (1.10) bounds the marginal productivity of labor from above at B. This function does not satisfy the Inada conditions. An agricultural sector with a constant returns to scale production function employing land and labor yields an ever shrinking agricultural sector.

US technology stock through the wage equality of the manufacturing and research sectors (1.6) and the interest rate $r = \rho$,

$$L_M = \frac{M}{\tilde{I}}\Psi^{-1}\left[\frac{M}{\tilde{I}}\right](\tilde{H}^C)^{-\beta}\frac{\rho}{\alpha}. \tag{1.12}$$

With L_A and L_M determined, the number of researchers follows from the labor endowment.[13]

To characterize the transition path, a numerical solution for the steady-state without an agricultural sector is first developed. The labor forces of the two economies are taken to be of size 0.20. For parameter values of $\alpha = 1/3$ and $\rho = 0.05$, the allocation of labor to manufacturing and research is 0.15 (75%) and 0.05 (25%), respectively. As the growth rate of inventions in the US is equal to the size of its research labor force (i.e., $(\partial\tilde{I}/\partial t)/\tilde{I} = \tilde{L}_R$), the set of available frontier technologies grows at a rate of 5%; the same growth rate is in turn found for China's imitated technology stock and the manufacturing outputs of the two countries. Specifying $A = 1$, the steady-state value of new inventions or imitations in both economies is given a numerical value of $V = 0.22$.

Examining more closely the technology transfer mechanism, the size of the Chinese research population living in the US is modeled as 0.001 (or 2% of the US researcher total). Taking an estimate of $\beta = 0.3$ from the empirical exercises in Section 1.4 and assuming a depreciation on human capital of $\delta = 0.15$, the steady-state human-capital stock is given a numerical value of $\tilde{H}^C = 0.0067$. Finally, a functional form for $\Psi[M/\tilde{I}]$ must be specified to estimate the share of US technologies imitated by China. The form $\Psi[M/\tilde{I}] = 0.5 \cdot (1 - M/\tilde{I})$ retains the properties of $\Psi' < 0$ and $\Psi[1] = 0$ and yields a steady-state imitation share $M/\tilde{I} = 0.10$.

Turning to the transition path simulations to this steady-state, the solid line in Figure 1.1 describes the evolution of China's economy from the initial conditions of 90% employment in agriculture for China and 1% of US technologies imitated.[14] The size of the agricultural sector corresponds to an initial stock of imitated technologies $M(0)$ in China, while the gap to the US frontier determines the technologies $\tilde{I}(0)$ in the US; the parameter $B = 0.2$ is also specified. The US, assumed to be in steady-state growth, evolves exogenously with a 5% growth rate in

[13] Even if the Chinese human-capital stock for US technologies is stable during the transition period, the relative proportion of labor devoted to manufacturing versus research is not at its steady-state level due to adjusting fraction of imitated technologies $M/\tilde{I}$.

[14] Full employment in agriculture is an unstable equilibrium if Chinese human capital to US technologies exists. Once some labor is devoted to manufacturing and research, China's economy will eventually transition completely out of agriculture if tacit knowledge with respect to the US frontier is maintained.

its technology stock. From (1.11), (1.12), and the labor endowment, the evolution of China's economy is subsequently characterized.

The leftmost panel describes the allocation of labor along the transition path. For the baseline simulation, the Chinese human-capital stock with respect to US technologies remains at its steady-state level of 0.0067. Initially, limited labor is devoted to manufacturing or research. In this general equilibrium, the small market size of final-goods producers depresses the value of new innovations and the researchers employed, even though the large gap to the US frontier makes it very productive to imitate new intermediate products. Likewise, the labor demand of final-goods producers is limited due to the small technology stock in China.

As the number of technologies steadily expands, however, the agricultural sector shrinks and more labor is allocated to both manufacturing and research. This industrialization sustains itself as the growth in market size increases the value of new innovations, while the larger technology base increases the labor demand of final-goods manufacturers. The sector reallocation quickens as the economy approaches the transition point ($t = 16$). Around this transition, the researcher share of China's labor force reaches its peak, before gradually declining to its steady-state value of 25%. The manufacturing labor share also surges around the transition, and continues to grow to its steady-state share of 75%.

The middle panel presents several growth rates evident in China during the transition. Initial growth is slow due to the inertia of the large agricultural sector. Around the transition point, however, growth in China's imitated technology stock surges due to the extensive labor resources devoted to research and the still sizeable gap to the US frontier. This high rate of technology adoption translates into higher growth in both manufacturing output and labor productivity. The manufacturing output growth is not due solely to labor productivity gains, however, as the growth in employment contributes approximately the same amount. After the transition, the growth rates decline to their steady-state rates of 5%.

Finally, the rightmost panel exhibits several levels with respect to the US frontier. As evident in research labor share, the linear preferences of consumers affords a substantial investment in the imitation of new technologies around the transition in return for higher future consumption. During this convergence period, China's technology gap to the US frontier is substantially narrowed. In the steady-state, the fraction of US technologies imitated by China translates directly into the steady-state fraction of US manufacturing output achieved, and the transition path dynamics take the same shape. Note, however, that China's total output level (including agriculture) relative to the US declines slightly during the transition period due to the investment in imitation. Except in the immediate vicinity of the transition point, China's output does not fall but instead fails to maintain pace with the US economy in steady-state growth. After this investment period, however, the share rises sharply to a long-run level equal to the manufacturing output share.

From this baseline, the dotted line in Figure 1.1 plots a second transition path for an exogenous increase in the number of Chinese researchers living in the US from 0.001 to 0.004 on date $t = 3$.[15] While these simulations are meant to be illustrative, the fourfold rise from 2% to 8% of the US research community is roughly in line with the growing Chinese research contribution in several high-tech industries for the period studied in the empirical analysis below. As the top left panel shows, this exogenous increase does not immediately translate into a fourfold increase in the Chinese human-capital stock with respect to US technologies. The human-capital stock instead grows over time with the higher rate of tacit knowledge gain in each period following the US Chinese researcher growth.

As the boost in technology transfer is realized, however, the transition from agriculture proceeds at a more rapid pace. The growth rate of manufacturing output spikes upward due to both higher growth in imitated technologies and more labor reallocation. An economy without an agricultural sector would only experience output growth due to labor productivity gains. In the new steady-state, the fourfold increase in China's human-capital stock results in an approximate 40% levels gain in imitated technologies and output; the percentage of US technologies imitated is also higher. China's growth rate and allocation of labor, though, are the same as in the simulation without the exogenous increase in scientific integration.

In summary, technology transfer to economies with large agricultural sectors can increase manufacturing output through both labor productivity gains, as in the steady-state scenario, and through employment growth. In this particular framework, the productivity gains and employment gains are of roughly similar magnitude. In alternative models, however, output

[15]The simulations abstract from any growth in the overall size of the US labor force due to this inflow.

growth would come only through labor reallocation.[16] After evaluating the core steady-state specification (1.8), Section 1.4 empirically evaluates how the responses of economies with large agricultural sectors differ from those with constrained labor resources.

1.3 Ethnic Patenting and International Citations Analysis

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 initial analysis of knowledge flows using international patent citation records. The ethnic patenting data are then joined with foreign output metrics in the next section to evaluate (1.8) directly.

1.3.1 Ethnic Patenting Records

Ethnic technology development in the US is quantified through the NBER Patent Data File (Hall, Jaffe, and Trajtenberg 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. The match rate is 99% for US patent records, and the process affords the distinction of nine ethnicities: Chinese, English, European, Hispanic, Indian, Japanese, Korean, Russian, and Vietnamese.

Table 1.1 describes the 1985-1997 US sample. The trends demonstrate a growing ethnic contribution to US technological development, especially among Chinese and Indian scientists. Also matching popular perceptions, 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).

[16] For example, specifications with constant outside wages and a fixed stock of physical capital. As the technology transfer increases the marginal product of labor, producers hire more labor to bring the marginal product of labor back down to the external wage. The agricultural sector's production function (1.10) instead allows the marginal product of labor in agriculture to increase in step with the manufacturing sector's wage.

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.[17] Figure 1.2 illustrates the evolving ethnic contribution to US technology development as a percentage of patents granted by the USPTO, while Figure 1.3 provides a more detailed glimpse of ethnic shares by broad technology groups. While the name-matching procedure certainly misclassifies the ethnicities of some inventors, the aggregate trends important for this study appear remarkably accurate.

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. The rows of Table 1.2 present the matched characteristics for countries and regions 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 a large percentage of foreign records (overall matching rate of 98%). Second, the estimated inventor compositions are quite reasonable, with the own-ethnicity contributions in all but three regions being greater than 80%. Similar to the US, own-ethnicity contributions should be less than 100% due to foreign researchers.[18,19]

1.3.2 International Patent Citation Analysis

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. This first exercise simply compares the ethnic composition

[17] The estimated European ethnic contribution is naturally higher than the immigrant contribution measured by foreign born.

[18] Details on the matching algorithms and additional descriptive statistics are presented in Chapter 3.

[19] A further supplement to the NBER patent data is important to highlight. The USPTO issues patents by technology categories rather than by industries. Combining the work of Johnson (1999) and Silverman (1999), concordances are developed between the USPTO classification scheme and the three-digit industries in which new inventions are manufactured or used. The main estimations focus on industry-of-use, affording a composite view of the technological opportunity developed for an industry. Studies of advanced economies find accounting for these inter-industry R&D flows important (e.g., Scherer 1984). Keller (2002a) reports inter-industry R&D flows aid productivity growth significantly within OECD countries, equal to half or more of the own-industry development. Estimations with manufacturing industries support the using-industry specifications.

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.[20]

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.

More than 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.[21] 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., the English ethnicity 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

[20]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.

[21]Wooldridge (Ch. 19, 2002) describes the statistical properties of the Negative Binomial regression.

category 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 first column of Table 1.3 reports the incidence rate ratio for all citations, finding 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 coefficient is statistically different from one, the level where own-ethnicity citations have the same frequency as citations of other ethnicities. 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 are also reported in Table 1.3. Common ethnicity appears most important for international technology diffusion in the first few years after a patent is made, peaking in a citation lag of four to five years.[22]

1.4 Output and Productivity Analysis

The international patent citation exercises confirm knowledge diffusion occurs at an uneven rate across countries and further verify that knowledge networks are important for short-run technology transfer from the US. The focus of this study, however, is whether greater tacit knowledge with respect to all US innovations translates into economic improvements for foreign countries. To evaluate this proposition, the US ethnic patenting trends are joined with additional data on foreign manufacturing industries. The combined dataset is first described, and an empirical extension of specification (1.8) that accounts for the features of the combined dataset is developed and estimated.

[22] Thompson and Fox-Kean (2004) criticize the USPTO categories as being too broad to control effectively for technology specialization. Thompson (2005) proposes an alternative approach that compares inventor-added citations to those added by the USPTO examiner, a distinction only made after 2000. Estimations using Thompson's technique and dataset yield a quantitatively similar role for own-ethnicity in international citations.

Jaffe, Trajtenberg, and Fogarty (2002) provide additional documentation on inferring communication channels from patent citations, while Jaffe, Trajtenberg, and Henderson (1993), Peri (2004), Hu and Jaffe (2004), and Agrawal, Cockburne, and McHale (2004) are examples of applications in an international distance context.

1.4.1 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 1.4 describes the sample and lists the three-digit ISIC industries.[23] 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 1.4 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.

The UNIDO dataset is inappropriate for studies of industry creation or destruction due to its unbalanced panel and industry aggregation. Recognizing this limitation and in order to enhance the quality of estimations, country-industry observations must maintain ten employees and one US ethnic patent per annum. These minimums exclude poor quality data, but raising or removing these hurdles does not significantly affect the findings.[24]

[23]The UNIDO collects data at the three-digit and four-digit industry levels of the International Standard Industrial Classification (ISIC). This presentation focuses on the three-digit aggregation, but the four-digit data delivers similar results. While sacrificing industries (28 versus 80), the three-digit dataset contains more countries (43 versus 20), better coverage of Chinese economies, and more capital data.

[24]Specifications employing alternative UNIDO data on industry value-added and establishments mirror the output and employment results presented below. The 1985-1997 period balances data inclusion with maintaining a consistent sample, as data for earlier or later years are quite limited. Similar outcomes are evident if all 1980-2000 data are employed or if the sample is restricted to a 1985-1997 balanced panel of continually surveyed countries and industries.

1.4.2 Output and Productivity Estimation Framework

The combined dataset affords an industry-level analysis of technology transfer with multiple countries and ethnicities. Extending (1.8) to industry i and country c of ethnicity e,

$$\ln(Y_{ci}) = \phi_{ci} + \ln(\tilde{I}_i) + \beta \ln(\tilde{L}_{R,ei}). \tag{1.13}$$

While analytically convenient, this steady-state description must be adapted for the empirical exercises. In particular, the ethnic human-capital stocks for US technologies change over the 1985-1997 period (and are indeed the source of identification for the β parameter). The citation regressions in Table 1.3 highlight that ethnic ties have an important lag structure, especially for the first five years of knowledge dissemination. Rewriting (1.13) in discrete time to model this five-year dependency,

$$\ln(Y_{cit}) = \phi_{ci} + \ln(\tilde{I}_{it}) + \beta \ln\left(\sum_{s=1}^{5} \tilde{L}_{R,ei,t-s}\right). \tag{1.14}$$

Unfortunately, existing data on the ethnicity of the US scientific workforce are limited. While Census estimates provide strong cross-sectional descriptions (i.e., country-of-birth, industry), they lack the necessary longitudinal detail. Other resources like the Current Population Survey and National Science Foundation reports offer only coarse cross-sectional distinctions. Ethnic patenting data, however, are a solid foothold for estimating the scientific research of ethnicities in the US. 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 (1.14),

$$\ln(Y_{cit}) = \phi_{ci} + \ln(\tilde{I}_{it}) + \beta \ln\left(\tilde{I}_{it_0}^{-1} \sum_{s=1}^{5} \tilde{I}_{ei,t-s}^{Flow}\right).$$

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

$$\ln(Y_{cit}) = \alpha + \beta \ln(PAT_{eit}^{US}) + \phi_{ci} + \eta_{it} + \epsilon_{cit}, \tag{1.15}$$

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 (1.15) 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 trends common to the countries in the sample (e.g., higher worldwide demand for computers and pharmaceuticals) and fluctuations in patent statistics due to changes in USPTO resources (Griliches 1990).[25]

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

[25] Industry-year effects extract industry-specific price movements (e.g., the rapid decline of computer prices). The UNIDO converts output data from foreign currencies to nominal US dollars using average yearly exchange rates (IMF International Financial Statistics Series rf).

over the last two decades is difficult to interpret.[26] Instead, (1.15) 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.

Finally, the five-year patent sums PAT_{eit}^{US} are developed for each ethnicity-industry from the patent database. Multiple nations map into the nine ethnicities available with the ethnic-name database, and the same industry-level patenting series from the US is applied to each country within an ethnicity (i.e., Mexican or Chilean scientists cannot be separated from the Hispanic total). 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, Duflo, and Mullainathan (2004). Robustness checks also examine whether the large European and Hispanic blocs significantly influence the results.[27,28]

1.4.3 Basic Output and Productivity Regressions

Table 1.5A reports the primary results for (1.15). The first row demonstrates that output consistently rises with strong scientific integration to the US. As both variables in logs, the 0.241 coefficient in the upper-left corner finds a 0.24% increase in foreign output with a 1% increase in US ethnic research. As discussed earlier, 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 patent-weighted regressions are consistently larger than the unweighted

[26] For example, Kortum and Lerner (2000), Kim and Marshcke (2004), and Hall (2004).

[27] Some country to ethnicity mappings are debatable (e.g., placing Spain and Portugal with European rather than Hispanic, including the Scandinavian countries in European), as is the inclusion of communist countries. The results are robust to these marginal reclassifications. The ethnic matching procedure does not attempt to distinguish the Filipino ethnicity from Hispanic, as was done in an earlier version of this paper, but the results are very similar if the division is made.

[28] The five-year sum gives equal weight to each year. Regressions weighting the lagged community sizes by the coefficients from the international citation exercises yield similar results.

regressions as this scheme 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 marginally lower for the output weights than the patent weights due to their greater emphasis on traditional economic sectors (e.g., food products, textiles). Both approaches, however, yield more consistent results by focusing attention on larger countries and industries.

Many empirical analyses first difference the levels specification (1.15) for estimation,

$$\Delta \ln(Y_{cit}) = \alpha + \beta \Delta \ln(PAT_{eit}^{US}) + \eta_{it} + \hat{\epsilon}_{cit}, \tag{1.16}$$

where $\hat{\epsilon}_{cit} = \epsilon_{cit} - \epsilon_{cit-1}$. 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. Table 1.5B demonstrates that the first-differences form yields similar results to the levels specification; both specifications are presented below.

1.4.4 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 1.6 undertakes three such tests, finding continued support for output expansion with stronger scientific integration.

Panel A begins by replicating the base foreign output regressions from Table 1.5. 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 National Science Foundation 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. (These Ph.D. trends and the reverse causality question are extensively studied in Section 1.5's immigration analysis.)

More generally, Panels C and D incorporate into (1.15) 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 and trend manufacturing gains, countries entering trade agreements or multi-national bodies (e.g., World Trade Organization), 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 world-wide industry trends are removed). The triple combination of country-industry, industry-year, and country-year fixed effects is a very stringent test, as much of the variation is removed from the sample. While the positive correlations are lost in the unweighted specifications, the weighted regressions continue to support the conclusion of foreign output growth with stronger scientific integration. The decline in coefficient magnitudes suggests, moreover, that a substantial portion of the growth in each ethnicity's US research community is uniform across industries (i.e., the within-industry Chinese contributions to US computer and pharmaceutical research expand at similar rates).

1.4.5 Foreign Country Capital-Labor Controls

Table 1.7 next explores the role of capital development in explaining the labor productivity growth evident in Table 1.5. Section 1.2's theory only models non-durable intermediate inputs, a simplification that removes the need to track two state variables. Labor productivity grows with capital deepening as well as technology adoption, however, and it is important to distinguish the two. The UNIDO data unfortunately lack capital records for several countries.[29] Thus, Panel B of Table 1.7 first re-estimates the basic labor productivity regression with the observations that have capital data. The results are close to the Full Sample replicated in Panel A, although some statistical significance is lost with the reduced sample size. Panel C finally incorporates into the estimating equation the contemporaneous log capital-labor ratio K_{cit}/L_{cit},

$$\ln(Y_{cit}/L_{cit}) = \alpha + \beta \ln(PAT_{eit}^{US}) + \gamma \ln(K_{cit}/L_{cit}) + \phi_{ci} + \eta_{it} + \epsilon_{cit},$$

The β coefficients in Panel C are quite stable to the introduction of capital stocks, indicating that the technology transfer operates beyond just capital accumulation.[30]

1.4.6 Sector Reallocation and Sample Composition

The theoretical model in Section 1.2 delineates how the benefits of technology transfer depend upon the following country's stage of development. In the specified framework, countries with large agricultural sectors realize gains in output from technology transfer due to both labor productivity development and employment reallocation. In an alternative framework where the outside wage is held constant, the output growth comes only through labor shifts. On the other hand, following economies with full manufacturing employment only realize output growth through labor productivity enhancements.

To test these predictions, Table 1.4 lists the 1980 share of national employment in agriculture for each economy.[31] The three smallest agricultural sectors are found in Hong Kong (1%),

[29]Sufficient capital data are only available for the countries noted in Table 1.4 and do not always cover the years listed in the UNIDO3 Panel column. Capital stocks are estimated using the perpetual inventory method with a depreciation rate of 15%. Initial stocks are developed using 1980 and 1981 investments, and subsequent investments are deflated using weighted deflators taken from the NBER-CES Manufacturing Productivity Database (Bartelsman and Gray 1996). Breaks in the capital series for Chile (1987, 1988), Macao (1987), Mexico (1992, 1993), Panama (1986), and Peru (1993) are bridged in the reported regressions; the results are robust to instead dropping the years after the breaks.

[30]Technology improvements and investments may occur together if new technologies are embodied in machines that are purchased and installed. Additional tests suggest embodied technical change may be present, but the results are not consistent across specifications. Foreign investment may be important for realizing the productivity gains from knowledge ties to the frontier country.

[31]Agricultural shares are from the United Nations Statistical Division and Sun, Fulginiti, and Peterson (2003).

Singapore (2%), and Belgium (3%), while the three largest sectors are India (70%), Vietnam (73%), and Mainland China (74%). A modified form of (1.15) interacts the ethnic scientific community regressor with this pre-period agricultural share,

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

where the main effect for the agricultural share is absorbed into the country-industry fixed effects. A positive γ coefficient indicates output growth due to scientific integration is stronger in countries with larger agricultural workforces in 1980.[32]

Tables 1.8A and 1.8B report the results from these interacted regressions. In both the levels and first-differences specifications, 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, 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.

Finally, the UNIDO dataset is a diverse group of countries and industries, and it is informative to identify which observations are most responsible for the aggregate findings. Table 1.8 investigates this question for the patent-weighted regressions in Columns 4-8. 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), it would be important to note if ethnic differences in high-tech industries alone are responsible for the positive correlations. The fourth column excludes the computer and pharmaceutical industries from the Full Sample and finds that while the main effect on output declines, the coefficient pattern is very similar. In general, dropping these two industries from the samples below does not significantly affect the outcomes discussed.

Chinese economies, more often than not, are also the centerpieces of technology transfer

[32] Both of the main effects are demeaned prior to the interaction to restore the β coefficient to close to its base level.

stories. Column 5 excludes Mainland China from the sample and finds very similar results. The stability of the interactions is especially comforting as Mainland China had the largest 1980 agricultural share of the sample. Unreported regressions further find that the parameter estimates do not depend significantly on the inclusion of any one country in the sample. Column 6 demonstrates, however, that excluding the full Chinese ethnicity can be important. In the both the levels and first-differences specifications, the main effect of output increasing is lost due to opposite movements in labor productivity and employment. 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 main effect is sensitive to their inclusion. Reassuringly for the sector reallocation finding, the interactions remain in their predicted directions in both specifications despite excluding the three most extreme economies.

The Full 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 patents were awarded to Japanese inventors living outside of the US during this period.[33] Positive correlations of foreign productivity 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, Columns 7 excludes Japan, European countries, and Russia from the Full Sample. The sharper contrast of the Chinese and Hispanic economies increases the productivity main effects, but the overall coefficient pattern is again evident. Finally, the UNIDO descriptive statistics noted that European and Hispanic countries each account for about 40% of the sample. The last columns drops Hispanic countries, finding results similar to the Full Sample estimations, although the output interaction is diminished.

In summary, the sample composition adjustments find positive benefits to scientific integration with the US are evident throughout the panel studied. While manufacturing output growth is pervasive, it is especially strong in economies with large agricultural sectors that facilitate sector reallocation. In countries with minimal agricultural employment, output growth comes through labor productivity gains.

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

1.5 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 1.2 following an industrialization push in China. China's government temporarily subsidizes invention until condition (1.9) no longer holds. As $I > \tilde{I}\Psi[M/\tilde{I}](\tilde{H}^C)^\beta$, it is more profitable for researchers in China to invent rather than imitate; China's output growth and sector reallocation are now driven solely by domestic innovations. In the US, Chinese researchers 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 US Chinese can register their imitations with the US patent office, a positive β coefficient will be found in the core estimating equations even though China's manufacturing gains no longer depend on its research community in the US. In fact, data trends will show contemporaneous accelerations in the growth of foreign output and US ethnic patenting.[34]

The US population of Chinese researchers is a foothold for establishing greater confidence in the direction of technology flows as they only influence China's development through their transmission of knowledge regarding US innovations. If the size of this research population is

[34] China's economy still depends on previously imitated products, as well as new inventions. How the system evolves from this initial disturbance depends on the relative populations of the Chinese researchers in the US and China. If the US Chinese group is sufficiently small, they will continue imitating a large invention stock developed abroad forever, and the Chinese human-capital stock with respect to US inventions will decline to zero. China's researchers will continue inventing, and the gap between China's researcher productivity for invention versus imitation will become entrenched. On the other hand, if the US Chinese research community is sufficiently large relative to China's, the declining imitation productivity will require at least some US Chinese resume direct invention to maintain full employment. In this scenario, the initial reverse technology flows yield to either a sustained mixing strategy, with Chinese researchers in both countries inventing and imitating, or the US Chinese resuming the leading role.

exogenously determined by immigration restrictions, a reduced-form strategy for the size of the ethnic research community can be developed within the quotas system. 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 Chinese inflows of researchers to the US to be $QUOTA_{R_C,t}$. Assuming that only the previous three years of immigration matter for a research stock[35], a reduced-form immigration estimator for ethnic scientific integration to the US is modelled as

$$\ln(IMM^{RF}_{R_C,t}) = \ln\left[\sum_{s=1}^{5}(QUOTA_{R_C,t-s} + QUOTA_{R_C,t-s-1} + QUOTA_{R_C,t-s-2})\right]. \quad (1.17)$$

The summation over the previous five years maintains the human-capital modelling technique employed with the ethnic patenting dataset. This section designs and implements an empirical version of (1.17) using exogenous changes in US immigration quotas.

Before proceeding, it is worth outlining why the US quotas are employed for a reduced-form estimator rather than in an instrumental-variables specification. The unobserved regressor in this study is the human-capital stock of each ethnicity with respect to US technologies. The patent metrics employed in Section 1.4 proxy, albeit imperfectly, for this scientific integration. Immigration quotas directly influence the size of ethnic research communities in the US, and thus the unobserved human-capital stocks. Scientific bonds, however, can operate through other channels besides formal patenting and the informal or tacit knowledge of new technologies that the patent metrics represent. As the exclusion restriction required for two-stage least squares does not hold, coefficient estimates from using the immigration estimator as an instrumental variable would be upward biased. The reduced-form approach, however, offers a direct check for the patent-based findings using exogenous changes in the populations of immigrant researchers living in the US.

1.5.1 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%

[35] The immigration reform examined below focuses on a very sharp surge in immigration that makes this assumption reasonable.

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.[36] 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).

US immigration law applies two distinct quotas to numerically restricted immigrants.[37] 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.[38] 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.[39] Moreover, the relative admissions of high-skilled professionals to low-skilled workers significantly increased within the employment-based admissions.

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. Figures 1.4 and 1.5 use records from the Immigration and Naturalization Service (INS) to detail the response. Figure 1.4 plots the number of ISEs granted permanent residency in the US from 1983-1997 for selected eth-

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

[37]US immigrants are admitted through numerically restricted categories, governed by the quotas discussed in this section, and numerically unrestricted categories (e.g., immediate relatives of US citizens). The reduced-form estimator centers on the numerically restricted categories that admit 75% of ISEs (versus 43% of all immigrants). Jasso, Rosenzweig, and Smith (1998) outline US immigration policy and the 1990 Act; they further discuss behavioral responses to changes in quotas. ISE inflows through the unrestricted categories are stable in the years surrounding the 1990 reform.

[38]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.

[39]The employment limit increased to 140,000, but 120,120 corresponds to the previously restricted categories.

nicities (summed over 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. Figure 1.5, on the other hand, shows low-skilled immigration during the same period. While Chinese and Indian immigration are substantially higher than Hispanic immigration for science and engineering, the opposite is true for low-skilled immigration. Moreover, low-skilled immigration did not respond to the 1990 Act.[40]

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, National Science Foundation 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 1.6 exhibits the percentage intending to remain in the US 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.

[40]Immigration trends are developed from immigrant-level INS records. The permanent residency admissions include ISEs already working in the US on temporary visas. The trends for "new arrival" ISE are very similar. Temporary visas can only be renewed once, so the total shift in ISE population should include workers gaining permanent residency. The analysis below does not depend on this distinction. Science and engineering categories are defined as Engineers, Natural Scientists, and Mathematical and Computer Scientists; low-skilled categories are Administrative Support, Farming, Laborer, Precision Production and Repair, Service, and Sales occupations.

1.5.2 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 1.4 through 1.6 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. Let $ISE\%_{ct_0}^{Adm}$ be the mean ISE arrivals from country c divided by an approximate country-level numerical limit for employment-based workers during the 1983-1990 pre-period. The theoretical numerical limit is taken to be the 20,000 country limit multiplied by the 20% worldwide allocation given to employment-based applications (i.e., 54,000/270,000).[41]

Define $POST_t$ as a indicator variable taking the value of zero from 1983-1990 and one for 1991 and after (i.e., the 1990 Act's effective date). Regressing annual ISE admissions ISE_{ct}^{Adm} on an interaction of $ISE\%_{ct_0}^{Adm}$ with $POST_t$ quantifies the immigration response of constrained countries,

$$ISE_{ct}^{Adm} = \alpha + \gamma ISE\%_{ct_0}^{Adm} \cdot POST_t + \phi_c + \eta_t + \epsilon_{ct}. \tag{1.18}$$

The main effect for $ISE\%_{ct_0}^{Adm}$ is absorbed by the country fixed effects ϕ_c, along with levels differences between nations in US immigration. The year effects η_t remove aggregate changes in US permanent residency admissions and control for the main effect of $POST_t$.

The γ coefficient in (1.18) will be positive and significant if raising the two numerical limits spurred ISE immigration from previously constrained countries (i.e., high values of $ISE\%_{ct_0}^{Adm}$). Table 1.9 shows this to be true, and economies with high values of $ISE\%_{ct_0}^{Adm}$ become the treatment group regardless of actual responses. From the waiting list and 1983-1990 flow data presented in Table 1.9, the treated groups are determined to be India, Mainland China, the Philippines, and Taiwan.[42] The reduced-form immigration estimator (1.17) then takes the

[41] The total employment immigration column in Table 1.9 demonstrates the theoretical limit works quite well. The scientific percentages are even larger than they initially seem since family members of employment-based admissions count towards the two quotas. The specific years selected for the pre-period are not important.

[42] Hong Kong is not included in the treatment group as its immigration status was not affected by the 1990 reform. The main results are robust to instead defining the treatment group at the ethnicity level, although the additional variation inherent in the country-level approach enhances performance in falsification exercises.

form

$$\ln(IMM_{cit}^{RF}) = \ln\left[\sum_{s=1}^{5}(QUOTA_{c,t-s}^{Eff} + QUOTA_{c,t-s-1}^{Eff} + QUOTA_{c,t-s-2}^{Eff})\right], \tag{1.19}$$

where $QUOTA_{ct}^{Eff}$ 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 (i.e., low $ISE\%_{ct_0}^{Adm}$), and their effective quotas are held constant at the pre-reform theoretical limit. For constrained countries with high $ISE\%_{ct_0}^{Adm}$ values, 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 the shift 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 research productivity, and others. If these types of concerns are overwhelming, regressions of US ethnic patenting on the reduced-form estimator will yield weak coefficients. The right-hand side of Table 1.9 shows instead that they are quite strong 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 Table 1.9 shows 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. Figure 1.7 gives some weight to this omitted variable concern for Mainland China. Mainland China was on a clear upward trend in science and engineering Ph.D. graduates in the US prior to the 1990 Act.[43] A similar expansion of researchers at home is likely and may have directly impacted manufacturing development. These concerns are evaluated empirically below.

1.5.3 Reduced-Form Results

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

$$\ln(Y_{cit}) = \alpha + \beta \ln(IMM_{ct}^{RF}) + \phi_{ci} + \eta_{it} + \epsilon_{cit}, \tag{1.20}$$

with $\ln(IMM_{ct}^{RF})$ defined by (1.19). Table 1.10 exhibits the main results in a format similar to that of Table 1.5. The reduced-form estimator suggests foreign output increases with an elasticity of 0.3-0.4 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 1.10's estimates across weighting schemes 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-

[43] On a logarithmic scale, Figure 1.7 exhibits a smooth trend for Mainland China from 1985-1991 with a marginal decrease in the growth rate thereafter.

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 1.11 next turns to robustness checks on the output growth finding, with the first row simply replicating the core regression set. As a test of the foreign human-capital development story, Panel B incorporates Figure 1.7's trends in foreign graduates from US science and engineering Ph.D. programs. Both the levels and first-differences specifications hold up well in the augmented specification. Given the specific concern regarding Figure 1.7's trend growth in Mainland China's Ph.D. graduates, it is reassuring that this country can be again excluded in Panel C with only minor shifts in the outcomes. As before, the results are also robust to dropping any other country, the computer and drug industries, the full Chinese ethnicity, advanced or Hispanic economies, and so forth.[44]

Finally, 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 reform for the identification of the β parameter. Despite losing about about half of their size, the coefficients remain economically and statistically significant in the augmented specification. Given the stringency of this test, this strong performance provides confidence against the estimator reflecting a spurious correlation.

Table 1.12 completes the immigration analysis by incorporating into (1.20) 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 60%-90% of their value and are still statistically different

[44]The robustness of the Full Sample results to excluding India is important. The INS quotas design does not consider shifts in the US ethnic populations of temporary workers (e.g., the H-1B program). The temporary visa program up to the mid-1990s looked quite different from today. Fewer visas were issued, and the most significant occupation and country were medical professionals and the Philippines, respectively. An explosion in Indian temporary workers, mostly for systems analysis and computer programming jobs, began in the 1990s. From 1989-1999, India's share of temporary visas issued rose from 9% to 48% (e.g., Lowell 2000).

Temporary visas can only be renewed once, for a maximum stay of six years, so long-term growth in ethnic research communities requires permanent immigration. Outside of India, the trends for temporary visas are fairly stable for the period studied, and the science and engineering component appears small compared to permanent residency changes. While the jump in India's temporary visa community could affect the final few years of the 1985-1997 period, the results do not depend on its inclusion.

from zero. Moreover, the standard errors for the placebo estimators are 100%-300% larger than those of the true estimator, and the placebo estimators 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 E, on the other hand, shows better performance with the 1995 counterfactual. Table 1.12 thus supports the conclusion of stronger scientific integration leading to foreign output growth, but also highlights 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 robustness checks on sample composition, ethnic time trends, and so on strongly suggest spurious correlations are not solely responsible for the outcomes measured. Overall, the reduced-form regressions support Section 1.4's conclusion that foreign manufacturing output increases with stronger ethnic scientific integration to the US frontier.

1.6 Conclusions

The international diffusion of new innovations from frontier countries is necessary for broad economic development. Even when the codified details of new technologies can be easily disseminated, successful adoption may be complicated by the difficult exchange of the associated tacit knowledge. This project considers the role and importance of knowledge networks for exchanging this practical information through the observable channel of ethnicity, specifically examining the ties between US ethnic research communities and their home countries. A new tool is developed for studying the role of ethnic scientists and engineers in the US in the technology transfer process by applying an ethnic-name database to individual patent records. The resulting cross-sectional and longitudinal detail affords new insights about how knowledge diffuses across countries.

First, ethnic knowledge networks are important for explaining international patent citation patterns, with inventors living outside of the US citing US-based inventors of their own ethnicity with a 50% higher rate. The core specifications further suggest a stronger US ethnic research community boosts foreign manufacturing output with an elasticity of about 0.3. These estimates are robust to multiple specification checks, and the pattern of results is consistent with a model of sector reallocation from agriculture to manufacturing. In economies with large agricultural sectors, manufacturing output expansions occur primarily through employment growth; in advanced economies with full manufacturing employment, the output gains are achieved via higher labor productivity. Finally, a reduced-form strategy using exogenous changes in US immigration law also finds qualitatively similar effects. These findings suggest that US ethnic communities play an important role in technology diffusion to their home countries, and more generally that inadequate access to the tacit knowledge complementing new frontier innovations can slow development in following countries.

This paper is a first step for characterizing the complex role of tacit knowledge in technology diffusion, and two promising extensions are currently being pursued with the ethnicity approach. One project examines the relative importance of trade and FDI channels for the technology transfer considered in this study.[45] Second, a companion study explores ethnicity's role in local knowledge diffusion within the US using the ethnic citation data. Special attention is given to high-tech clusters like Silicon Valley and Boston's Route 128, and ongoing research concentrates on the linkages between these high-tech clusters and foreign inventors (e.g., international collaborations between inventors of the same ethnicity). These extensions will further characterize how knowledge networks shape the effective technology frontier for emerging economies.

[45] Rauch (2001) discusses the importance of business and social networks in trade, and Rauch and Trindade (2002) demonstrate a substantial trade boost from Chinese networks. Technology diffusion is also facilitated by foreign direct investment and multinational enterprises (e.g., Branstetter 2004, Singh 2003).

Bibliography

[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", NBER Working Paper 9950 (2004).

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

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

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

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

[7] Bartelsman, Eric, and Wayne Gray, "The NBER Manufacturing Productivity Database", NBER Technical Working Paper 205 (1996).

[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] Blonigen, Bruce, and Rossitza Wooster, "CEO Turnover and Foreign Market Participation", NBER Working Paper 9527 (2003).

[10] Borjas, George, "Do Foreign Students Crowd Out Native Students from Graduate Programs?", NBER Working Paper 10349 (2004).

[11] Branstetter, Lee, "Is Foreign Direct Investment a Channel of Knowledge Spillovers? Evidence from Japan's FDI in the United States", Working Paper (2004).

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

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

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

[15] Dumais, Guy, Glenn Ellison, and Edward Glaeser, "Geographic Concentration as a Dynamic Process", CES Working Paper 98-3 (1997).

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

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

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

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

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

[21] Hall, Bronwyn, "Exploring the Patent Explosion", NBER Working Paper 10605 (2004).

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

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

[24] 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).

[25] Hu, Albert, and Adam Jaffe, "Patent Citations and International Knowledge Flow: The Cases of Korea and Taiwan", *International Journal of Industrial Organization*, forthcoming 2004.

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

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

[28] 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.

[29] 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.

[30] Jasso, Guillermina, Mark Rosenzweig, and James Smith, "The Changing Skill of New Immigrants to the United States: Recent Trends and Their Determinants", NBER Working Paper 6764 (1998).

[31] Johnson, Daniel, "150 Years of American Invention: Methodology and a First Geographic Application", Wellesley College Economics Working Paper 99-01 (1999). Data currently reside at http://faculty1.coloradocollege.edu/~djohnson/uships.html.

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

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

[34] Kannankutty, Nirmala, and R. Keith Wilkinson, "SESTAT: A Tool for Studying Scientists and Engineers in the United States", National Science Foundation SRS Special Report (1999).

[35] Kapur, Devesh, and John McHale, "Sojourns and Software: Internationally Mobile Human Capital and High-Tech Industry Development in India, Ireland, and Israel", Working Paper (2003).

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

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

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

[39] Kim, Jinyoung, and Gerald Marschke, "Accounting for the Recent Surge in U.S. Patenting: Changes in R&D Expenditures, Patent Yields, and the High Tech Sector", *Economics of Innovation and New Technologies*, 13:6 (2004), 543-558.

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

[41] Kortum, Samuel, and Joshua Lerner, "Assessing the Contribution of Venture Capital to Innovation", *RAND Journal of Economics*, 31:4 (2000), 674-692.

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

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

[44] Lowell, B. Lindsay, "H1-B Temporary Workers: Estimating the Population", The Center for Comparative Immigration Studies Working Paper 12 (2000).

[45] 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.

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

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

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

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

[50] Peri, Giovanni, "Determinants of Knowledge Flows and their Effect on Innovation", Working Paper (2004).

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

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

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

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

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

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

[57] Romer, Paul, "Increasing Returns and Long-Run Growth", *The Journal of Political Economy*, 94:5 (1986), 1002-1037.

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

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

[60] 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).

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

[62] Scherer, Frederic, "Using Linked Patent Data and R&D Data to Measure Technology Flows," in Griliches, Zvi (ed.) *R & D, Patents and Productivity* (Chicago, IL: University of Chicago Press, 1984).

[63] 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.

[64] Singh, Jasjit, "Multinational Firms and International Knowledge Diffusion: Evidence from Patent Citation Data", Working Paper (2003).

[65] 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.

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

[67] Sun, Ling, Lilyan Fulginiti, and Wesley Peterson, "Accounting for Agricultural Decline with Economic Growth in Taiwan", Working Paper (2003).

[68] Thompson, Peter, "Patent Citations and the Geography of Knowledge Spillovers: What do Patent Examiners Know?", Working Paper (2004).

[69] Thompson, Peter, and Melanie Fox-Kean, "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment", *The American Economic Review*, forthcoming (2004).

[70] Wooldridge, Jeffrey, *Econometric Analysis of Cross Section and Panel Data* (Cambridge, MA: MIT Press, 2002).

Figure 1.1: Transition Path Simulations for China

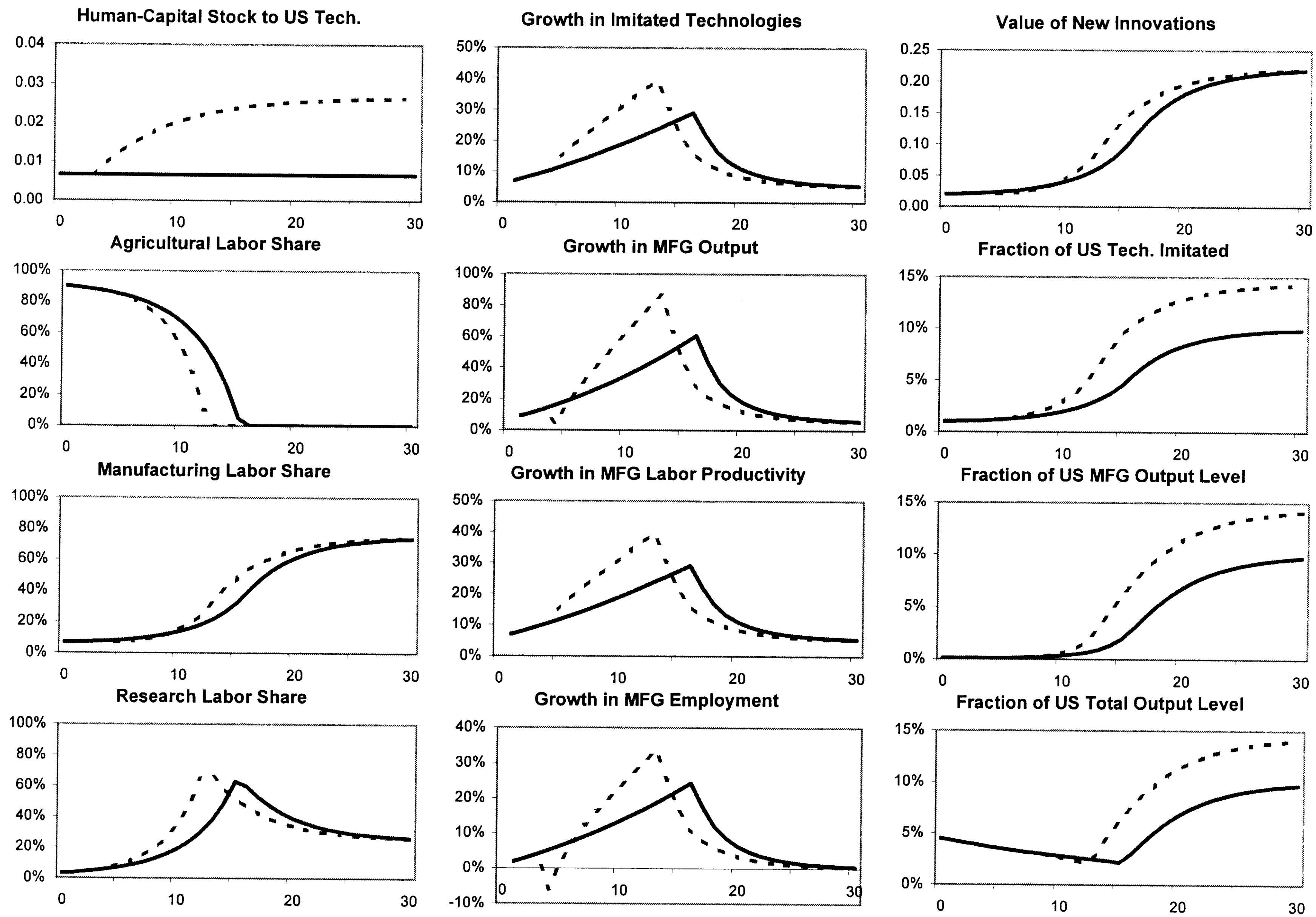


Figure 1.2: US Ethnic Patenting

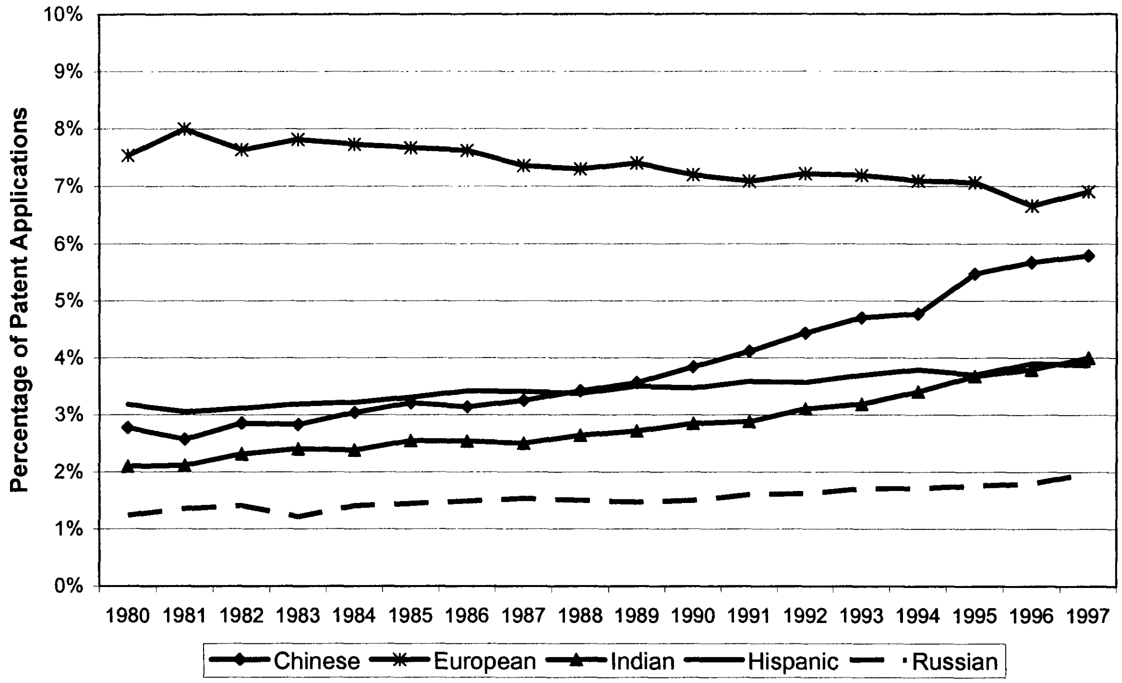


Figure 1.3: Ethnic Share by Technology

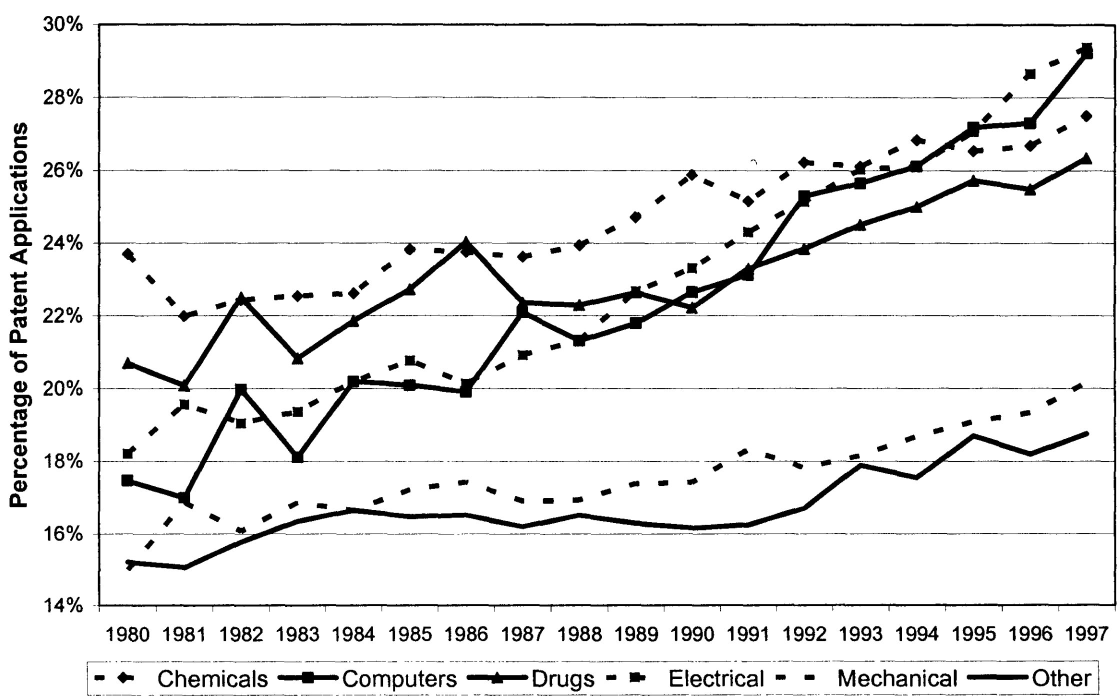


Figure 1.4: Science & Engineering Immigration

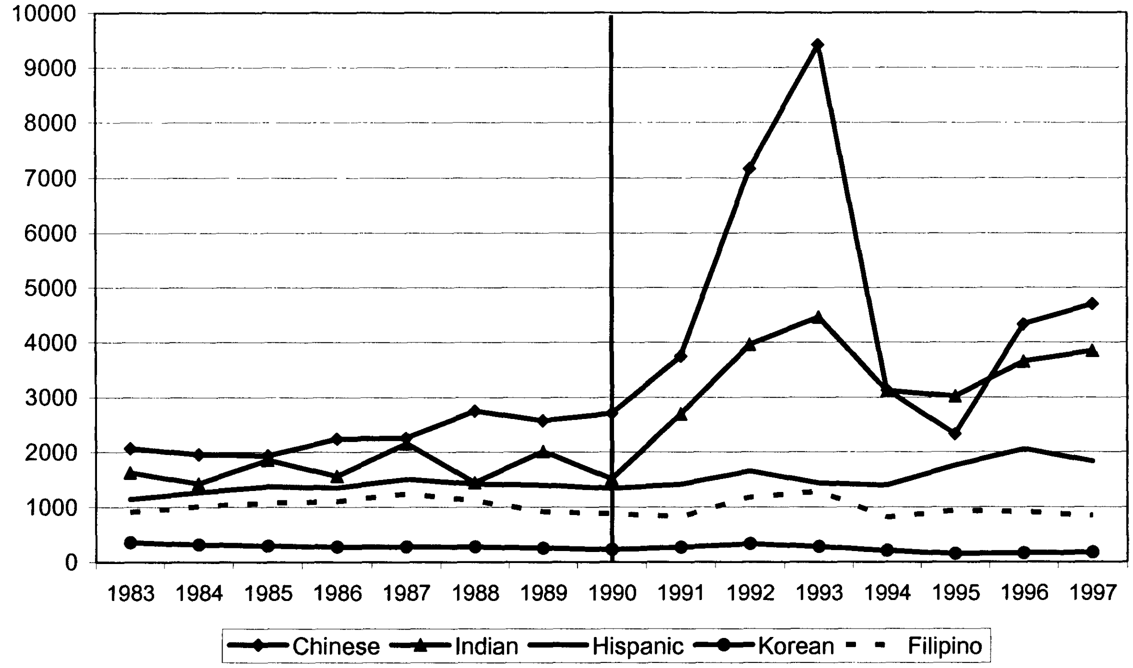


Figure 1.5: Low-Skilled Immigration

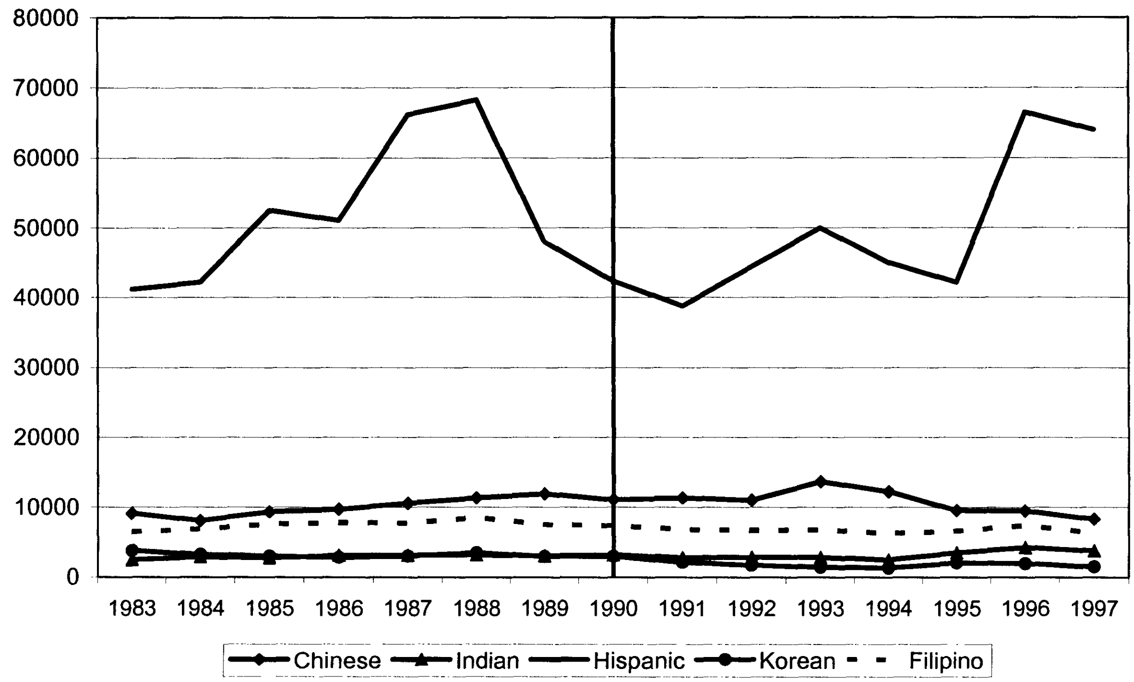


Figure 1.6: US SE Ph.D. Graduates Staying

Percentage of Graduates from Country Expecting to Stay in US

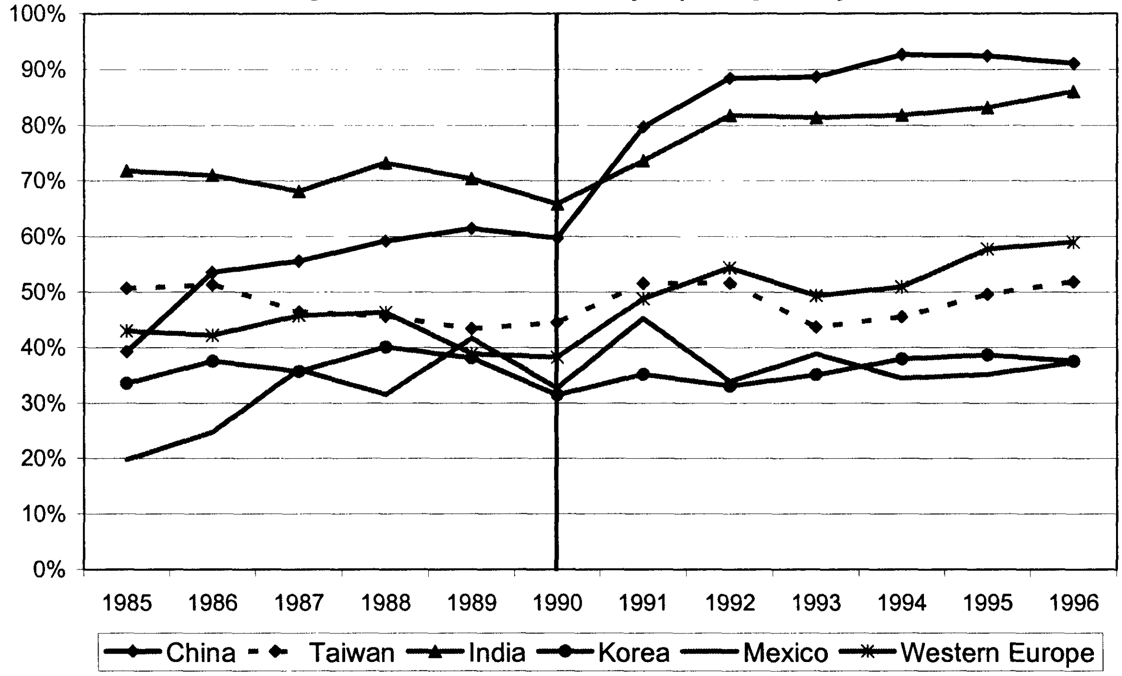


Figure 1.7: US SE Ph.D. Graduates

Graduates by Country of Origin

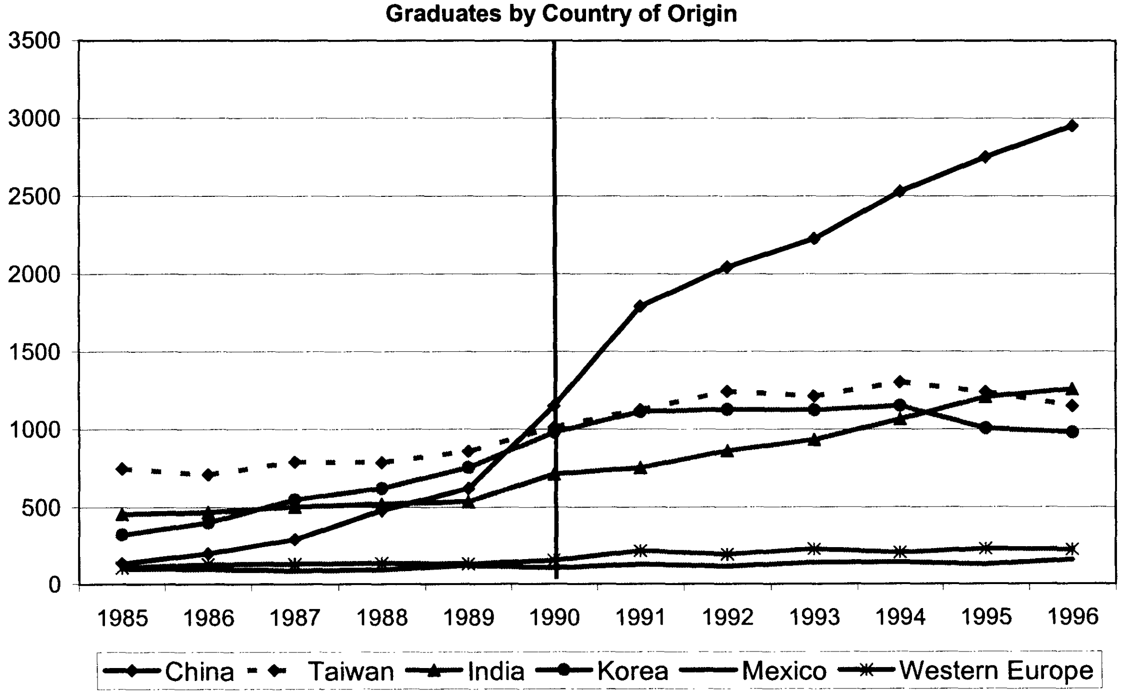


Table 1.1: Descriptive Statistics for US Ethnic Patents

	Ethnicity of Inventor (Percent Distribution)								
	English	Chinese	European	Hispanic	Indian	Japanese	Korean	Russian	Vietnamese
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)
1990 Bachelors	87.6	2.7	2.3	2.4	2.3	0.6	0.5	0.4	1.2
1990 Masters	78.9	6.7	3.4	2.2	5.4	0.9	0.7	0.8	1.0
1990 Doctorate	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 groupings listed in Table 4; English provides a residual for the Census statistics.

Table 1.2: Descriptive Statistics for Foreign Country Patent Records

		Match	Estimated Ethnicity of Country's Inventors (Percent Distribution)								
	Obs.	Rate	English	Chinese	European	Hispanic	Indian	Japanese	Korean	Russian	Vietnam.
United Kingdom	118,475	99	**87**	1	4	3	2	0	0	2	0
China, Singapore	33,309	99	3	**89**	1	1	0	1	5	0	1
Western Europe	757,185	96	18	1	**73**	6	1	0	0	2	0
Hispanic Nations	12,985	98	14	1	9	**72**	1	1	0	2	0
India	2,083	82	8	1	4	5	**80**	0	0	2	0
Japan	1,043,105	98	0	0	0	0	0	**100**	0	0	0
South Korea	26,553	100	4	11	0	0	0	1	**83**	0	0
Russia	29,870	94	5	1	2	9	0	0	0	**83**	0
Vietnam	9	100	5	18	16	0	0	0	0	0	**62**

Notes: China includes Mainland China, Hong Kong, Macao, and Taiwan. Western Europe includes Austria, Belgium, Denmark, Finland, France, Germany, Italy, Luxembourg, Nethlerlands, Norway, Poland, Sweden, and Switzerland. Hispanic Nations includes Argentina, Belize, Brazil, Chile, Columbia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Philippines, Portugal, Spain, Uruguay, and Venezuela. Russia includes former Soviet Union countries.

Table 1.3: Incidence Rate Ratios for Foreign Citations of US Domestic Patents

	All	Citations Separated by Citation Lags									
	Citations	Lag 0-1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
Incidence Rate Ratio	1.496	1.315	1.402	1.485	1.524	1.332	1.399	1.348	1.320	1.332	1.159
for Same Ethnicity	(0.052)	(0.089)	(0.085)	(0.085)	(0.088)	(0.079)	(0.087)	(0.086)	(0.088)	(0.092)	(0.084)

Notes: Regressions contain 116,640 cells as defined in the text. Regressions include fixed effects for ethnicity of citing inventor, ethnicity of cited inventor, technology class of citing inventor, technology class of cited inventor, and an indicator variable for same industry.

Table 1.4: UNIDO Industry Sample

Country	1980 Agr. Share	UNIDO3 Panel	Output (m) Level	Output (m) Growth	Labor Prod. (k) Level	Labor Prod. (k) Growth	Employment (k) Level	Employment (k) Growth	Capital (m) Level	Capital (m) Growth
Single Ethnic Mappings:										
India	70%	85-97	117,950	6%	16	3%	7,354	2%	46,740	4%
Japan	11%	85-97	2,053,048	7%	206	8%	9,998	-1%	415,195	8%
South Korea	37%	85-97	230,942	14%	88	13%	2,626	1%	88,873	14%
Russia	16%	93-97	109,729	12%	10	22%	11,685	-8%		
Soviet Union	16%	85-89	1,087,914	7%	35	8%	31,434	-1%		
Chinese Economies:										
China, Mainland	74%	85-97	327,173	11%	8	9%	38,940	3%		
Hong Kong	1%	85-97	30,520	3%	66	12%	535	-9%	6,628	3%
Macao	6%	85-97	1,209	8%	26	10%	49	-2%	235	1%
Singapore	2%	85-97	37,830	16%	117	12%	309	3%	8,477	8%
Taiwan	8%	85-96	145,055	11%	68	11%	2,141	0%		
European Economies:										
Austria	10%	85-97	73,524	5%	125	5%	595	0%	22,001	5%
Belgium	3%	85-92, 95-97	31,958	5%	131	7%	247	-2%	19,809	7%
Denmark	7%	85-91	38,198	9%	93	11%	411	-1%	8,788	7%
Finland	12%	85-97	52,510	4%	141	8%	386	-4%	18,868	1%
France	8%	85-96	517,276	8%	130	10%	4,006	-2%	107,758	4%
Germany	7%	91-97	870,625	7%	147	7%	5,920	0%		
Germany, East		85-92	233,905	12%	81	12%	2,902	0%		
Germany, West		85-89	734,523	12%	115	12%	6,391	0%	51,571	-6%
Italy	13%	85-94, 96-97	390,266	7%	134	7%	2,897	0%	79,391	6%
Luxembourg	5%	85-97	2,952	3%	137	5%	22	-1%	730	1%
Netherlands	6%	85-97	117,868	6%	178	7%	670	-1%	29,146	6%
Norway	8%	85-97	37,467	4%	149	6%	256	-2%	10,402	-1%
Poland	30%	90-97	54,895	6%	21	7%	2,650	-1%	18,749	1%
Sweden	6%	85-97	93,727	6%	140	7%	678	-1%	23,192	4%
Switzerland	6%	86-96	37,827	7%	142	8%	270	-2%		

Table 1.4: UNIDO Industry Sample (continued)

	1980	Unido 3	Output (m)		Labor Prod. (k)		Employment (k)		Capital (m)	
Country	Agr. %	Panel	Level	Growth	Level	Growth	Level	Growth	Level	Growth
Hispanic Economies:										
Argentina	13%	85-90, 93-96	66,160	11%	73	14%	938	-3%		
Bolivia	53%	85-97	1,474	7%	41	1%	36	6%		
Brazil	37%	90, 92-95	127,807	11%	61	17%	2,105	-5%		
Chile	21%	85-97	20,604	10%	72	5%	278	5%	3,964	9%
Columbia	40%	85-97	20,099	5%	41	3%	487	2%	4,917	-1%
Costa Rica	35%	85-97	3,264	5%	26	1%	127	4%		
Cuba	24%	85-89	10,531	-1%	20	-3%	524	2%	6,097	0%
Ecuador	40%	85-97	4,372	3%	41	2%	107	2%	2,797	1%
Honduras	57%	90-95	989	8%	12	-10%	90	22%		
Mexico	36%	85-97	61,612	4%	60	6%	1,021	-2%	11,111	2%
Panama	29%	85-94, 96-97	1,468	4%	44	3%	33	1%	445	-3%
Peru	40%	85-92, 94-96	13,944	8%	55	9%	255	-1%	2,320	5%
Philippines	52%	85-97	23,238	11%	27	6%	857	5%	5,512	4%
Portugal	26%	85-97	36,365	8%	43	9%	816	-1%		
Spain	18%	85-97	201,951	8%	108	7%	1,858	2%	35,005	7%
Uruguay	17%	85-97	4,648	6%	37	8%	130	-1%		
Venezuela	15%	85-97	24,174	1%	59	2%	417	0%	13,775	1%

Notes: Values are in 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. Belize, Dominican Republic, El Salvador, Guatemala, Latvia, Lithuania, Nicaragua, Paraguay, and Vietnam are not included due to lack of data. For countries in the sample, insufficient observations or severe quality concerns excluded observations in Bolivia (353 in 1985, 355 and 382 in 1987), Brazil (1985), Costa Rica (371, 385 in 1997), Ecuador (352 in 1994, 354 in 1995, 313 in 1997), Honduras (1981-1989), Hong Kong (369 in 1996), Macao (314) and Venezuela (314 in 1996, 371 in 1995). Series breaks are modeled for Argentina (1990), Austria (1985), China (1989), Denmark (1989), Italy (1994), Mexico (1993), and Portugal (1989) for distinct levels shifts over the 1985-1997 period usually due to changes in variable definitions.

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 1.5A: UNIDO Levels Specification

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
		A. Log Foreign Output	
Log US Ethnic Research Community	0.241	0.420	0.400
	(0.134)	(0.242)	(0.156)
		B. Log Foreign Labor Productivity	
Log US Ethnic Research Community	0.215	0.383	0.310
	(0.094)	(0.191)	(0.132)
		C. Log Foreign Employment	
Log US Ethnic Research Community	0.026	0.037	0.090
	(0.146)	(0.212)	(0.148)
Industry x Year FE	X	X	X
Country x Industry FE	X	X	X
Observations	9912	9912	9912

Notes: Standard errors are clustered at the ethnicity-industry level.

Table 1.5B: UNIDO First-Differences Specification

	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
	(0.057)	(0.135)	(0.075)
		B. Δ Log Foreign Labor Productivity	
Δ Log US Ethnic Research Community	0.087	0.214	0.217
	(0.050)	(0.116)	(0.073)
		C. Δ Log Foreign Employment	
Δ Log US Ethnic Research Community	0.003	0.127	0.068
	(0.037)	(0.085)	(0.048)
Industry x Year FE	X	X	X
Observations	8736	8736	8736

Notes: Standard errors are clustered at the ethnicity-industry level.

Table 1.6A: UNIDO Country Controls - Levels

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Base Foreign Output Regression		
Log US Ethnic Research Community	0.241	0.420	0.400
	(0.134)	(0.242)	(0.156)
Observations	9912	9912	9912
	B. Including Foreign Ph.D. Students in US		
Log US Ethnic Research Community	0.225	0.294	0.337
	(0.268)	(0.229)	(0.265)
Log Foreign Ph.D. Students in US	0.016	0.100	0.054
	(0.093)	(0.089)	(0.097)
Observations	8914	8914	8914
	C. Including Country Time Trends		
Log US Ethnic Research Community	0.024	0.078	0.090
	(0.120)	(0.205)	(0.137)
Observations	9912	9912	9912
	D. Including Country-Year Effects		
Log US Ethnic Research Community	0.021	0.243	0.094
	(0.119)	(0.292)	(0.164)
Observations	9912	9912	9912
Industry x Year FE	X	X	X
Country x Industry FE	X	X	X

Notes: Standard errors are clustered at the ethnicity-industry level.

Table 1.6B: UNIDO Country Controls - First-Differences

	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
	(0.057)	(0.135)	(0.075)
Observations	8736	8736	8736
	B. Including Foreign Ph.D. Students in US		
Δ Log US Ethnic Research Community	0.061	0.313	0.210
	(0.035)	(0.075)	(0.066)
Δ Log Foreign Ph.D. Students in US	0.038	0.050	0.053
	(0.070)	(0.082)	(0.074)
Observations	7780	7780	7780
	C. Including Country Time Trends		
Δ Log US Ethnic Research Community	0.000	0.130	0.153
	(0.062)	(0.104)	(0.069)
Observations	8736	8736	8736
	D. Including Country-Year Effects		
Δ Log US Ethnic Research Community	-0.092	0.149	-0.022
	(0.049)	(0.109)	(0.060)
Observations	8736	8736	8736
Industry x Year FE	X	X	X

Notes: Standard errors are clustered at the ethnicity-industry level.

Table 1.7A: UNIDO Capital-Labor - Levels

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Base Foreign Productivity Regression		
Log US Ethnic Research Community	0.215	0.383	0.310
	(0.094)	(0.191)	(0.132)
Observations	9912	9912	9912
	B. Restricted Capital Sample		
Log US Ethnic Research Community	0.132	0.445	0.295
	(0.140)	(0.267)	(0.184)
Observations	5604	5604	5604
	C. Including Capital-Labor Ratio		
Log US Ethnic Research Community	0.121	0.332	0.238
	(0.128)	(0.257)	(0.170)
Log Foreign Capital-Labor Ratio	0.212	0.244	0.216
	(0.065)	(0.054)	(0.059)
Observations	5604	5604	5604
Industry x Year FE	X	X	X
Country x Industry FE	X	X	X

Notes: Standard errors are clustered at the ethnicity-industry level.

Table 1.7B: UNIDO Capital-Labor - First-Differences

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Base Foreign Productivity Regression		
Δ Log US Ethnic Research Community	0.087	0.214	0.217
	(0.050)	(0.116)	(0.073)
Observations	8736	8736	8736
	B. Restricted Capital Sample		
Δ Log US Ethnic Research Community	0.044	0.194	0.166
	(0.064)	(0.148)	(0.089)
Observations	4866	4866	4866
	C. Including Capital-Labor Ratio		
Δ Log US Ethnic Research Community	0.049	0.191	0.164
	(0.062)	(0.148)	(0.087)
Δ Log Foreign Capital-Labor Ratio	0.111	0.033	0.083
	(0.029)	(0.043)	(0.035)
Observations	4866	4866	4866
Industry x Year FE	X	X	X

Notes: Standard errors are clustered at the ethnicity-industry level.

Table 1.8A: UNIDO Sectoral Reallocation Regressions - Levels

				Patent Weights				
	No Weights	Patent Weights	Output Weights	Excluding Computers and Drugs	Excluding Mainland China	Excluding All Chinese Economies	Excluding Advanced Economies	Excluding Hispanic Economies
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Log Foreign Output								
Log US Ethnic Reseach Community	0.237	0.431	0.414	0.105	0.431	0.034	0.512	0.463
	(0.136)	(0.248)	(0.164)	(0.163)	(0.279)	(0.391)	(0.263)	(0.278)
Log US Ethnic Comm. x 1980 Agriculture Share	0.667	0.340	0.606	0.653	0.343	0.810	0.307	0.076
	(0.252)	(0.419)	(0.312)	(0.279)	(0.440)	(0.605)	(0.466)	(0.429)
B. Log Foreign Labor Productivity								
Log US Ethnic Reseach Community	0.220	0.360	0.294	0.261	0.294	-0.153	0.772	0.399
	(0.069)	(0.142)	(0.099)	(0.175)	(0.130)	(0.306)	(0.159)	(0.096)
Log US Ethnic Comm. x 1980 Agriculture Share	-0.855	-0.665	-0.654	-0.707	-0.876	-0.362	-0.512	-0.510
	(0.108)	(0.182)	(0.138)	(0.214)	(0.199)	(0.298)	(0.223)	(0.262)
C. Log Foreign Employment								
Log US Ethnic Reseach Community	0.017	0.072	0.120	-0.155	0.138	0.187	-0.260	0.063
	(0.122)	(0.177)	(0.114)	(0.100)	(0.200)	(0.198)	(0.172)	(0.197)
Log US Ethnic Comm. x 1980 Agriculture Share	1.521	1.005	1.260	1.360	1.220	1.172	0.818	0.586
	(0.220)	(0.336)	(0.252)	(0.273)	(0.334)	(0.416)	(0.340)	(0.286)
Industry x Year FE	X	X	X	X	X	X	X	X
Country x Year FE	X	X	X	X	X	X	X	X
Observations	9912	9912	9912	9067	9653	8669	6259	5453

Notes: Standard errors are clustered at the ethnicity-industry level.

Table 1.8B: UNIDO Sectoral Reallocation Regressions - First-Differences

	No Weights	Patent Weights	Output Weights	Patent Weights: Excluding Computers and Drugs	Patent Weights: Excluding Mainland China	Patent Weights: Excluding All Chinese Economies	Patent Weights: Excluding All Advanced Economies	Patent Weights: Excluding All Hispanic Economies
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Δ Log Foreign Output								
Δ Log US Ethnic Research Community	0.043	0.315	0.252	0.083	0.315	0.019	0.425	0.295
	(0.063)	(0.155)	(0.087)	(0.094)	(0.179)	(0.132)	(0.129)	(0.209)
Δ Log US Ethnic Comm. x 1980 Agriculture Share	0.765	0.442	0.647	0.754	0.322	0.793	0.349	0.283
	(0.188)	(0.359)	(0.247)	(0.193)	(0.357)	(0.386)	(0.377)	(0.388)
B. Δ Log Foreign Labor Productivity								
Δ Log US Ethnic Research Community	0.105	0.225	0.228	0.109	0.173	-0.083	0.498	0.219
	(0.048)	(0.108)	(0.069)	(0.106)	(0.097)	(0.119)	(0.102)	(0.099)
Δ Log US Ethnic Comm. x 1980 Agriculture Share	-0.284	-0.191	-0.216	-0.167	-0.572	-0.169	-0.237	-0.053
	(0.099)	(0.165)	(0.122)	(0.202)	(0.161)	(0.192)	(0.169)	(0.217)
C. Δ Log Foreign Employment								
Δ Log US Ethnic Research Community	-0.062	0.091	0.024	-0.026	0.142	0.102	-0.073	0.076
	(0.038)	(0.085)	(0.048)	(0.049)	(0.105)	(0.096)	(0.063)	(0.121)
Δ Log US Ethnic Comm. x 1980 Agriculture Share	1.049	0.633	0.863	0.922	0.894	0.962	0.586	0.336
	(0.149)	(0.271)	(0.202)	(0.190)	(0.264)	(0.308)	(0.290)	(0.241)
Industry x Year FE	X	X	X	X	X	X	X	X
Observations	8736	8736	8736	7991	8518	7616	5549	4821

Notes: Standard errors are clustered at the ethnicity-industry level.

Table 1.9: Immigration Quotas Reduced-Form Preliminaries

	Regressions of Immigration Response to 1990 Act (Thousands)				Regressions of Log US Ethnic Patents on Log US Immigration Quotas		
	Scientists	Business	Total		No Weights	Patent Weights	Output Weights
Base% x Post	4.669	4.842	3.279	Log US Immigration	0.217	0.256	0.213
	(0.183)	(0.127)	(0.125)	Quotas Estimator	(0.102)	(0.127)	(0.104)
Observations	2310	2310	2310	Observations	9912	9912	9912
	1983-1990 Percent of Theoretical Employment Quota for Country				**Employment Visa Waiting List January 1992**		
	Scientists	Business	Total		High-Skill	Skilled	Low-Skill
Hong Kong	20.5%	15.6%	102.6%	The Philippines	6795	9550	5995
India	18.5%	5.7%	83.3%	Mainland China	3266	1942	2976
Taiwan	18.2%	10.8%	102.0%	India	3132	1156	1131
United Kingdom	11.7%	13.9%	103.7%	Taiwan	2065	2411	1613
Iran	8.4%	4.5%	54.1%	Nigeria	1854	166	298
Mainland China	6.5%	5.3%	57.1%	Great Britain	1841	2521	714
The Philippines	4.6%	8.4%	96.4%	Canada	1587	2107	191
Canada	3.8%	9.5%	67.7%	Hong Kong	811	1350	885
South Korea	2.2%	5.0%	69.0%	Iran	804	1536	927
Pakistan	1.8%	1.4%	13.0%	Japan	787	1634	800
Israel	1.7%	1.6%	24.5%	South Korea	539	1656	5466
World Average	0.8%	0.8%	8.8%	Total	50,003	32,452	87,806

Notes: Immigration response regressions include country and year effects. Hong Kong is excluded from the regressions due to its special US immigration treatment. Patent regressions include country-year and industry-year effects; standard errors are clustered at the ethnicity level.

Table 1.10A: Quotas Levels Specification

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Log Foreign Output		
Log US Immigration Quotas Estimator	0.360	0.419	0.368
	(0.193)	(0.284)	(0.225)
	B. Log Foreign Labor Productivity		
Log US Immigration Quotas Estimator	-0.024	0.039	0.011
	(0.202)	(0.220)	(0.200)
	C. Log Foreign Employment		
Log US Immigration Quotas Estimator	0.384	0.380	0.357
	(0.120)	(0.159)	(0.125)
Industry x Year FE	X	X	X
Country x Industry FE	X	X	X
Observations	9912	9912	9912

Notes: Standard errors are clustered at the ethnicity level.

Table 1.10B: Quotas First-Differences Specification

	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
	(0.064)	(0.090)	(0.078)
	B. Δ Log Foreign Labor Productivity		
Δ Log US Immigration Quotas Estimator	0.054	0.135	0.086
	(0.072)	(0.074)	(0.076)
	C. Δ Log Foreign Employment		
Δ Log US Immigration Quotas Estimator	0.240	0.236	0.234
	(0.043)	(0.070)	(0.042)
Industry x Year FE	X	X	X
Observations	8736	8736	8736

Notes: Standard errors are clustered at the ethnicity level.

Table 1.11A: Quotas Country Controls - Levels

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Base Foreign Output Regression		
Log US Immigration Quotas Estimator	0.360	0.419	0.368
	(0.193)	(0.284)	(0.225)
Observations	9912	9912	9912
	B. Including Foreign Ph.D.s in US		
Log US Immigration Quotas Estimator	0.335	0.362	0.317
	(0.170)	(0.313)	(0.225)
Log Foreign Ph.D. Students in US	0.012	0.131	0.085
	(0.094)	(0.104)	(0.105)
Observations	8914	8914	8914
	C. Excluding Mainland China		
Log US Immigration Quotas Estimator	0.335	0.418	0.358
	(0.228)	(0.342)	(0.267)
Observations	9653	9653	9653
	D. Including Ethnic Time Trend		
Log US Immigration Quotas Estimator	0.244	0.172	0.139
	(0.133)	(0.094)	(0.116)
Observations	9912	9912	9912
Industry x Year FE	X	X	X
Country x Industry FE	X	X	X

Notes: Standard errors are clustered at the ethnicity level.

Table 1.11B: Quotas Country Controls - First-Differences

	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
	(0.064)	(0.090)	(0.078)
Observations	8736	8736	8736
	B. Including Foreign Ph.D.s in US		
Δ Log US Immigration Quotas Estimator	0.280	0.350	0.297
	(0.058)	(0.112)	(0.079)
Δ Log Foreign Ph.D. Students in US	0.033	0.054	0.054
	(0.069)	(0.090)	(0.078)
Observations	7780	7780	7780
	C. Excluding Mainland China		
Δ Log US Immigration Quotas Estimator	0.210	0.306	0.252
	(0.113)	(0.149)	(0.128)
Observations	8518	8518	8518
	D. Including Ethnic Time Trend		
Δ Log US Immigration Quotas Estimator	0.183	0.202	0.172
	(0.132)	(0.117)	(0.131)
Observations	8736	8736	8736
Industry x Year FE	X	X	X

Notes: Standard errors are clustered at the ethnicity level.

Table 1.12A: Quotas Falsifications - Levels

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Base Foreign Output Regression		
Log US Immigration	0.360	0.419	0.368
Quotas Estimator	(0.193)	(0.284)	(0.225)
	B. Including 1987 Counterfactual		
Log US Immigration	0.222	0.265	0.220
Quotas Estimator	(0.100)	(0.119)	(0.105)
1987 Counterfactual	0.206	0.231	0.221
Quotas Estimator	(0.213)	(0.276)	(0.243)
	C. Including 1995 Counterfactual		
Log US Immigration	0.347	0.514	0.409
Quotas Estimator	(0.261)	(0.399)	(0.311)
1995 Counterfactual	0.036	-0.260	-0.113
Quotas Estimator	(0.201)	(0.327)	(0.246)
Industry x Year FE	X	X	X
Country x Industry FE	X	X	X
Observations	9912	9912	9912

Notes: Standard errors are clustered at the ethnicity level.

Table 1.12B: Quotas Falsifications - First-Differences

	No Weights	Patent Weights	Output Weights
	(1)	(2)	(3)
	A. Base Foreign Output Regression		
Δ Log US Immigration	0.294	0.370	0.320
Quotas Estimator	(0.064)	(0.090)	(0.078)
	B. Including 1987 Counterfactual		
Δ Log US Immigration	0.262	0.335	0.287
Quotas Estimator	(0.061)	(0.071)	(0.069)
Δ 1987 Counterfactual	0.296	0.326	0.311
Quotas Estimator	(0.182)	(0.226)	(0.205)
	C. Including 1995 Counterfactual		
Δ Log US Immigration	0.266	0.439	0.345
Quotas Estimator	(0.079)	(0.102)	(0.076)
Δ 1995 Counterfactual	0.069	-0.170	-0.063
Quotas Estimator	(0.154)	(0.099)	(0.118)
Industry x Year FE	X	X	X
Observations	8736	8736	8736

Notes: Standard errors are clustered at the ethnicity level.

Chapter 2

Heterogeneous Technology Diffusion and Ricardian Trade Patterns

Summary 2 *This study tests the importance of Ricardian technology differences for international trade. Panel regressions find technology growth increases manufacturing exports. To establish a causal relationship between technology and trade, instrumental-variables specifications exploit uneven technology diffusion from the US through ethnic scientific networks. The instrumented elasticity of export growth to the exporter's technology development is 0.9 in the preferred specification. Supplemental specifications show this elasticity is robust to incorporating the importer's technology development and to controlling for the Rybczynski effect due to factor accumulation. An exogenous reform of US immigration law also confirms the results are not due to reverse causality. The findings suggest technology differences are an important determinant of trade patterns.*

2.1 Introduction

Trade among countries due to technology differences is a core principle in international economics. Countries with heterogeneous technologies focus on producing goods in which they have comparative advantages; subsequent exchanges afford higher standards of living than are possible in isolation. This Ricardian finding is the first lesson in most undergraduate courses on trade, and it still undergirds many modelling frameworks on which recent theoretical advances build (e.g., Dornbusch, Fischer, and Samuelson 1977, Eaton and Kortum 2002). In a famous response to Stanislaw Ulam's challenge to name a true and nontrivial theory in the social sciences, Paul Samuelson chose this principle of comparative advantage due technology differences.

While empirical tests of this framework date back to its original proponent David Ricardo (1817), the underlying technology differences across countries are very difficult to quantify. Proxies for technology (e.g., total factor productivity) allow substantial progress but do risk confounding heterogeneous technologies with other country-specific determinants of trade, especially in cross-sectional exercises. In principle, the relationship between technology and trade is best estimated through panel data models that remove time-invariant characteristics of each nation (e.g., distances, colonial history) and afford explicit controls of the time-varying determinants deemed important (e.g., factor accumulation, economic development, trading blocs). Of course, quantifying the dynamics of uneven technology advancement across countries is an even more challenging task, and whether the uncovered partial correlations represent causal parameters still needs to be addressed.[1]

This study develops this form of empirical environment by exploiting differences across countries in their access to the US technology frontier. Recent research emphasizes the importance of ethnic scientists and entrepreneurs living in the US for the diffusion of US technologies to their home countries. These frontier expatriates facilitate the transfer of the codified details of new innovations, but perhaps more importantly also convey the tacit knowledge required for successful adoption. The first chapter of this thesis (Kerr 2005a) finds that a larger ethnic research community in the US improves technology diffusion to foreign countries of the same ethnicity. That is, new computer technologies flow faster to Chinese economies than to Latin America if the Chinese computer research community in the US is stronger than the Hispanic community. Moreover, the foreign countries realize substantial manufacturing output and productivity gains from the stronger scientific integration. As invention is disproportionately concentrated in the US, these ethnic channels significantly influence the technology

[1] The panel exercises in this paper are closest in spirit to Harrigan (1997b), who evaluates the importance of both technology and factor supply differences across countries for determining industry specialization. This paper differs, however, in its direct study of trade flows, its substantial attention to non-OECD economies, and in its IV analysis using heterogeneous technology diffusion. Other tests of the Ricardian model are McDougall (1951, 1952) and Stern (1962).

opportunities of imitating economies.[2]

This uneven technology diffusion through ethnic networks offers an empirical foothold for evaluating the importance of technology differences across countries in explaining trade patterns. The empirical specifications, however, must be carefully designed to isolate technology's role, and the next section of this paper utilizes the multi-country Ricardian model of Eaton and Kortum (2002) to guide the form of the estimating equations. Eaton and Kortum construct a special theoretical framework that relates trade flows among countries to the technology capabilities, distances, and input costs of each economy. A simple application in Section 2.2 replaces Eaton and Kortum's random technology parameters with country-specific technology capabilities that depend upon a frontier country's technology state and the technology follower's human-capital stock with respect to the frontier innovations. The follower's human-capital stock is acquired through scientists of the following country's ethnicity who work in the frontier economy. Reduced-form expressions thereby relate the trade patterns of the technology follower to its ethnic scientific community in the technology leader.

After the appropriate estimating specifications are developed, Section 2.3 describes the dataset constructed for this project. Ethnic scientists working in the US are identified by applying an ethnic-name database to individual US patent records (e.g., identifies inventors with Chinese versus Hispanic names). The matched dataset describes the 1980-1997 ethnic composition of US inventors with unparalleled cross-sectional and longitudinal detail, and the sizes of ethnic research communities are determined at the industry level by aggregating individual patent records. These research communities are joined with detailed export data for foreign countries (e.g., US Chinese computer research is paired with China's trade in the computer industry) in an econometric framework that follows from theoretical model. Total factor productivity (TFP) indices are also developed as proxies for aggregate technology states.

The fourth section presents the main empirical results using bilateral manufacturing exports aggregated over industries from 1980-1997.[3] Following the reduced-form equations developed

[2] Kerr (2005a) provides additional references on the role of ethnic networks in transmitting new technologies. Other sources of heterogeneous technology frontiers are geographic distances to major R&D nations (e.g., Keller 2002b), the innovative efforts of trading partners (e.g., Grossman and Helpman 1991, Coe and Helpman 1995, Coe, Helpman and Hoffmaister 1997), or international patenting decisions (e.g., Eaton and Kortum 1999). Keller (2004) reviews the technology transfer literature.

[3] As discussed below, exports to the US are excluded from the trade patterns examined in this paper due to

in Section 2.2, tests of the Ricardian theory first regress these bilateral exports on the exporter's ethnic human-capital stocks for US technologies as measured in the ethnic patenting dataset. Panel fixed effects remove time-invariant determinants of bilateral trade and global developments in technology and trade; gravity covariates are also included to isolate technology's role. Export volumes rise with an elasticity of about 0.6 to better human capital for the US technology frontier, with the coefficient statistically different from zero. The results are robust to a number of sample decomposition exercises and specification variants. The strong elasticity of exports to the exporter's integration to the US frontier is also preserved when the importer's technology integration is added as an additional regressor. Moreover, the elasticity estimates for the importer's technology regressor are smaller and not statistically different from zero. The combined pattern suggests countries export more manufacturing goods when they develop a comparative advantage due to uneven technology diffusion from the US.

These initial reduced-form estimations assume the following order of events: technologies are developed in the US, ethnic scientists in the US transmit the technologies to their respective countries, and trade patterns are determined. Reverse causality, however, is an important concern, with a plausible alternative being that foreign human-capital development is responsible for both the export growth, perhaps with industry reallocations due to the Rybczynski effect, and the emigration of ethnic researchers to the US. If true, the ethnic human-capital stock for US technologies measured through the patenting data would not be a valid instrument for the exporter's technology set. To begin addressing this issue, Section 2.4 continues by developing a second estimator using exogenous, differential changes in the sizes of US ethnic research communities following the US Immigration Act of 1990. Reduced-form regressions with this immigration quotas estimator also find positive export growth following stronger scientific integration with the US, and the strong contrast with the importer's integration is again evident. While the new elasticity estimate of 0.5 is not directly comparable to the ethnic patenting estimator, the directions of the two exercises support each other.

After establishing these two estimators for heterogeneous technology diffusion, Section 2.4 concludes with the full OLS and IV regressions of bilateral export volumes on the exporter's and

potential network effects operating alongside technology transfer. Kerr (2005b) separately analyzes the role of ethnic scientific networks in US bilateral trade.

importer's technology states. Aggregate technology capabilities are proxied with country-level TFP indices. In the OLS regressions, growth in the exporter's technology set correlates with positive export growth, but the elasticity estimates are sensitive whether the gravity covariates are included; the importer's TFP coefficients are much weaker. IV regressions instrument for technology development in both countries using the heterogeneous diffusion from the US through ethnic scientific networks. First-stage regressions with both the patent-based and quotas-based instruments find a robust growth in foreign technology levels with a stronger US ethnic research community.

The second-stage regressions exhibit instrumented elasticities of exports to the exporter's technology development that range from 0.6 to 1.1 with the patent-based instruments; the higher estimates are for specifications that exclude the (potentially endogenous) gravity covariates. The range of elasticities evident with the quotas-based instruments is 1.1 to 1.6. In all cases these elasticity estimates are statistically different from zero. By contrast, the instrumented elasticities of exports to the importer's technology state are smaller and not consistently different from zero, especially when the gravity covariates are included. The conclusion from these country-level exercises is that Ricardian technology differences are important determinants of trade patterns.

Section 2.5 extends these core tests of the Ricardian model by exploiting the additional industry and geographic variation available in the combined trade and ethnic patenting dataset. These extensions are of interest in their own right, and more importantly provide additional confidence that the measured role for technology development is not reflecting an omitted factor accumulation. In contrast to the Ricardian framework, Heckscher-Ohlin-Vanek (HOV) models describe trade as resulting from factor differences across countries (e.g., labor, capital, natural resources).[4] During the period studied, some countries experienced significant growth in their skilled labor forces and physical capital stocks, as well as their technology sets, and the former could lead to significant growth in manufacturing exports due to the Rybczynski effect.

Section 2.5 tests this alternative hypothesis by contrasting industries within each country

[4]See Heckscher (1919), Ohlin (1933), and Vanek (1968). Dornbusch, Fischer, and Samuelson (1980) provide a classic HOV model, while Schott (2003) and Romalis (2004) offer state-of-the-art extensions and empirical tests. Trefler (1994, 1996), Harrigan (1997b) and Davis and Weinstein (2001) also jointly explore technology and factor differences as determinants of trade.

of similar factor input intensities. Technology's important role is preserved in these detailed matching exercises. Exploiting the within-country variation also ensures the findings are robust to other country-level explanations like nations entering trade agreements or multinational bodies (e.g., the World Trade Organization), asynchronous business cycles, and so on. Section 2.5 concludes by analyzing the geographic margin of trade expansion with technology improvements. Export growth is strongest in bordering and nearby countries, but positive growth is evident at all distances.

The results of this project confirm technology is an important determinant of trade; moreover, it is relevant for explaining changes in trade patterns over time. Section 2.6 concludes this paper by discussing future projects that will utilize the uneven transmission of technologies through ethnic networks to characterize how Ricardian technology differences shape international exchanges.

2.2 Theoretical and Estimating Frameworks

This section develops the estimating equations employed in Section 2.4's empirical analysis. It begins by briefly sketching a recent multi-country Ricardian model of Eaton and Kortum (2002). This framework is unique in relating trade to technology differences across several countries, and a simple application builds into this theory ethnic research networks and heterogeneous technology diffusion. From this analysis, reduced-form specifications are developed that relate bilateral exports to ethnic human-capital stocks with respect to frontier technologies. The second half of this section in turn manipulates these theoretical specifications into a framework suitable for empirical analysis.[5]

2.2.1 Theoretical Framework

The world consists of N countries producing and consuming a continuum of goods $j \in [0, 1]$. Consumers maximize utility in each period by purchasing these goods in quantities $Q(j)$ ac-

[5]See also Alvarez and Lucas (2004). The Dornbusch, Fischer, and Samuelson (1977) approach to Ricardian trade is not readily extended to multiple countries, although some local comparative statics are feasible (e.g., Wilson 1980).

cording to a constant elasticity of substitution (CES) objective function,

$$U = \left(\int_0^1 Q(j)^{(\sigma-1)/\sigma} dj \right)^{\sigma/(\sigma-1)}, \tag{2.1}$$

subject to prices determined below. $\sigma > 0$ is the elasticity of substitution across goods for the consumers. Consumers earn wage w and consume their full wages in each period. Accordingly, time subscripts are omitted throughout most of this discussion.

Countries are free to produce or trade all goods. Inputs can move among industries within a country but not across countries. Industries are characterized by identical Cobb-Douglas production functions employing labor with elasticity α and the continuum of produced goods, also aggregated with (2.1), with elasticity $1 - \alpha$. Factor mobility and identical production functions yield constant input production costs across goods within each country, $c_i(j) = c_i \ \forall j$.

Technology differences exist across countries, so that country i's efficiency in producing good j is $z_i(j)$. With constant returns to scale in production, the unit cost of producing good j in country i is $c_i/z_i(j)$. While countries are free to trade, geographic distance results in "iceberg" transportation costs so that delivering one unit from country i to country n costs $d_{ni} > 1$ units in i. Thus, the delivery to country n of good j made in country i costs

$$p_{ni}(j) = \left(\frac{c_i}{z_i(j)} \right) d_{ni}. \tag{2.2}$$

An increase in country i's efficiency for good j lowers the price it must charge. Perfect competition allows consumers to buy from producers in the country offering the lowest price (inclusive of shipment costs). Thus, the price that consumers in country n pay for good j is

$$p_n(j) = \min[p_{ni}(j); i = 1, ..., N]. \tag{2.3}$$

The technology determining the efficiency $z_i(j)$ is modelled as the realization of a random variable Z_i drawn from a country-specific probability distribution $F_i(z) = \Pr[Z_i < z]$. Draws are independent for each industry j within a country. The core innovation of Eaton and

Kortum's model is to use Fréchet functional distribution to model technologies,

$$F_i(z) = e^{-T_i \cdot z^{-\theta}}, \tag{2.4}$$

where $T_i > 0$ and $\theta > 1$. The country-specific parameter T_i determines the location of the distribution, while the common parameter θ determines the variation within each country's distribution. By the law of large numbers, a larger T_i raises the average efficiency of industries for country i, and therefore its absolute advantage for trade. A larger θ, on the other hand, implies a tighter distribution for industries within every country and thereby limits the scope for comparative advantage across nations.

To model heterogeneous technology diffusion through ethnic ties to the frontier economy, the technology location parameter is specified as

$$T_i = \tilde{T} \cdot (H_i)^{\tilde{\beta}_H}. \tag{2.5}$$

$\tilde{T}$ is the exogenously determined frontier technology stock.[6] H_i is the human-capital stock of country i with respect to the frontier innovations, including both the codified and tacit knowledge required for successful adoption. This human-capital stock depreciates at a rate δ, and the population of researchers of country i's ethnicity ($\tilde{L}_i$) working in the frontier country replenishes it: $\partial H_i / \partial t = -\delta H_i + \tilde{L}_i$. If the number of expatriate researchers is constant, the steady-state human-capital stock of country i with respect to frontier inventions is $\delta^{-1}\tilde{L}_i$. The elasticity $\tilde{\beta}_H$ is empirically estimated below.

2.2.2 Estimating Framework

The Fréchet distribution (2.4) allows prices from equations (2.2) and (2.3) to be determined. The probability that country i is the lowest-cost producer of an arbitrary good for country n is $\pi_{ni} = T_i(c_i d_{ni})^{-\theta} / \sum_{k=1}^{N} T_k(c_k d_{nk})^{-\theta}$.[7] With a continuum of goods, π_{ni} is also the fraction of

[6] Variables referring to the frontier economy are generally denoted by a tilda.

[7] The distribution of prices country i presents to country n is $G_{ni}(p) = \Pr[P_{ni} \leq p] = 1 - F_i(c_i d_{ni}/p) = 1 - \exp(-T_i(c_i d_{ni})^{-\theta} p^{\theta})$. Country n buys from the lowest cost producer of each good, so that its realized price distribution is $G_n(p) = \Pr[P_n \leq p] = 1 - \prod_{i=1}^{N}[1 - G_{ni}(p)] = 1 - \exp(-p^{\theta} \sum_{i=1}^{N} T_i(c_i d_{ni})^{-\theta})$. The probability is $\pi_{ni} = \Pr[P_{ni}(j) \leq \min\{P_{ns}(j); s \neq i\}] = \int_0^{\infty} \prod_{s \neq i}[1 - G_{ns}(p)] dG_{ni}(p)$. See Eaton and Kortum (2002) for the

goods country n purchases from country i. Country n's average expenditure per good does not vary by source country, so that the fraction of country n's expenditure on goods from country i is also

$$\frac{X_{ni}}{X_n} = \frac{T_i(c_i d_{ni})^{-\theta}}{\sum_{k=1}^{N} T_k(c_k d_{nk})^{-\theta}}, \tag{2.6}$$

where X_n is total expenditure in country n. Holding input prices constant, technology growth in country i increases its exports to country n through entry into industries in which it was previously uncompetitive. Looking across import destinations for an industry in which it already exports, country i also becomes the lowest-cost producer for more distant countries it could not previously serve due to the markup of transportation costs. Condition (2.6) also shows how trading costs d lead to deviations in the law of one price.

For the case of frictionless trade or constant trading costs ($d_{nk} = d \; \forall n, k$)[8], condition (2.6) can be rearranged into the structural estimating equation for year t,

$$\ln(X_{nit}) = \ln(T_{it}) - \theta \ln(c_{it}) + \ln(X_{nt}) - \ln\left(\sum_{k=1}^{N} T_{kt}(c_{kt})^{-\theta}\right).$$

Panel estimations of bilateral exports are used to evaluate this structural relationship. A vector of year effects η_t control for the world price and technology aggregate $\ln\left(\sum_{k=1}^{N} T_{kt}(c_{kt})^{-\theta}\right)$; the year effects also remove uniform growth in trade volumes and price changes during the period studied. A vector of cross-sectional effects ϕ_{ni} further extract time-invariant determinants of bilateral exports from country n and country i (e.g., distances, colonial ties). Aggregate expenditure in the importing country (X_{nt}) is proxied by the importer's GDP level in year t.

To integrate the estimations with the empirical trade literature, most specifications further add the gravity covariates of the exporter's GDP and both countries' GDP per capita. The exporter's GDP per capita also captures broad changes in the input costs c_{it}.[9] These gravity covariates are interacted following Frankel (1997) and Rauch (2002). The primary estimating

full derivation of the price index.

[8]The assumption of constant trading costs is unimportant for the upcoming panel estimations if the number of countries is large. The numerator's bilateral distance $d_{ni}^{-\theta}$ is time-invariant and absorbed into the cross-sectional effects. The error from modelling the denominator's world price and technology aggregate with year effects is small, $\lim_{N\to\infty} [\partial \ln(\sum_{k=1}^{N} T_{kt}(c_{kt}d_{nk})^{-\theta})/\partial \ln(T_i)] = 0$.

[9]Input costs can be endogenized through a specification of the labor market. This study does not undertake this step, instead measuring export growth due technology transfer net of input costs increases from general-equilibrium wage pressure.

equation is

$$\ln(X_{nit}) = \alpha + \beta_T \ln(T_{it}) + \gamma \ln(GDP/CAP_{it} \cdot GDP/CAP_{nt}) + \zeta \ln(GDP_{it} \cdot GDP_{nt}) + \phi_{ni} + \eta_t + \epsilon_{nit}, \tag{2.7}$$

where the theoretical elasticity of one between technology and bilateral exports is empirically evaluated with the β_T coefficient.[10]

The OLS specification (2.7) is evaluated in Section 2.4 with the aggregate technology parameter T_{it} measured through country-level TFP indices. While this Ricardian framework clearly assigns a causal relationship of export growth to technology development, in practice the empirical estimation of (2.7) can be confounded by reverse causality or omitted variable biases. Moreover, the possible simultaneous accumulation of factor endowments and new technologies is particularly worrisome for isolating the Ricardian impetus for trade from relative factor scarcities. Technology is the only channel promoting export growth in this framework due to identical factor endowments and no intertemporal factor accumulation.[11]

Anticipating these issues, this section closes with how heterogeneous technology transfer from the frontier economy (2.5) provides a foothold for establishing causality when these complications are introduced. The human-capital stock of country i with respect to the frontier innovations (H_i) affects country i's exports only through technology transfer and can thus serve as an instrument for country i's technology in the structural specification (2.7). This human-capital stock is acquired through researchers of country i's ethnicity ($\tilde{L}_i$) working in the frontier economy, and two instruments are developed in Section 2.4 using ethnic scientific communities in the frontier US economy. Substituting (2.5) into the structural equation (2.7) yields the

[10]The interaction of the gravity covariates is further discussed below. The gravity relationship in the Eaton and Kortum model arises due to technology differences interacting with distances and production costs. An alternative derivation through Armington or monopolistic competition builds on imperfect substitution among goods for consumers. The former predicts trade growth at the extensive margin with technology development, while the latter two predict trade growth at the intensive margin.

[11]Differences in preferences or non-homothetic utility functions can also promote trade, but Hunter and Markusen (1988) and Hunter (1991) cap these stimulants at 20% of world trade. The specified production function also abstracts from trade due to increasing returns to scale (e.g., Helpman and Krugman 1985, Antweiler and Trefler 2002).

reduced-form contribution of these communities for exports from their home countries,

$$\begin{aligned} \ln(X_{nit}) &= \alpha + \beta_H \ln(H_{it}) + \gamma \ln(GDP/CAP_{it} \cdot GDP/CAP_{nt}) \\ &\quad + \zeta \ln(GDP_{it} \cdot GDP_{nt}) + \phi_{ni} + \eta_t + \epsilon_{nit}, \end{aligned} \tag{2.8}$$

where the log frontier technology state $\tilde{T}_t$ is separated from the ethnic human-capital stock H_{it} and absorbed into the year effects η_t and $\beta_H = \beta_T \cdot \tilde{\beta}_H$.[12] The empirical exercises below commence with this reduced-form relationship, and then analyze (2.7) in a two-stage least squares framework.

2.3 Dataset Preparation

To test empirically these Ricardian predictions, a dataset is prepared combining bilateral trade data, foreign-country technology measures, and US ethnic human-capital stocks. The core industry-level and country-level variation exploited in this study is dictated by the ethnic patenting metrics developed first. Once this panel foundation is established, the mapping in of trade and foreign-country TFP measures is easily motivated.

2.3.1 US Ethnic Human-Capital Stocks

This paper exploits technology differences across countries arising due to uneven technology diffusion from the US through ethnic scientific networks. The ethnic human-capital stocks with respect to US technologies are developed through the NBER Patent Data File (Hall, Jaffe, and Tratjenberg 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 inventor ethnicities, a commercial database of ethnic first names and surnames is mapped into the inventor records. The match rate is 99% for US patent records, and the process affords the

[12] Time-invariant differences in access to the frontier technology state, $T_{it} = \tilde{T}_t \cdot \Upsilon_i \cdot (H_{it})^{\tilde{\beta}_H}$, are absorbed into the cross-sectional effects ϕ_{ni} in the log specification (2.8). These differences could arise due to geographic distance from the US (e.g., Keller 2002), heterogeneous production techniques (e.g., Davis and Weinstein 2001), and so on.

distinction of nine ethnicities: Chinese, English, European, Hispanic, Indian, Japanese, Korean, Russian, and Vietnamese.

Table 2.1 describes the 1980-1997 US sample. The trends demonstrate a growing immigrant contribution to US technology development, especially among Chinese and Indian scientists. Also matching popular perceptions, 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.[13] Figure 2.1 illustrates the evolving immigrant contribution to US technology development as a percentage of patents granted by the USPTO, while Figure 2.2 provides a more detailed glimpse of immigrant contributions by broad technology groups.[14]

From this matched database, the ethnic human-capital stocks H_i to the US frontier are easily developed. Recall that these stocks depend upon the number of frontier researchers undertaking inventive activity of country i's ethnicity ($\tilde{L}_i$): $\partial H_i/\partial t = -\delta H_i + \tilde{L}_i$. Define the patenting productivity of a US researcher to be $\tilde{P}$, so that the measured patenting of ethnicity i in year t is expected to be $\tilde{P}_{it}^{Flow} = \tilde{P} \cdot \tilde{L}_{it}$. Inverting this expression suggests the size of the ethnic research community can be inferred annually using the observed number of ethnic patent applications divided by the constant researcher productivity ($\tilde{L}_{it} = \tilde{P}_{it}^{Flow} / \tilde{P}$).

With the population of US ethnic researchers quantified annually, subject to a multiplicative constant, the human-capital stocks H_i could be estimated through the perpetual inventory method. The empirical results presented below, however, take a slightly different approach. In an examination of international patent citations, Kerr (2005a) finds ethnic scientific networks aid direct communications among inventors. Inventors living outside of the US cite US inventors of their own-ethnicity approximately 50% more often than other US-based inventors, even after controlling for technology classes. Moreover, by considering different time lags between the

[13] The estimated European ethnic contribution is naturally higher than the immigrant contribution measured by foreign born.

[14] The third chapter (Kerr 2005c) further details the ethnic patenting dataset and provides additional descriptive statistics. A quality assurance exercise matching the ethnic-name database to foreign patent records registered in the US is also presented. The ethnic-name procedure assigns ethnicities to 98% of foreign inventor records, and the average own-ethnicity contribution is 88%. Similar to the US, own-ethnicity contributions should be less than 100% due to expatriate researchers.

filing dates of the cited and citing patents, the ethnic bias is shown to be most important in the first five years of the diffusion process. The human-capital stocks H_i are accordingly modelled by aggregating the number of ethnic patents over the previous five years, with the panel fixed effects controlling for any changes in patenting productivity $\tilde{P}$ of US researchers.[15,16]

These ethnic human-capital stocks for US innovations are developed at the four-digit level of the International Standard Industrial Classification (ISIC) system. This framework distinguishes 81 manufacturing industries at a level of detail that straddles the two-digit and three-digit levels of the US Standard Industrial Classification system. Table 2.A1 in the appendix lists the ISIC industries employed.[17]

2.3.2 Export Volumes

Bilateral exports are taken from the World Trade Flows Database (WTF), compiled by Statistics Canada and Feenstra (2000). This rich data source documents the product-level values of bilateral trade for most countries from 1980-1997. These product flows are aggregated into the four-digit ISIC industries developed in the US patent dataset, and exporting countries are grouped into the eight non-English ethnicities that are identifiable with the ethnic-name database. Five ethnicities map to a single country, while the Chinese, European, and Hispanic ethnicities have larger blocs. Table 2.2 lists the countries studied and their summary characteristics.[18]

[15] In an influential survey, Griliches (1990) notes that annual fluctuations in US patents granted occur due to changes in USPTO personnel resources, as well as technology development. Over the last two decades, US patent grants have increased dramatically. While several explanations for this increase are put forth (e.g., Kortum and Lerner 2000, Kim and Marshcke 2004, and Hall 2004), it is clear that the number of patents awarded has grown faster than the growth in research scientists would suggest. The time effects account for changes in the underlying patenting productivity, effectively contrasting ethnic shares of US patents granted.

[16] Similar results are found if the perpetual inventory method is employed. The two disadvantages of this technique are the assignments of a depreciation rate δ for human capital and initial human-capital stocks. The latter is particularly burdensome since the ethnicity of inventors can only be determined after 1975.

[17] The USPTO issues patents by technology categories rather than by industries. Combining the work of Johnson (1999) and Silverman (1999), concordances are developed between the USPTO classifications and the four-digit ISIC industries in which new inventions are manufactured or used. The main estimations focus on industry-of-use, affording a composite view of the technological opportunity developed for an industry. Studies of advanced economies find accounting for these inter-industry R&D flows important (e.g., Scherer 1984, Keller 2002a). Estimations with manufacturing industries support the using-industry specifications.

[18] The empirical analysis verifies the multiple country mappings do not unduly influence the results. Some country to ethnicity mappings are debatable (e.g., placing Spain and Portugal with European rather than Hispanic, including the Scandinavian countries in European), as is the inclusion of communist countries. The results are robust to these marginal reclassifications.

The first pair of columns present the mean 1980-1997 multilateral export volumes and growth rates for each country (in nominal US dollars). Exports to all countries other than the US are considered in this paper; trade relations with the US are excluded, however, due to the strengthening network effects — which are also thought to increase trade flows — operating alongside the heterogeneous technology transfer.[19] The base panel is thus asymmetric in the sense that exporters are limited to countries of non-English ethnicities contained in the ethnic-name database, but in many specifications their bilateral exports to other countries are included (e.g., exports to English or African countries). Variations on the base panel below restrict the sample to be only bilateral flows among the eight non-English ethnicities, particularly when the importer's technology growth is being contrasted with the exporter's technology development.

The forty-four economies account for 53% and 64% of global manufacturing exports in 1980 and 1997, respectively, with countries of English ethnicity accounting for most of the residual (the US export share is 12% in 1980 and 13% in 1997). Not surprisingly, the largest exporters are European nations (especially Germany) and Japan, while the smallest exporters are found in Latin America. Vietnam (25%), Hong Kong (15%), Korea (13%), and Mainland China (13%) experience the strongest compound annual growth in nominal exports; only Venezuela demonstrates an absolute decline in trade volumes over the seventeen years.

2.3.3 TFP Indices and Development Indicators

This study employs country-level total factor productivity (TFP) as a proxy for the technology location parameter T_i. Section 2.2's theory abstracts from capital stocks by specifying a Cobb-Douglas production function of labor (with elasticity α) and intermediate inputs from the CES aggregator. Capital accumulation, however, is important for explaining economic development, particularly the rapid advances made by several East Asian economies (e.g., Young 1992, 1995; Ventura 1997). Accordingly, TFP indices are developed for each country c relative to the US frontier with a production function employing capital and labor,

$$TFP_{c,US} = \frac{Y_c}{Y_{US}} \left(\frac{L_{US}}{L_c}\right)^{\alpha} \left(\frac{K_{US}}{K_c}\right)^{1-\alpha}, \tag{2.9}$$

[19] Kerr (2005b) explores how ethnic scientific networks facilitate bilateral trade with the US. See also Rauch (2001) and Rauch and Trindade (2002).

with the year subscripts omitted. This index is superlative (i.e., it is exact for the Cobb-Douglas production function).[20] As this index is also transitive (i.e., $TFP_{ac} = TFP_{ab} \cdot TFP_{bc}$), the choice of the base country is irrelevant, with the US baseline providing intuition only. The second pair of columns in Table 2.2 presents the country-level TFP indices calculated from the Penn World Tables (PWT) using data on aggregate GDP, workers, and capital stocks.[21]

The final pair of columns exhibit GDP per capitas for each country that are also used in the estimations below. Comparing the TFP indices to the GDP per capitas highlights several important points about the former. First, the mean country TFPs relative to the US range from <1% for Honduras, Nicaragua, and Bolivia to 34% for Germany and 46% for Japan. In the GDP per capita series, the four lowest 1980-1997 means are India (\$1364), Mainland China (\$1653), Honduras (\$1747), and Nicaragua (\$2007), while the four highest are Switzerland (\$19,258), Norway (\$17,829), Denmark (\$17,487), and Japan (\$17,120). The levels and rank correlations are 55% and 68%, respectively. The core difference between the two series is that the relative TFP metrics of larger to smaller countries (e.g., Mainland China, India, and Brazil vis-à-vis Scandinavian nations) tends to be greater than the relative GDP per capita metrics. This difference may reflect how workers are measured in the underlying PWT data or may be due to an increasing discrepancy in using the US labor share. The levels of both series are inconsequential, however, as the estimations employ cross-sectional fixed effects.

The correlation in growth rates is much higher at 94%, and their rankings are closely tied. The three fastest growing economies in both series are Mainland China, Taiwan, and Korea; Vietnam's GDP per capita also experiences a fast 10% growth rate, but TFP measures are unavailable due to insufficient capital data. Likewise, TFP metrics cannot be constructed for Belize, Cuba, and the former Soviet Union. While these four economies are excluded from the IV regressions, they are retained in the reduced-form exercises that form the bulk of this study.

Estimating country-level TFP indices is fraught with peril: the industrial structures of

[20] Caves, Cristensen, and Diewert (1982) derive the general TFP index for the translog family. The Cobb-Douglas index (2.9) is a special case where labor shares are constant.

[21] See Heston, Summers, and Aten (2002). Real GDP levels are adjusted for PPP differences and for price movements using the Laspeyres method (rgdpl, pop). Real capital stocks are measured beginning-of-year and calculated from aggregate investment (ki) using the perpetual inventory method. Initial stocks are developed for 1970, and a depreciation rate of 15% is employed. Worker estimates are developed by the PWT from International Labour Organization data, with linear interpolation between census or survey dates for each country (rgdpwok). The share of labor α is taken to be 0.67.

countries differ, inputs are only coarsely measured, pricing differentials and exchange rates are difficult to handle, and so on (e.g., Harrigan 1997a, 1997b). Measurement error is clearly present in the TFP metrics constructed. However, errors confined to cross-sectional levels or aggregate yearly fluctuations (e.g., due to mismeasurement in the US baseline) do not affect the analysis due to the log specification and panel fixed effects. Idiosyncratic errors across countries with time, however, are not captured by the panel effects and bias the estimated elasticities towards zero. This scope for this concern, however, is also limited in that the TFP indices are primarily used in the IV regressions, which circumvent measurement error in the endogenous regressor.

2.4 Country-Level Estimations

This combined dataset is a unique laboratory for evaluating Ricardian technology differences in international trade. This section evaluates country-level regressions of bilateral export volumes using the derived specification (2.7), while the next section concentrates more on the industry and geographic dimensions of the data. These later extensions provide a better platform for distinguishing heterogeneous technology diffusion from shifts in factor endowments than the country-level regressions. Accordingly, most discussion of the factor content of trade is postponed until then.

The study deviates from the typical arrangement of empirical papers by presenting the reduced-form specification (2.8) first, with the OLS and IV regressions using the exporting country's TFP index held until the section's end. This ordering takes early advantage of the richness of the ethnic patenting dataset; it further segues into the introduction of a second estimator for US ethnic human-capital stocks that exploits exogenous changes in US immigration law. Both of these estimators are used in the IV analysis. Throughout this section, technology development is found to be a strong determinant of exports.

2.4.1 US Ethnic Human-Capital Estimations

Table 2.3 evaluates the reduced-form specification (2.8). The dependent variable for each regression is the log dollar value of bilateral exports from country n to country i.[22] The base regression includes the log of the exporting country's human-capital stock H_i with respect to US technologies, as well as the interacted gravity covariates of the exporter's and importer's log GDP per capitas and log GDPs. Although not reported in the table, vectors of cross-sectional and year fixed effects are included as well. Finally, to account for the multiple countries mapping into the Chinese, European, and Hispanic ethnicities, standard errors are clustered at the ethnicity level. This conservative cross-sectional clustering further addresses the serial-correlation concerns of Bertrand, Duflo, and Mullainathan (2004).

As both variables are in logs, the 0.569 coefficient in the upper-left corner finds a 0.6% increase in the value of bilateral exports with a 1% increase in ethnic human-capital stocks with respect to US technologies. Stronger scientific integration with the US frontier clearly correlates with an increase in exports to other countries. The gravity covariates are important for isolating technology's role, as the growth in human-capital stocks could otherwise be confounded with other sources of growth in country size or standards of living that independently promote trade. The economic development of the exporter, however, is clearly endogenous — technology transfer in Section 2.2's model simultaneously increases the exporter's GDP and trade. The second column finds the estimated elasticity increases to 1.3 if the gravity covariates are excluded.

Figures 2.3 and 2.4 provide a graphical view of these two regressions if export volumes are summed across importers into single exporter-year observations,

$$\ln(X_{it}) = \alpha + \beta_H \ln(H_{it}) + \gamma \ln(GDP/CAP_{it}) + \zeta \ln(GDP_{it}) + \phi_i + \eta_t + \epsilon_{it}.$$

The data points in Figure 2.3 are the residuals of total non-US multilateral exports and US ethnic human-capital stocks after the country and year fixed effects and the exporter's gravity

[22]Nominal dollar values are employed, with the vector of year effects extracting aggregate price changes. This approach is motivated by the expenditure specification derived in Section 2.2. Section 2.5 demonstrates, however, that the results are robust to individually deflating each industry's trade with the NBER shipments deflators (Bartlesman and Gray 1996) before aggregating to country-wide exports.

covariates are removed. The slope of the trend line through Figure 2.3's residuals is akin to the β_H coefficient estimated in the first column. Figure 2.4, on the other hand, follows the second column and does not partial out the gravity covariates. Without the gravity covariates, the strong relationship is driven by the exceptional performance of Chinese economies, Korea, and India. Removing development levels, on the other hand, leads to a more nuanced picture with a broader set of nations contributing to the positive relationship. The sample composition exercises below directly evaluate the importance of different ethnic groups to the measured elasticity.[23]

The constructed dataset builds on two rich sources for economic data, and thus trade and ethnic patenting data are available for all observations (subject to the ethnicities identifiable with the ethnic-name database). The base specification, however, excludes observations if the gravity covariates are missing or if the export volume is zero. Columns 3 and 4 evaluate whether these restrictions are overly influencing the measured elasticity. First, gravity covariates are not always available for very small importers, resulting in a 22% loss from the maximum possible sample size of non-zero bilateral exports. Column 3 demonstrates, however, that this attrition is unimportant as the elasticity estimate in the larger sample is very close to Column 2.

Second, approximately 25% of the possible bilateral exchanges have zero values. In some cases these zero values reflect explicit restrictions on trade, but most are due to a combination of geographic distance and small country sizes. Zero values are undefined in a log specification like (2.8), and the base regression drops these observations. To ensure this procedure is not dictating the results, Column 4 retains the zero-valued observations through recoding and including appropriate indicator variables. The results are again very close to the base regression in Column 1, indicating that ignoring zero-valued observations is not overly influencing the estimated elasticities. Moreover, Figures 2.3 and 2.4 also find a positive relationship at the exporter-year level, where no zero values are encountered.

As highlighted in the data description section, the base sample is asymmetric in the sense

[23]Figures 2.3 and 2.4 exclude Belize, Cuba, the former Soviet Union, and Vietnam, all of which are included in Table 2.3's reduced-form analysis. This done for consistency with TFP graphs examined later, when these countries are dropped due to missing TFP indices. The exports of the former Soviet Union and Vietnam also demonstrate strong correlations to their US ethnic human-capital stocks. For visual ease, the figures also exclude three other trade outliers (Panama 1988, Paraguay 1991, Venezuela 1991); the inclusion or exclusion of these observations does not have a noticeable effect on the results.

that it only considers bilateral exports from the forty-four countries with ethnicities identifiable with the ethnic-name database; the countries receiving the imports, however, can be of any ethnicity (including English). Column 5 restricts the base sample to consider only bilateral exports within the countries associated with the ethnic patenting dataset. There is no significant difference from Column 1.

More importantly, this balanced panel serves as a platform for further characterizing the direction of trade due to technology differences. Column 6 introduces the similarly constructed importer's human-capital stock with respect to the US frontier. The strong elasticity of manufacturing exports to the exporter's scientific integration is preserved. On the other hand, the elasticity to the importer's integration is smaller and not statistically different from zero. A related regression of bilateral imports on the two ethnic human-capital stocks equivalently finds that imports respond positively to the foreign country's technology development and less to own-country ethnic human-capital stocks for US technologies.

Finally, Columns 7 and 8 present two reduced-form regressions that correspond directly to the IV regressions (with gravity covariates) later in this section. Belize, Cuba, the former Soviet Union, and Vietnam are excluded from the IV regressions since TFP indices are not available. While this attrition is only 10% of the countries in the sample, the number of non-English ethnicities is reduced from eight to six. Comparing Column 7 to Column 1, the elasticity of bilateral exports to the exporter's US ethnic human-capital stocks is somewhat weaker but still statistically significant in the smaller sample. Column 8, however, finds a substantial reduction in the measured elasticity of exports to the importer's human-capital stocks for US technologies, although the large standard errors in both cases indicate the estimate is imprecisely measured. This contrast is further discussed in the IV section.[24]

2.4.2 Sample Composition

The assembled dataset is a diverse set of countries and experiences. While Figures 2.3 and 2.4 suggest the unique outcomes of a single exporter or ethnicity (e.g., Mainland China, India) are not solely responsible for the positive correlations, Columns 9-16 formally test the robustness

[24] The standard errors for the importer's technology regressor in Columns 6 and 8 are calculated through dual regressions clustering on the importer's ethnicity.

of the measured elasticity to sequentially dropping each ethnicity. The β_H coefficient remains strong and statistically significant throughout; the regressions are individually discussed below to highlight further the underlying features of the dataset.

First, Column 9 drops the four Chinese economies. Perhaps surprisingly, the measured elasticity strengthens when these economies are excluded (although the coefficient is not statistically different from the base regression). The US Chinese research community exhibits the strongest growth over the 1980-1997 period, and therefore receives significant attention in the fixed effect estimations. All four economies exhibit strong export growth during the period studied and contribute to a high elasticity in the regressions without the gravity covariates (evident in Figure 2.4). After the gravity covariates are included, however, Taiwan's export growth is weaker than the expanding US Chinese human-capital stock would have predicted (evident in Figure 2.3). Dropping this ethnicity thus raises the estimated elasticity from the base regression. Individually excluding each Chinese economy also results in 10%-20% positive or negative shifts, but the results are remarkably robust to dropping this special case.

The constructed 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 patents were awarded to Japanese inventors living outside of the US during this period.[25] Positive correlations of export growth to US ethnic research may simply be capturing reverse technology flows, intra-company patenting, or defensive patenting from these advanced economies. Columns 10 and 13 demonstrate, however, that excluding the large European bloc or Japan only results in small increases in the estimated elasticity. Hispanic countries also account for about 45% of the sample, and Column 11 demonstrates that the results are robust to dropping these nations too. Individual European and Hispanic countries can similarly be excluded.[26]

The last five columns turn to the ethnicities with single country mappings. India and Korea, like the Chinese ethnicity, are widely noted for their export growth and US research

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

[26] An alternative strategy to mapping multiple countries into the Chinese, European, and Hispanic ethnicities is to run the export regressions at the ethnicity level (thereby also removing intra-ethnicity trade). As these sample composition exercise suggest, this approach yields similar results to the country-level analysis.

presence during the 1980-1997 period. Excluding these two countries in Columns 12 and 14 leads to only small declines in the estimated elasticity. Finally, the seventeen years covered by this study witnessed the integration of the former Soviet Union and Vietnam into the world trading community. During this period, the US research presence of these two ethnicities also increased, especially Vietnamese which roughly quadrupled its ethnic share of US patent applications (from a low initial position of 0.1%). Columns 15 and 16 show these groups contribute to the estimated elasticity, but that the measured effect is not being driven by their integration alone.

Finally, unreported regressions divide the sample into two smaller, overlapping time periods: 1980-1992 and 1985-1997. The positive dependence of exports on technological advancement is clearly present in both subsamples, but the relationship is stronger in the later years (elasticities of 0.4 and 1.0, respectively). This higher elasticity stems from the stronger growth of Asian research communities in US high-tech industries and their associated export growth in the later years. Section 2.5's industry-level regressions further evaluate the different elasticities in high-tech versus low-tech industries. In summary, the sample composition adjustments find positive export growth from scientific integration with the US is evident throughout the panel studied.

2.4.3 US Immigration Reform Estimations

The reduced-form estimations demonstrate a strong correlation between the growth of US ethnic scientific communities and their home country's export development. In preparation for the IV specifications, it is important to question whether these correlations capture causal relationships. Of particular concern is reverse causality. A plausible alternative to technology transfer from the US is that foreign human-capital development leads to both export growth abroad, perhaps with industry shifts due to the Rybczynski effect, and the emigration of scientists to the US. If true, using US ethnic research communities measured through the patenting database as an instrument for foreign technology levels would not solve the endogeneity problem.[27]

US immigration law is a foothold for establishing greater confidence in the direction of

[27] Another concern may be omitted variables that simultaneously boost the productivity of US ethnic researchers and the exports of their home countries (e.g., Japanese technology diffusion through Asian business networks). The ability to contrast exporter and importer technology levels substantially weakens this concern, and most of the ensuing discussion centers on reverse causality.

causation as it only influences foreign exports (to countries other than the US) through the sizes of US ethnic research communities and their associated technology transfer. If the populations of immigrant scientists and engineers (ISEs) are exogenously determined by legal restrictions, a second reduced-form strategy for ethnic human-capital stocks with respect to the US frontier can be developed within the US quotas system. US immigration law does not control the population sizes of foreigners in the US, but it does control the inflow of new immigrants. Define the quota on ISE inflows from country c to the US to be $QUOTA_{ct}$. Assuming that only the previous three years of immigration matter for a research stock[28], a reduced-form immigration estimator for ethnic scientific integration to the US is modelled as

$$\ln(IMM_{ct}^{RF}) = \ln\left[\sum_{s=1}^{5}(QUOTA_{c,t-s} + QUOTA_{c,t-s-1} + QUOTA_{c,t-s-2})\right]. \tag{2.10}$$

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 empirical version of (2.10) using exogenous changes in US immigration quotas from the Immigration Act of 1990 (1990 Act).

The disproportionate influence of 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.[29] Yet, the US immigration system significantly restricted the inflow of ISEs from certain nations prior to its reform with the 1990 Act.

US immigration law applies two distinct quotas to numerically restricted immigrants.[30] Both of these quotas were increased by the 1990 Act, and their combined change dramatically

[28]The immigration reform examined below focuses on a very sharp surge in immigration that makes this assumption reasonable.

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

[30]US immigrants are admitted through numerically restricted categories, governed by the quotas discussed in this section, and numerically unrestricted categories (e.g., immediate relatives of US citizens). The reduced-form estimator centers on the numerically restricted categories that admit 75% of ISEs (versus 43% of all immigrants). Jasso, Rosenzweig, and Smith (1998) outline US immigration policy and the 1990 Act; they further discuss behavioral responses to changes in quotas. ISE inflows through the unrestricted categories are stable in the years surrounding the 1990 reform.

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.[31] 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.[32] Moreover, the relative admissions of high-skilled professionals to low-skilled workers significantly increased within the employment-based admissions.

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. Figures 2.5 and 2.6 detail the response. Figure 2.5 plots the number of ISEs granted permanent residency in the US from 1983-1997 for selected ethnicities (summed over 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.[33]

National Science Foundation 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 2.6 exhibits the percentage intending to remain in the US for available countries. The 60% to 90% jump for Mainland China from 1990 to 1992 is striking. Substantial increases are also apparent

[31] 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.

[32] The employment limit increased to 140,000, but 120,120 corresponds to the previously restricted categories.

[33] Kerr (2005a) shows that low-skilled immigration remained constant during the 1990 reform; temporary versus permanent immigration and the responses of other skilled occupations are also discussed. ISE trends are developed from immigrant-level INS records using the Engineers, Natural Scientists, and Mathematical and Computer Scientists occupations.

for India and Western Europe.

The second 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 2.5 and 2.6 to determine treatment and control groups. Kerr (2005a) undertakes a formal analysis of researcher immigration responses to the legislation change, finding the constrained countries to be India, Mainland China, the Philippines, and Taiwan. The reduced-form immigration estimator (2.10) then takes the form

$$\ln(IMM_{ct}^{RF}) = \ln\left[\sum_{s=1}^{5}(QUOTA_{c,t-s}^{Eff} + QUOTA_{c,t-s-1}^{Eff} + QUOTA_{c,t-s-2}^{Eff})\right], \qquad (2.11)$$

where $QUOTA_{ct}^{Eff}$ is the effective quota for country c in year t. Raising the numerical ceilings did not change the effective quota for nations unconstrained by the former immigration regime, and their effective quota is held constant at a pre-reform theoretical limit of the 20,000 country-level quota multiplied by the 20% employment allocation. The effective quota for the four constrained countries 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 the shift is moved forward to 1990 for Mainland China to account for the Chinese Student Protection Act.[34]

Table 2.4 documents the reduced-form specifications (2.8) using the immigration quotas estimator (2.11) in place of the ethnic patenting estimator. The format of Table 2.4 mirrors Table 2.3. Exports are found to increase with an elasticity of about 0.5 to the exporter's scientific integration with the US. Although the β_H coefficients from the two reduced-form estimators should not be directly compared, their similar qualitative directions do provide confidence that reverse causality is not responsible for the positive reduced-form elasticities

[34]The extremely large Chinese response and sharp decline in Figure 2.5 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 is factored into the timing of the reduced-form estimator.

evident in Table 2.3. The elasticity of exports to the importer's technology integration is negative with this quotas-based estimator. If the gravity covariates are excluded, however, this elasticity is instead positive and very small.[35]

2.4.4 TFP Estimations

With both instruments and their reduced-forms described, this subsection completes the country-level analysis with OLS and IV regressions of bilateral export volumes on the TFP indices (a proxy for the country technology location parameter T_i). Column 1 of Table 2.5 estimates the structural specification (2.7). While the TFP indices vary by country, the standard errors are still clustered by ethnicity for comparison to the IV regressions. In contrast to Tables 2.3 and 2.4, the bilateral export elasticity to the exporter's TFP development is much weaker (0.2) and statistically insignificant when the gravity covariates are included. Column 2, however, shows a much stronger response when the gravity covariates are excluded. Figures 2.7 and 2.8 illustrate these regressions in a form similar to Figures 2.3 and 2.4.

The large coefficient swing from Column 1 to Column 2 brings to the forefront an underlying data concern. Unlike the reduced-form estimations, the TFP indices and the gravity covariates are derived from the same PWT data and are highly colinear by construction (recall the 92% correlation in growth rates for the TFP indices and the exporters' GDP per capitas). As a result, the TFP elasticity is very sensitive to small specification changes. For example, substituting the non-interacted exporter's and importer's GDPs and GDP per capitas for the two interacted covariates restores the measured elasticity to 0.8. Moreover, it is questionable whether the gravity covariates should be included in the upcoming IV regressions anyway due to their endogeneity. Fortunately, the IV regressions are robust to including or excluding the gravity variables, and Table 2.5 maintains the same baseline estimation as the reduced-form exercises for consistency.

Examining the remainder of Table 2.5, the balanced panel specifications of Columns 5 and 6 find a stronger and statistically significant elasticity of exports to the exporter's TFP index, while the elasticity to the importer's TFP index is negative and statistically insignificant. Column 8 shows the positive elasticity derives from the latter part of the sample, while the

[35] Graphs for the immigrations quotas estimator are presented in the appendix.

sample decompositions find the Chinese and Indian outcomes are most responsible for the partial correlations evident between export and TFP growth. As noted in Section 2.3, TFP indices are unavailable for the former Soviet Union and Vietnam (as well as Belize and Cuba), and these two ethnicities are dropped from the sample considered in Figures 2.7 and 2.8 and Tables 2.5 and 2.6.

The core IV regressions using the US ethnic human-capital stocks instruments are presented in Table 2.6A. The first four columns document again the OLS permutations with and without the importer's TFP regressor and gravity covariates. As noted earlier, the inclusion of the gravity covariates substantially reduces the coefficient on the exporter's TFP regressor. Columns 5-8 of Table 2.6A instrument for the exporter's and importer's technology development with the ethnic human-capital stocks developed from the US patenting dataset. The first-stage regressions in Panels B and C demonstrate that ethnic scientific integration with the US improves the TFPs of their home countries. The gravity covariates are included in the first-stages of Columns 6 and 8, but the coefficients are not reported to conserve space. Figures 2.9 and 2.10 plot the first-stage regressions for the exporters, highlighting that technology transfer is most evident for the Chinese, Indian, and Korean ethnicities. The first-stage relationship is weaker among European economies and is not present for Latin America or Japan unless the gravity covariates are included.

In contrast to the OLS regressions, the instrumented elasticities for the exporter's technology state are strong and statistically significant regardless of whether the gravity covariates are included. The IV specifications are able to overcome the measurement error and colinearity problems from the coarse TFP proxy. The estimated elasticity for the importer's technology state, on the other hand, is not robust to the including the gravity covariates. Unfortunately, the null hypothesis that the two elasticities are equal cannot be rejected due to the large standard errors for the importer's technology state, but the greater importance of the exporter's technology does appear to hold true.

Columns 5-8 of Table 2.6B present the IV regressions using the immigration-based instruments. The instrumented export elasticities of 1.1-1.7 to the exporter's technology state are 50% larger than those using the patent-based instrument; all of the elasticities but Column 8 are again statistically significant. Moreover, the importer's technology development is not found to increase export volumes and is statistically different from the exporter's contribution in Column 7. Taken together, these IV regressions find strong support for Ricardian technology differences as a source of trade.[36]

A definitive explanation for the larger elasticity with the immigration-based instrument cannot be given, but two hypotheses are readily identified. First, the theoretical development in Section 2.2 assumes both constant country sizes and that the labor resources of each country are fully employed in manufacturing. Several countries in the sample, however, have large reservoirs of underutilized labor in agriculture, and the transition of these workers to manufacturing is important for characterizing their economic development (e.g., Harris and Todaro 1970). Kerr (2005a) finds that technology transfer from the US to these emerging economies produces a larger growth in manufacturing output compared to industrialized economies due employment growth from sector reallocation complementing labor productivity gains.

In the current framework, this transition process is equivalent to an increase in effective country size. If wage equality with the agricultural sector is also maintained, general-equilibrium increases in the input costs c_i for the manufacturing sector are also depressed. Both effects further promote growth in export volumes. The immigration-based instrument focuses attention on three economies undergoing this transition. The 1980 agriculture employment share for Mainland China, India, and the Philippines are 70%, 74%, and 52%, respectively, compared to 8% in Taiwan and a sample mean of 22%.[37] Additional sector reallocation to manufacturing in these economies following technology transfer from the US may explain part of the larger elasticity.

A second candidate explanation, however, is an omitted variable bias. The reduced variation inherent in the immigration estimator's design potentially exposes it to correlation with omitted factors that differentially affect the four treatment economies from the control group around the 1990 Act. To take one case, India undertook several trade reforms in the early 1990s as part of an IMF-supported adjustment program (e.g., Topalova 2004). While it is possible to argue these reforms are endogenous outcomes to facilitate the expansion of exports, it certainly

[36]Table 2.A2 in the appendix shows the IV results are robust to the sample decompositions undertaken in the reduced-form analysis. The only major departure is for the immigration quotas instrument when the Chinese ethnicity is excluded.

[37]Agricultural shares are from the United Nations Statistical Division and Sun, Fulginiti, and Peterson (2003).

conceivable that India's reforms are in fact independent and potentially bias technology's role with the immigration IV upward. The price of the exogeneity in the immigration quotas instrument is greater exposure to these types of concerns.

To summarize the material presented in this section, two estimators are developed to model heterogeneous technology diffusion from the US through ethnic scientific networks. The first uses ethnic patenting data to measure the human-capital stocks of each ethnicity with respect to US innovations, while the second exploits exogenous changes in the sizes of US ethnic scientific communities following the Immigration Act of 1990. In both reduced-form and IV specifications, the exporter's integration with the US frontier consistently yields growth in export volumes, while the importer's integration has an inconsistent and often negligible effect. Taken together, these country-level regressions provide strong evidence that Ricardian technology differences across countries are important for explaining trade patterns.

2.5 Empirical Extensions

The bilateral perspective taken in Section 2.4's country-level regressions affords a detailed contrast of the exporter's and importer's technology development. The large number of exporter-importer permutations, however, prohibits the use of much of the underlying industry and geographic variation available in the ethnic patenting and trade datasets. This section undertakes two extensions that highlight instead these other dimensions.

The first extension sums multilateral exports across destination countries to focus on industry-level exports for each country. Technology advancement is again found to increase exports at this disaggregated level. Moreover, this result is robust to including detailed country-level trends that remove aggregate changes for the manufacturing sectors in each nation. This finding argues against alternative hypotheses for the earlier results that would operate at the country-level (e.g., entry into trading blocs). Through matching exercises that further group industries according their capital and skilled-labor intensities, the industry-level analysis also provides confidence that a generalized Rybczynski effect due to factor accumulation is not responsible for the positive elasticities assigned to technology growth.

The second extension returns to the bilateral trade data to explore how the elasticity estimates differ by different geographic distances to import destinations. Export growth is strongest in bordering and nearby countries, but positive growth is evident at all distances.[38]

2.5.1 Industry Analysis

The industry-level analysis sums exports for each country i across import destinations by the ISIC industries listed in Table 2.A1. For comparison, non-US imports are also calculated for each economy. As in the bilateral exercises, some observations have zero values that are undefined in a log specification. The multiple specification checks undertaken in Table 2.7 (e.g., first differences, time trends) are best conducted with panels of uniform series length. Accordingly, country-industry observations are excluded if they do not maintain \$10k in non-US multilateral exports and imports throughout the 1980-1997 period. A second bar of one ethnic patent per annum is also applied. These hurdles focus the analysis on economically important interactions, but the results are robust to raising or removing these hurdles.

The basic industry-level regressions for country i and industry j in year t take the form

$$\ln(X_{ijt}) = \alpha + \beta \ln(H_{ijt}) + \gamma \ln(GDP/CAP_{it}) + \zeta \ln(GDP_{it}) + \phi_{ij} + \eta_{jt} + \epsilon_{ijt}. \qquad (2.12)$$

All of the industry-level specifications employ the detailed US ethnic human-capital stocks H_{ijt} that are constructed through the ethnic patenting database. The regressions retain the exporting country's gravity covariates and include vectors of country-industry fixed effects ϕ_{ij} and industry-year fixed effects η_{jt}. In Table 2.7, Column 1 of the top panel finds a β coefficient of 0.7 when estimating (2.12). As in the country-level regressions, positive integration with the US technology frontier increases manufacturing exports. The industry-level elasticity is statistically significant, with the standard errors now clustered at the ethnicity-industry level to reflect the more disaggregated data employed.

By itself, this industry-level regression provides two important checks on the earlier results. First, the data development in Section 2.3 asserted that five-year sums of ethnic patents could proxy for the underlying sizes of ethnic research communities because changes in the patenting productivity of US researchers would be absorbed into the aggregate year effects η_t. This

[38] Industry-level analyses do not follow directly from the Eaton and Kortum model, which requires random industry technologies around the country technology aggregate to satisfy the Fréchet distribution.

assumption may be cavalier, however, as the patenting rate per researcher likely increased faster in high-tech sectors than in more mundane fields. With ethnicities specializing in different industries (e.g., the US Chinese in computer research), this composition effect could potentially bias the results. By narrowing the focus to within-industry variation, this liability is minimized because the patenting productivity is now controlled for at the industry-year level by the fixed effects η_{jt}. Likewise, differential changes in industry prices (e.g., the rapid decline in computer prices) in the nominal export volumes are now captured.

The second column in Panel A weights the observations by each industry's patenting level from 1980-1982. This weighting scheme focuses more attention on high-tech industries, finding a small growth in the estimated elasticity. The stability of the elasticity across Columns 1 and 2 highlights, however, how pervasive the industry-level response of exports to technology integration is.

The country-level exercises found weak export elasticities to the importer's technology development. The analogous industry-level prediction is that non-US multilateral imports should respond less to scientific integration with the US frontier than exports. This Ricardian outcome is evident in Columns 3 and 4, which present the unweighted and weighted regressions for imports using a specification similar to (2.12). Unreported regressions further find that multilateral imports respond more to the scientific integration of other ethnicities to the US frontier. This contrast between exports and imports again emphasizes the role of technology in determining trade patterns. Moreover, the asymmetry argues against omitted trade agreements affecting the outcomes measured.

Many empirical analyses first difference a levels specifications for estimation,

$$\Delta \ln \left(X_{ijt}\right) = \alpha + \beta \Delta \ln \left(H_{ijt}\right) + \gamma \Delta \ln (GDP/CAP_{it}) + \zeta \Delta \ln \left(GDP_{it}\right) + \eta_{jt} + \tilde{\epsilon}_{ijt}, \qquad (2.13)$$

where $\tilde{\epsilon}_{ijt} = \epsilon_{ijt} - \epsilon_{ijt-1}$. The efficiency of a first-differences form versus the levels specification turns on whether the error term ϵ 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 (2.12) find moderate serial correlations of 0.5-0.6. Columns 5-8 of Panel A show the Ricardian findings are evident in a first-differences form too, although the measured export elasticity is weaker at 0.2-0.3.[39]

The lower panel in Table 2.7 presents a battery of specification checks to the base regressions from Panel A. The tabulated coefficients are from separate regressions that modify the base regressions as indicated in the left-hand column; the coefficients for the gravity covariates are again omitted to conserve space. The first row of Panel B excludes the computer and drug industries. 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), it would be important to note if ethnic differences in high-tech industries alone are responsible for the positive correlations. The minor elasticity changes indicate that they are not.

More importantly, the bottom four rows introduce aggregate ethnicity-level and country-level controls. The power of the industry-level analysis is its ability to isolate within-country variation, even after removing levels differences and global industry trends. A positive β coefficient in these regressions requires that China's exports in the computer industry grow faster than its exports in the pharmaceuticals industry if Chinese integration to the US computer frontier grows faster than its integration to the pharmaceutical frontier. Strong performance in these regressions places this same burden on any competing explanations.

The elasticity estimates are remarkably stable to these country-level controls. The specifications become more stringent as one moves down the rows from the linear time trends to nonparametric country-year effects; the coefficient estimates naturally decline as the additional variation is removed. Yet, a strong elasticity of export growth to scientific integration with the US frontier is maintained even when the triple combination of country-industry, industry-year, and country-year fixed effects are introduced. Moreover, the asymmetric growth of multilateral exports over multilateral imports is maintained throughout. The robust conclusion from Table 2.7 is that Ricardian technology differences are important determinants of industry-level

[39] As first differencing exacerbates the downward bias in estimated coefficients due to measurement error in the regressor, the levels form remains the preferred specification in this study. GLS estimations that correct the autoregressive parameter yield elasticity estimates bounded by the levels and first differences results presented. Lagged dependent variable specifications that test for mean reversion also demonstrate export growth with scientific integration.

patterns even after aggregate country trends are removed.

2.5.2 Testing for the Rybczynski Effect

In contrast to the Ricardian framework, Heckscher-Ohlin-Vanek (HOV) models describe trade as resulting from factor differences across countries (e.g., labor, capital, natural resources). During the period studied, some countries experienced significant growth in their skilled labor forces and physical capital stocks, as well as their technology sets, and the former could lead to growth in manufacturing exports due to the Rybczynski effect. Table 2.7's finding of strong technology elasticities using only within-country variation already argues against this interpretation for the manufacturing sector as a whole (e.g., a shift from agriculture to manufacturing due to capital accumulation). This subsection provides additional evidence that technology's role is not reflecting changes in trade patterns due to factor accumulations by using industry comparisons within the manufacturing sector itself.

The intuition behind the proposed test is straightforward. Under the Rybczynski effect, the accumulation of skilled workers in country i shifts country i's specialization towards manufacturing industries that employ skilled labor more intensively than other factors. By grouping manufacturing industries by their skilled-labor intensities, tests examine if technology's important role is preserved after time trends are removed for these industry groups within each country. To illustrate, both the computer and pharmaceutical industries are highly skill intensive. A general Rybczynski effect due to skilled worker accumulation in China would favor specialization and export growth in these industries equally. Additional confidence for technology's role is warranted if China's exports grow faster in the skill-intensive industry that receives the strongest technology transfer from the US *relative to its peer industries.*

To implement this matching exercise, industries are grouped into quintiles based upon their mean 1980-1997 factor intensities in the US. Three intensities are studied — the industry's capital-labor ratio, the share of non-production workers in the industry's labor force, and the industry's average wage. Table 2.A1 lists for each industry the quintile groupings assigned. Textiles rank in the lowest quintiles in all three classifications schemes, while chemicals and industrial machinery consistently fall into the top quintiles. Some differences do exist though. The correlations among quintile groupings are 76% for capital-labor and wage, 59% for wage and non-production share, and 37% for capital-labor and non-production share.

These detailed comparisons place a premium on the number of industries available for each country. Accordingly, the matching test employs a slight variant of (2.12) that regresses the share of country i's manufacturing exports in industry j on the share of its US ethnic human-capital stock in industry j,

$$\left(\frac{X_{ijt}}{X_{it}}\right) = \alpha + \beta\left(\frac{H_{ijt}}{H_{it}}\right) + \phi_{ij} + \eta_{jt} + \epsilon_{ijt}. \tag{2.14}$$

The regression retains the vectors of country-industry fixed effects ϕ_{ij} and industry-year fixed effects η_{jt}. This regression is similar to the specification that included country-year fixed effects, but without the log transformation the zero export values can be incorporated.

The first column in Table 2.8 finds an unweighted β coefficient of 0.7 with this modified specification. The next three columns incorporate into (2.14) the ethnicity time trends interacted with each quintile group. For all three factor intensity comparisons, the positive role of technology is preserved. The robustness of technology's role in this matching exercise is also evident in the weighted regressions of Columns 5-8.[40]

These findings suggest an omitted factor accumulation is not confounding the strong role for technology identified throughout this study. Of course, the test is not foolproof. In particular, specialized foreign development could produce both export shifts within intensity groups and the emigration of ethnic researchers trained in these specialized fields to the US. The earlier specifications employing the US immigration quotas argued against this reverse causality interpretation, but this estimator does not provide the industry-level variation necessary for this analysis. Nevertheless, this study concludes that the importance of Ricardian technology differences identified here are robust to alternative factor-based explanations of trade.[41]

[40] Vietnam is excluded as its US ethnic patenting is very small in the early years of the sample, producing several outliers in this industry-comparison exercise. The coefficient estimates are lower if Vietnam is included, but they remain statistically significant and convey the same results.

[41] The ideal test would simply remove factor-based trade from the export volumes studied. This is test is unattainable for several theoretical and practical reasons. First, while 2x2x2 HOV models (two countries, factors, and goods) cleanly predict a country exports goods that intensely use the factors in which the country is well endowed, this prediction does not hold universally in settings with multiple goods and factors (e.g., the critique of Leamer (1980) on Leontief's (1953) paradox). Likewise, bilateral trade patterns due to factor-based differences are only determined for special cases in a multi-country world (e.g., Romalis 2004). Thus, strong assumptions would be required for distinguishing factor-based trade in this empirical setting. Practically speaking, the data constraint is also prohibitive as factor data and industry input-output matrices are very poorly

2.5.3 Geographic Expansion

This study closes with some initial findings regarding the geographic margin of export growth due to technology advances. This margin is quantified in the reduced-form specification (2.8) for country-level bilateral exports. Table 2.9 begins by repeating the basic regression with the bilateral exports for which distances are available (using the Great Circle distances between capital cities). The sample is then divided into bordering countries in Column 2 and non-bordering countries in Columns 3-7. Non-bordering countries are further separated into five distance categories: 0-1500 km., 1501-3000 km., 3001-6000 km., 6001-9000 km, and greater than 9000 km. To give a feel for these demarcations, the distances from Beijing, China, to the capitals of Taiwan, Bangladesh, United Arab Emirates, and Spain are 1723 km., 3029 km., 5967 km., and 9229 km., respectively.

The estimated elasticities suggest export growth declines slightly with distance and in a non-linear way. The strongest elasticities are found for bordering countries, although the small sample size yields substantial standard errors, and non-bordering countries within 1500 km. The remaining groupings, however, are not very different from the average elasticity estimate and do not display an obvious trend. This latter stability is a comforting result as it suggests local shocks (e.g., regional trade agreements, macroeconomic cycles) are not responsible results. A complete characterization of the interaction between distance and export growth from technology advancement is left for future work.

2.6 Conclusions

While the principle of Ricardian technology differences as a source of trade is well established in the theory of international economics, empirical evaluations of its importance are relatively rare due to the difficulty of quantifying and isolating technology differences. This study exploits heterogeneous technology diffusion from the US through ethnic scientific networks to make additional headway using panel estimation techniques. Country-level regressions find bilateral exports respond positively to the exporter's technology development. This result is robust

measured for most of the countries and years covered by this study. See Davis and Weinstein (2001) for an application using richer OECD data.

to 1) including the importer's technology development as a regressor, 2) instrumenting for technology development using scientific integration with the US frontier, 3) testing for reverse causality using an exogenous reform of US immigration quotas, and 4) considering industry-level specifications that focus on within-country variation and test for factor accumulation alternatives. The results strongly support the conclusion that technology differences are an important determinant of trade.

Several promising extensions are currently being pursued using the additional geographic and industry variation discussed in Section 2.5. The first project seeks a fuller characterization of the industry and geographic margins of export expansion (including their joint interactions for intensive and extensive growth). This research will bring specific attention to industry entry and exit decisions and characterize the importance of different country sizes. A second project is exploring the impact of technology transfer for trade policies and political institutions (and vice versa). This additional work will further refine our understanding of how Ricardian technology differences influence trading patterns among countries.

Bibliography

[1] Alvarez, Fernando, and Robert Lucas, "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade", Working Paper (2004).

[2] Antweiler, Werner, and Daniel Trefler, "Increasing Returns and All That: A View from Trade", *The American Economic Review*, 92:1 (2002), 93-119.

[3] Bartelsman, Eric, and Wayne Gray, "The NBER Manufacturing Productivity Database", NBER Technical Working Paper 205 (1996).

[4] 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.

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

[6] Caves, Douglas, Laurits Cristensen, and Erwin Diewert, "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers", *Economic Journal*, 92:365 (1982), 73-86.

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

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

[9] Davis, Donald, and David Weinstein, "An Account of Global Factor Trade", *The American Economic Review*, 91:5 (2001), 1423-1453.

[10] Dornbusch, Rudiger, Stanley Fischer, and Paul Samuelson, "Comparative Advantage, Trade and Payments in a Ricardian Model with a Continuum of Goods", *The American Economic Review*, 67:5 (1977), 823-839.

[11] Dornbusch, Rudiger, Stanley Fischer, and Paul Samuelson, "Heckscher-Ohlin Trade Theory with a Continuum of Goods", *The Quarterly Journal of Economics*, 95:2 (1980), 203-224.

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

[13] Eaton, Jonathan, and Samuel Kortum, "Technology, Geography, and Trade", *Econometrica*, 70:5 (2002), 1741-1779.

[14] Feenstra, Robert, "World Trade Flows, 1980-1997", Working Paper (2000).

[15] Frankel, Jeffrey, *Regional Trading Blocs in the World Economic System* (Washington, DC: Institute for International Economics, 1997).

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

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

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

[19] Hall, Bronwyn, "Exploring the Patent Explosion", NBER Working Paper 10605 (2004).

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

[21] Harrigan, James, "Cross-Country Comparisons of Industry Total Factor Productivity: Theory and Evidence", Federal Reserve Bank of New York Research Paper 9734 (1997a).

[22] Harrigan, James, "Technology, Factor Supplies, and International Specialization: Estimating the Neoclassical Model", *The American Economic Review*, 87:4 (1997b), 475-494.

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

[24] Heckscher, Eli, "The Effect of Foreign Trade on the Distribution of Income", *Ekonomisk Tidskrift*, 21:2 (1919), 1-32.

[25] 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).

[26] Helpman, Elhanan, and Paul Krugman, *Market Structure and Foreign Trade* (Cambridge, MA: MIT Press, 1985).

[27] Heston, Alan, Robert Summers, and Bettina Aten, Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002.

[28] Hunter, Linda, "The Contribution of Non-Homothetic Preferences to Trade", *Journal of International Economics*, 30:3 (1991), 345-358.

[29] Hunter, Linda, and James Markusen, "Per Capita Income as a Determinant of Trade", in Robert Feenstra, ed., *Empirical Methods for International Trade* (Cambridge, MA: MIT Press, 1988).

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

[31] Jasso, Guillermina, Mark Rosenzweig, and James Smith, "The Changing Skill of New Immigrants to the United States: Recent Trends and Their Determinants", NBER Working Paper 6764 (1998).

[32] Johnson, Daniel, "150 Years of American Invention: Methodology and a First Geographic Application", Wellesley College Economics Working Paper 99-01 (1999). Data currently reside at http://faculty1.coloradocollege.edu/~djohnson/uships.html.

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

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

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

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

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

[38] Kerr, William, "Ethnic Scientific Communities and International Technology Diffusion", Working Paper (2005a).

[39] Kerr, William, "Ethnic Scientific Networks and the Facilitation of Trade", Working Paper (2005b).

[40] Kerr, William, "The Ethnic Composition of US Inventors", Working Paper (2005c).

[41] Kim, Jinyoung, and Gerald Marschke, "Accounting for the Recent Surge in U.S. Patenting: Changes in R&D Expenditures, Patent Yields, and the High Tech Sector", *Economics of Innovation and New Technologies*, 13:6 (2004), 543-558.

[42] Kortum, Samuel, and Joshua Lerner, "Assessing the Contribution of Venture Capital to Innovation", *RAND Journal of Economics*, 31:4 (2000), 674-692.

[43] Leontief, Wassily, "Domestic Production and Foreign Trade: The American Capital Position Re-examined", *Proceedings of the American Philosophical Society*, 97 (1953), 332-349.

[44] Leamer, Edward, "The Leontief Paradox, Reconsidered", *The Journal of Political Economy*, 88:3 (1980), 495-503.

[45] MacDougall, G.D.A., "British and American Exports: A Study Suggested by the Theory of Comparative Costs. Part I", *The Economic Journal*, 61:244 (1951), 697-724.

[46] MacDougall, G.D.A., "British and American Exports: A Study Suggested by the Theory of Comparative Costs. Part II", *The Economic Journal*, 62:247 (1952), 487-521.

[47] Ohlin, Bertil, *Interregional and International Trade* (Cambridge, MA: Harvard University Press, 1933).

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

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

[50] Ricardo, David, *On the Principles of Political Economy and Taxation* (London, UK: John Murray, 1817).

[51] Romalis, John, "Factor Proportions and the Structure of Commodity Trade", *The American Economic Review*, 94:1 (2004), 67-97.

[52] Scherer, Frederic, "Using Linked Patent Data and R&D Data to Measure Technology Flows," in Griliches, Zvi (ed.) *R & D, Patents and Productivity* (Chicago, IL: University of Chicago Press, 1984).

[53] Schott, Peter, "Across-Product versus Within-Product Specialization in International Trade", *The Quarterly Journal of Economics*, 119:2 (2004), 647-678.

[54] Schott, Peter, "One Size Fits All? Heckscher-Ohlin Specialization in Global Production", *The American Economic Review*, 93:2 (2002), 686-708.

[55] 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.

[56] Singh, Jasjit, "Multinational Firms and International Knowledge Diffusion: Evidence from Patent Citation Data", Working Paper (2003).

[57] 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.

[58] Stern, Robert, "British and American Productivity and Comparative Costs in International Trade", *Oxford Economic Papers*, 14:3 (1962), 275-296.

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

[60] Sun, Ling, Lilyan Fulginiti, and Wesley Peterson, "Accounting for Agricultural Decline with Economic Growth in Taiwan", Working Paper (2003).

[61] Topalova, Petia, "Trade Liberalization and Firm Productivity: The Case of India", IMF Working Paper (2004).

[62] Trefler, Daniel, "International Factor Price Differences: Leontief was Right!", *The Journal of Political Economy*, 101:6 (1993), 961-987.

[63] Trefler, Daniel, "The Case of the Missing Trade and Other Mysteries", *The American Economic Review*, 85:5 (1995), 1029-1046.

[64] Vanek, Jaroslav, "The Factor Proportions Theory: The N-Factor Case", *Kyklos*, 21:4 (1968), 749-756.

[65] Ventura, Jaume, "Growth and Interdependence", *The Quarterly Journal of Economics*, 112:1 (1997), 57-84.

[66] Wilson, Charles, "On the General Structure of Ricardian Models with a Continuum of Goods: Applications to Growth, Tariff Theory, and Technical Change", *Econometrica*, 48:7 (1980), 1675-1702.

[67] Young, Alwyn, "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore", *National Bureau of Economic Research Macroeconomics Annual*, 7 (1992), 13-63.

[68] Young, Alwyn, "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience", *The Quarterly Journal of Economics*, 110:3 (1995), 641-680.

Figure 2.1: US Ethnic Patenting

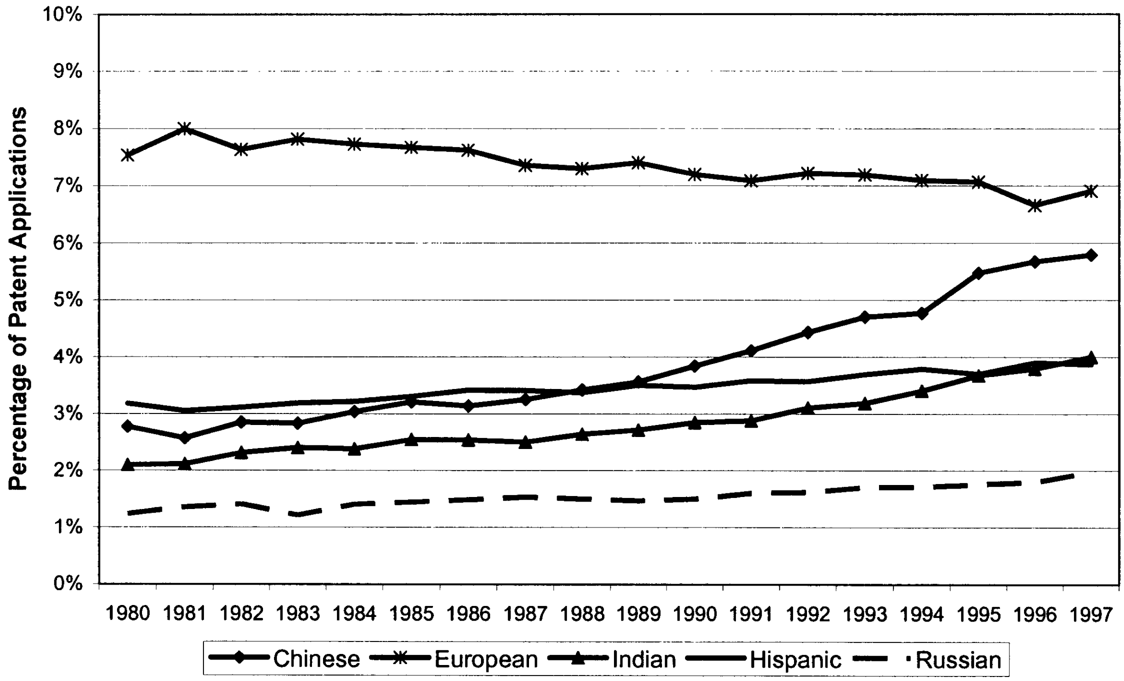


Figure 2.2: Ethnic Share by Technology

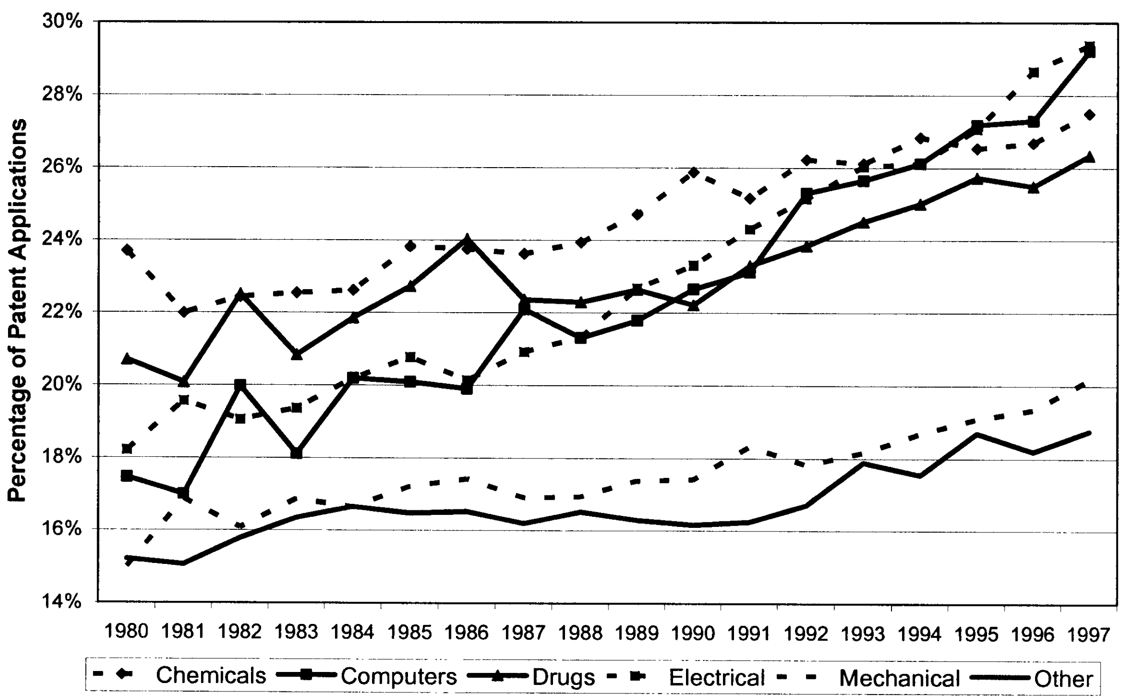


Figure 2.3: RF Multilateral Exports on US Ethnic HC Stocks

Panel Effects and Gravity Covariates Removed

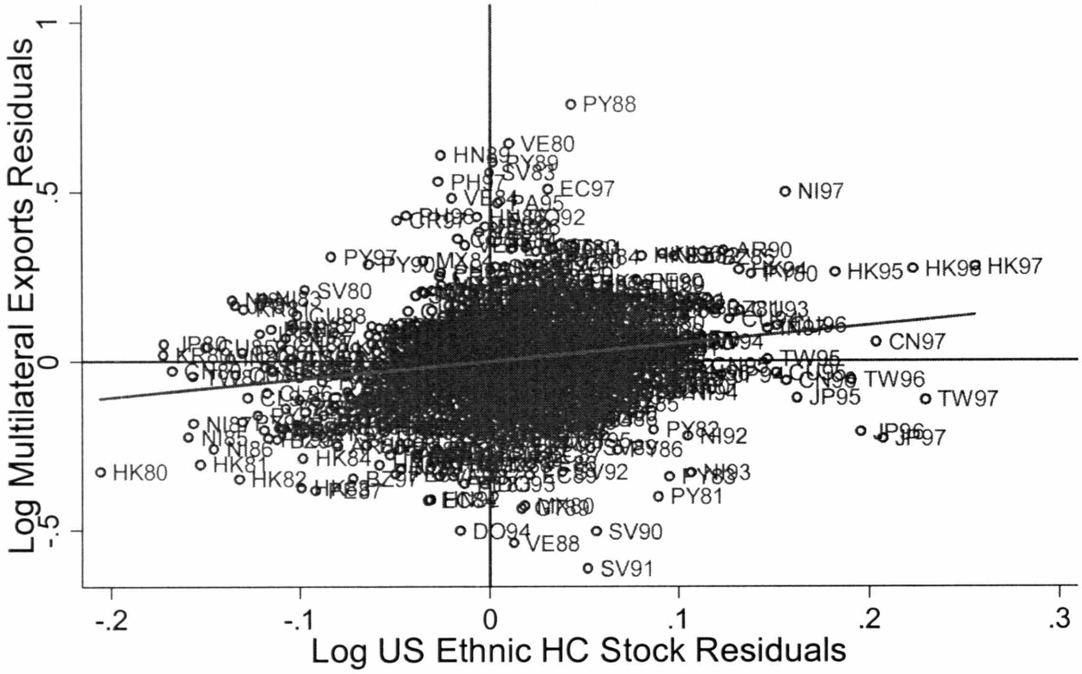


Figure 2.4: RF Multilateral Exports on US Ethnic HC Stocks

Panel Effects Removed

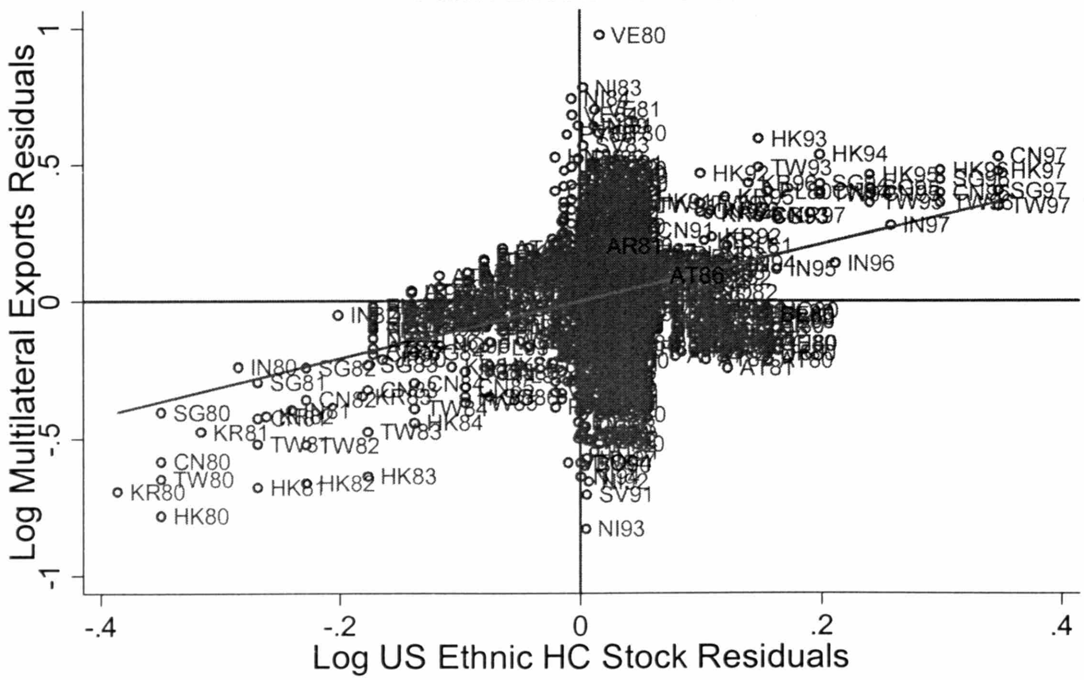


Figure 2.5: Science & Engineering Immigration

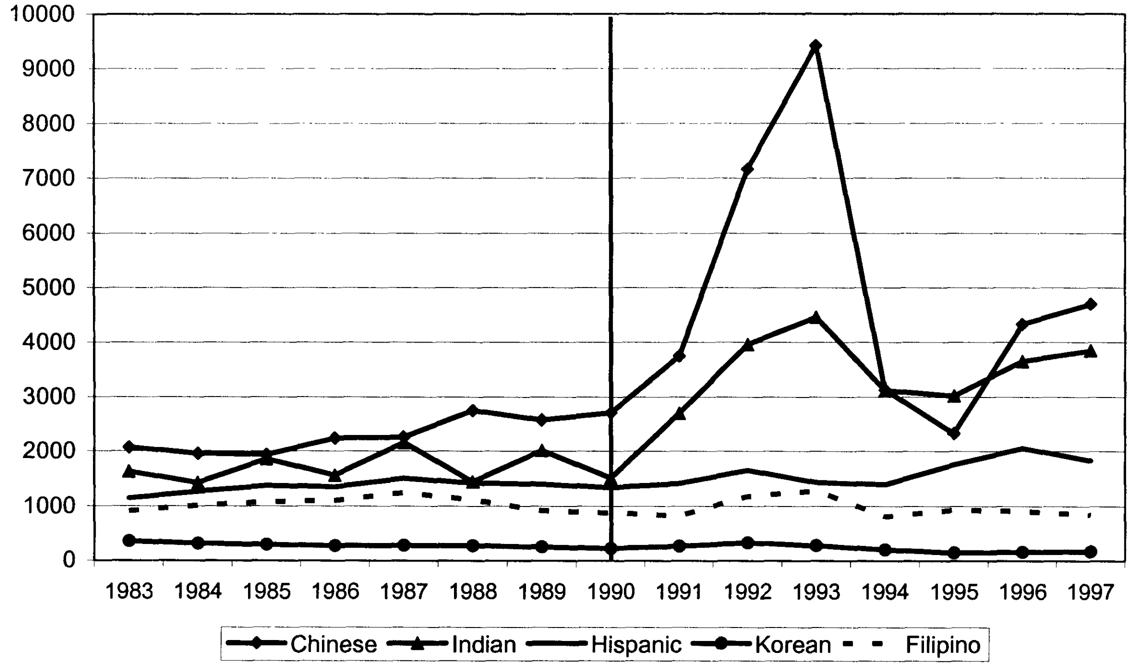


Figure 2.6: US SE Ph.D. Graduates Staying

Percentage of Graduates from Country Expecting to Stay in US

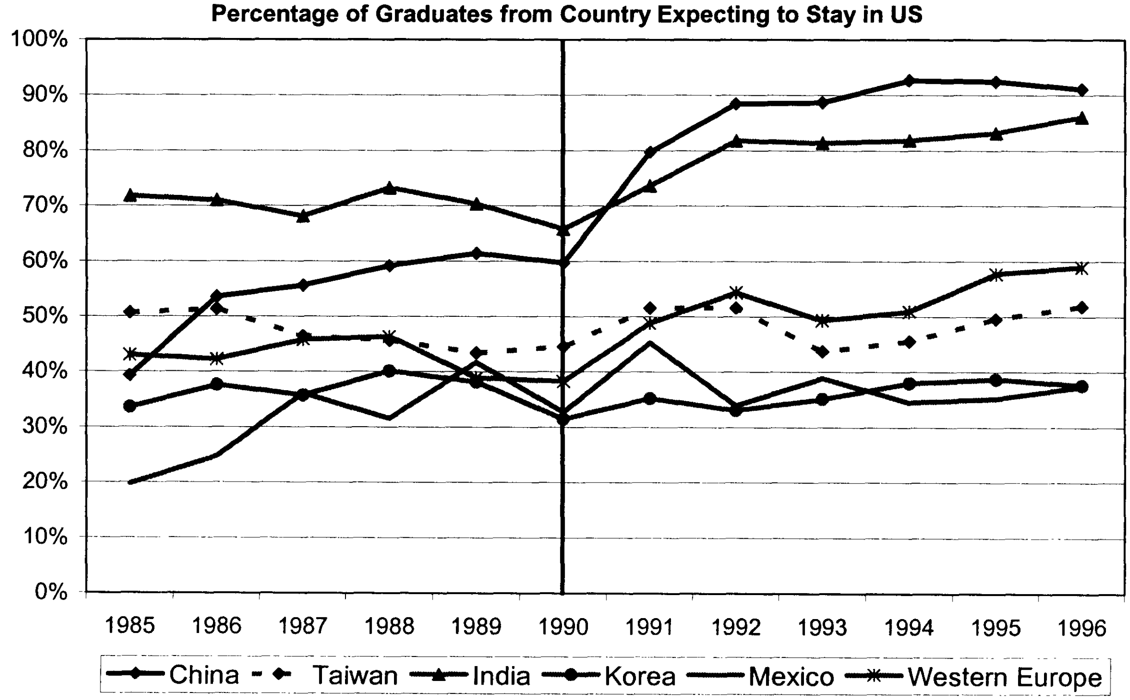


Figure 2.7: OLS Multilateral Exports on Exporter's TFP

Panel Effects and Gravity Covariates Removed

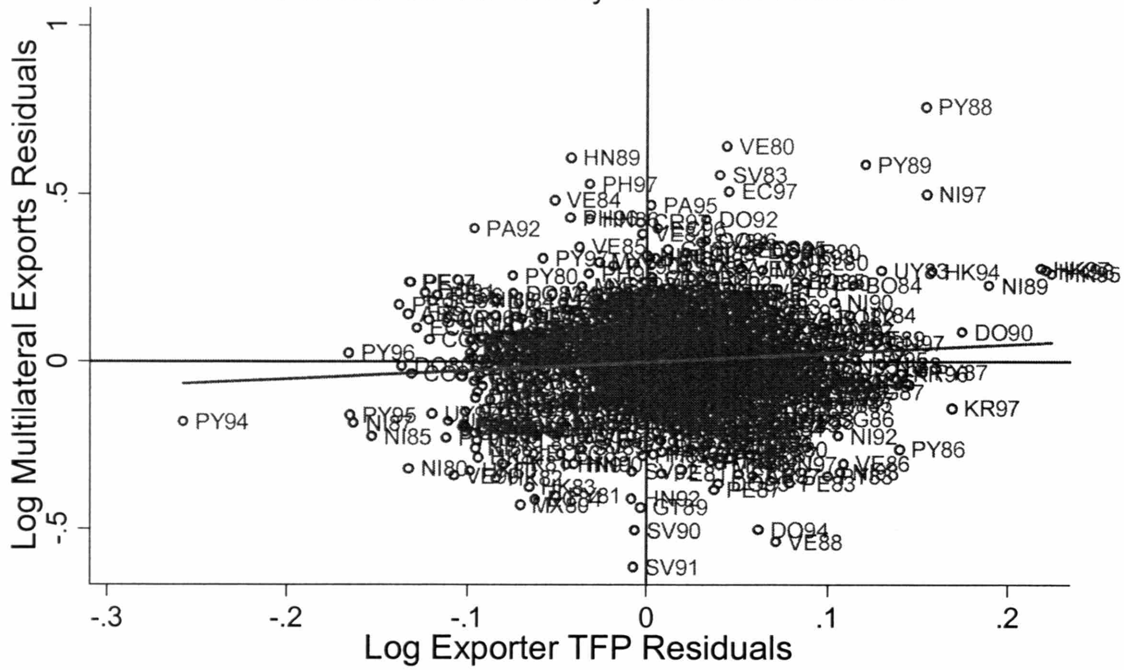


Figure 2.8: OLS Multilateral Exports on Exporter's TFP

Panel Effects Removed

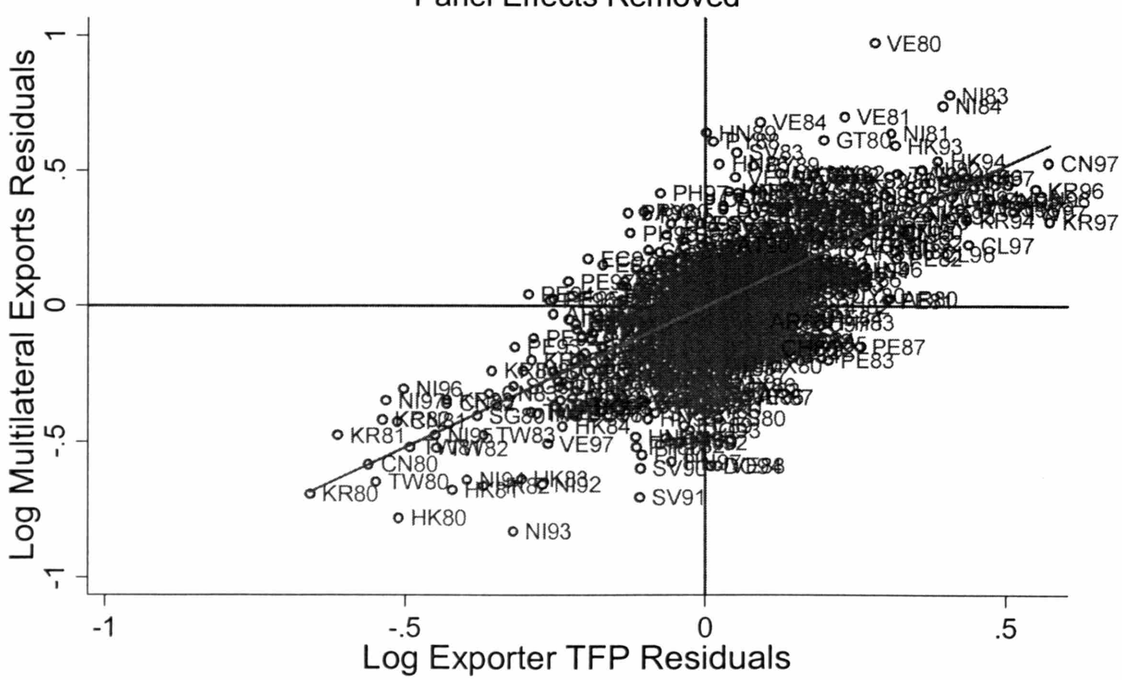


Figure 2.9: FS Exporter's TFP on US Ethnic HC Stocks

Panel Effects and Gravity Covariates Removed

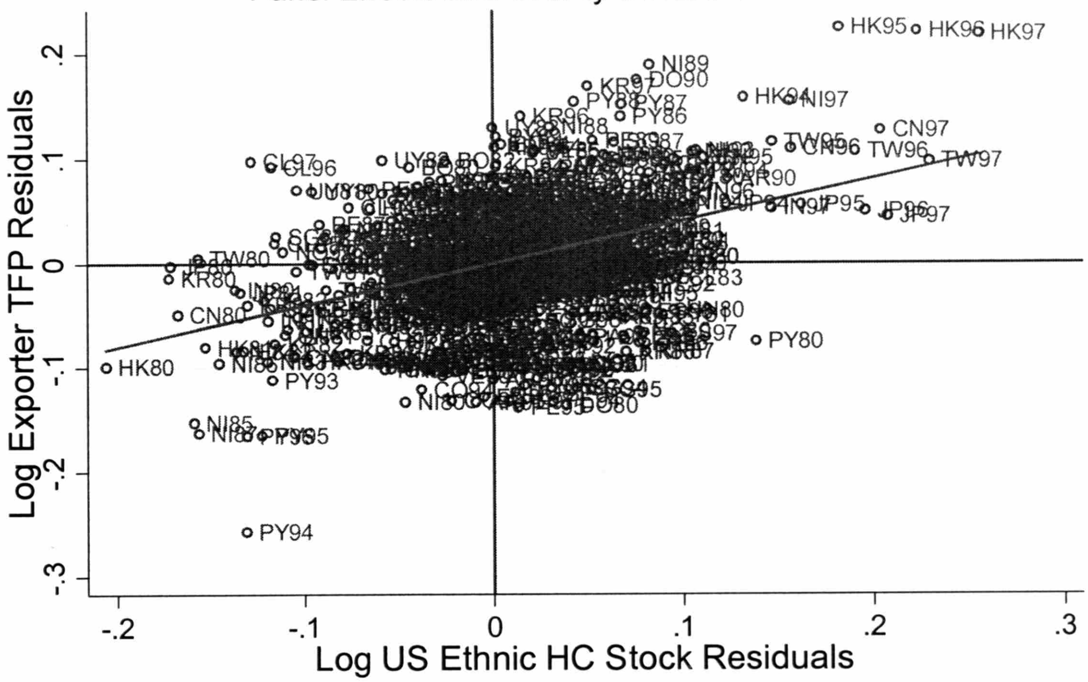


Figure 2.10: FS Exporter's TFP on US Ethnic HC Stocks

Panel Effects Removed

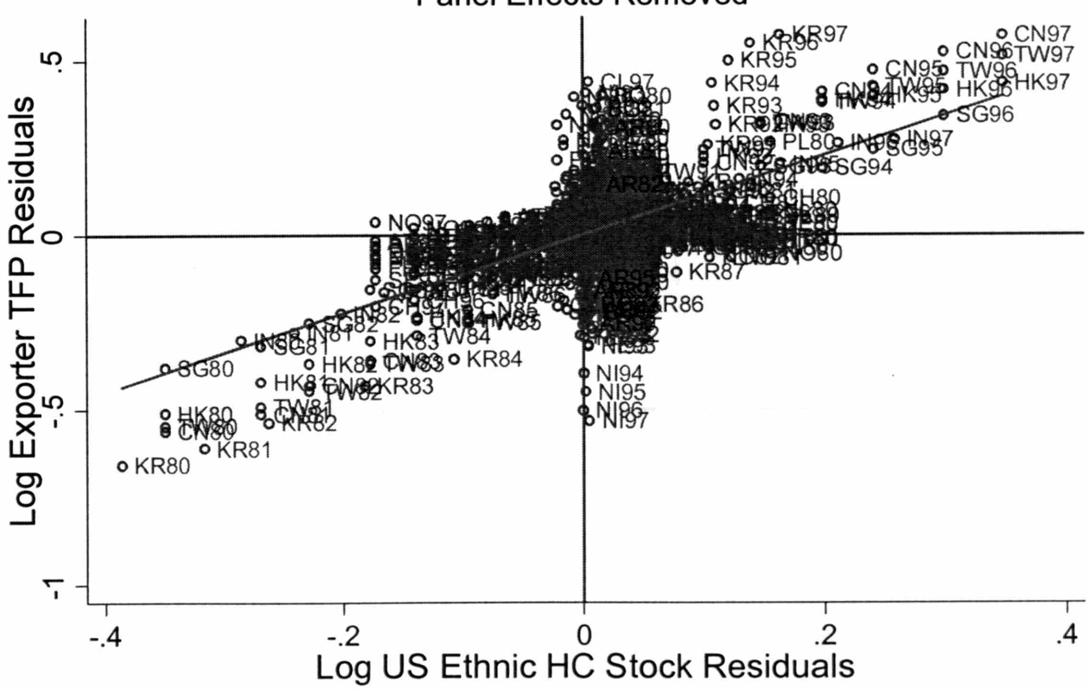


Figure 2.A1: RF Multilateral Exports on US Imm. Estimator

Panel Effects and Gravity Covariates Removed

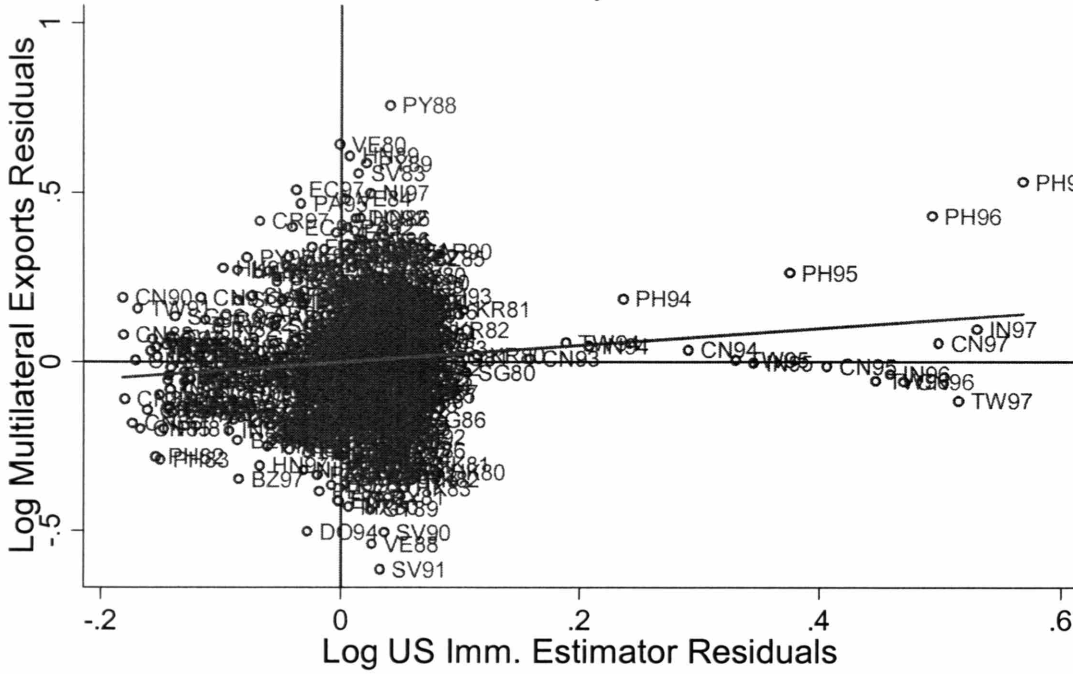


Figure 2.A2: RF Multilateral Exports on US Imm. Estimator

Panel Effects Removed

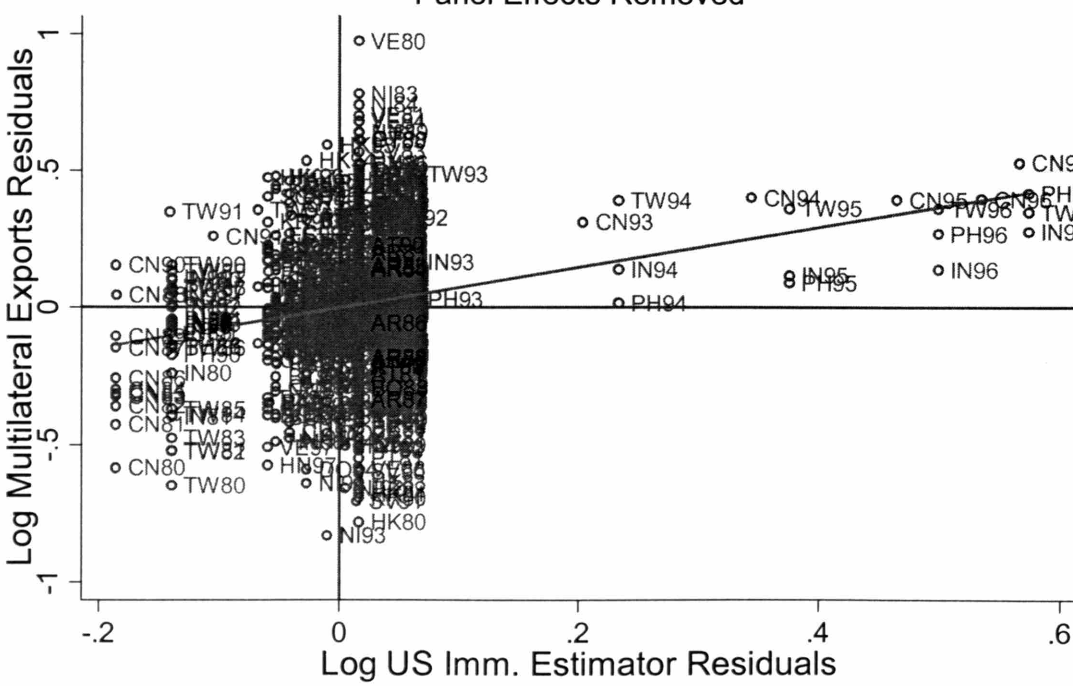


Figure 2.A3: FS Exporter's TFP on US Imm. Estimator

Panel Effects and Gravity Covariates Removed

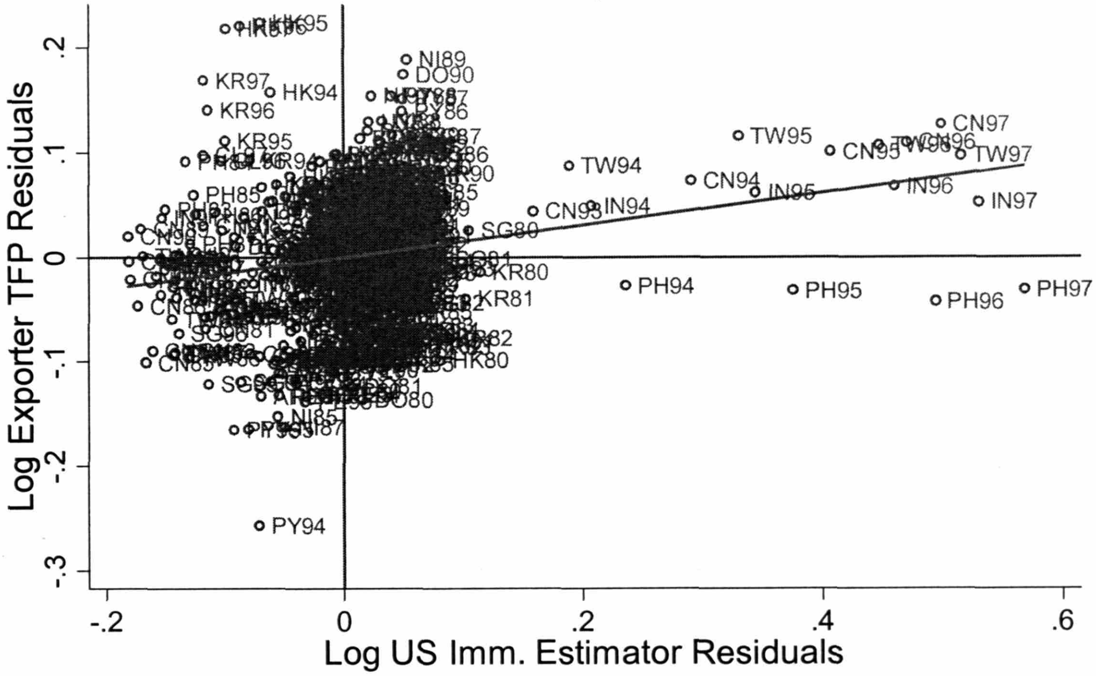


Figure 2.A4: FS Exporter's TFP on US Imm. Estimator

Panel Effects Removed

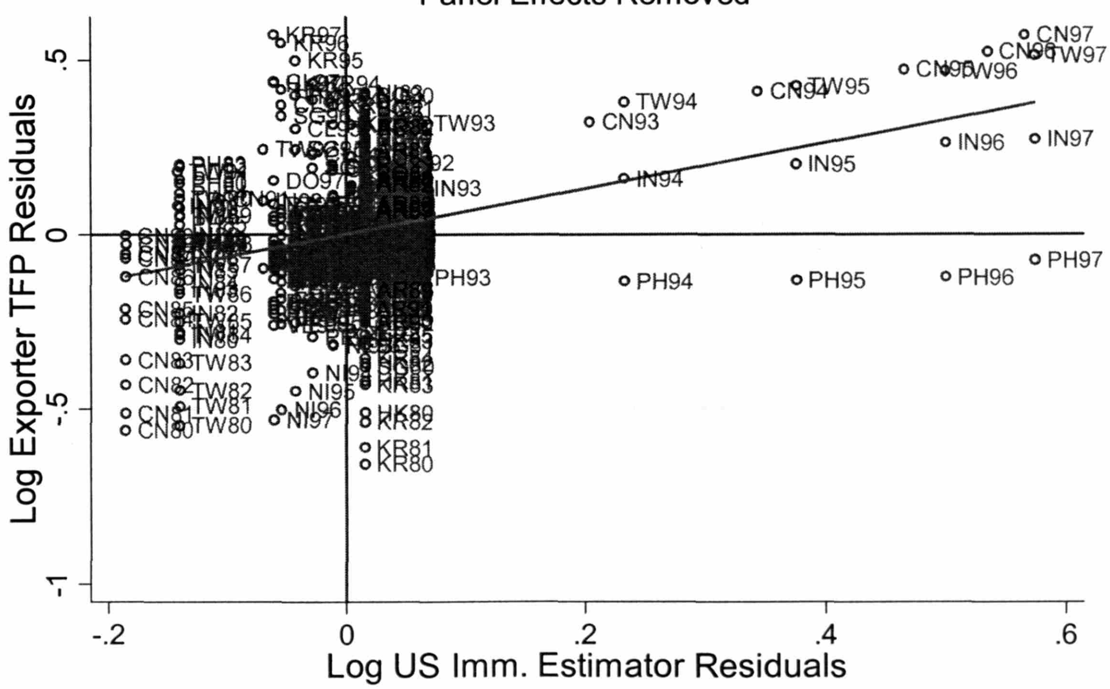


Table 2.1: Descriptive Statistics for US Ethnic Patents

	Ethnicity of Inventor (Percent Distribution)								
	English	Chinese	European	Hispanic	Indian	Japanese	Korean	Russian	Vietnamese
1980-1985 Share	81.0	3.0	7.7	3.1	2.4	0.7	0.6	1.3	0.1
1986-1990 Share	79.6	3.7	7.3	3.3	2.9	0.8	0.7	1.5	0.2
1991-1997 Share	76.4	5.4	6.9	3.7	3.7	0.9	0.8	1.7	0.4
Chemicals	75.0	6.1	7.7	3.5	4.0	0.9	0.8	1.6	0.3
Computers	75.9	6.0	6.3	3.4	4.4	0.9	0.8	1.7	0.6
Pharmaceuticals	75.8	5.0	7.6	4.0	3.7	1.1	1.0	1.6	0.2
Electrical	76.0	5.7	7.2	3.5	3.5	1.0	0.8	1.8	0.5
Mechanical	82.2	2.3	7.3	3.2	2.2	0.6	0.5	1.5	0.2
Miscellaneous	82.9	2.3	7.0	3.5	1.9	0.5	0.5	1.2	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)
1990 Bachelors	87.6	2.7	2.3	2.4	2.3	0.6	0.5	0.4	1.2
1990 Masters	78.9	6.7	3.4	2.2	5.4	0.9	0.7	0.8	1.0
1990 Doctorate	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). MSA percentages are for 1985-1997. 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 recoding 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 groupings listed in Table 2; English provides a residual for Census statistics.

Table 2.2: Bilateral Trade Sample

	Exports ($B)		TFP Index		GDP/Capita ($)			Exports ($B)		TFP Index		GDP/Capita ($)	
Country	Mean Level	Growth Rate	Mean Level	Growth Rate	Mean Level	Growth Rate	Country	Mean Level	Growth Rate	Mean Level	Growth Rate	Mean Level	Growth Rate
Single Ethnic Mappings:							*Hispanic Economies:*						
India	14.4	9%	5%	3%	1364	7%	Argentina	11.4	7%	8%	-3%	7711	4%
Japan	193.3	7%	46%	1%	17120	6%	Belize	0.1	3%	na	na	4466	4%
Korea	46.9	13%	10%	7%	8110	11%	Bolivia	0.7	1%	0%	-4%	2134	2%
Soviet Union	47.6	5%	na	na	7770	-3%	Brazil	26.8	6%	14%	-2%	5327	4%
Vietnam	2.2	25%	na	na	1300	10%	Chile	7.1	7%	2%	3%	5413	7%
							Columbia	4.0	5%	3%	-1%	3974	5%
Chinese Economies:							Costa Rica	1.0	7%	1%	-2%	4106	3%
China, Mainland	58.1	13%	7%	6%	1653	10%	Cuba	1.4	1%	na	na	5278	0%
Hong Kong	64.5	15%	5%	5%	16962	8%	Dom. Republic	0.3	5%	1%	0%	2846	5%
Singapore	43.3	11%	4%	4%	14499	9%	Ecuador	1.6	4%	1%	-3%	3252	2%
Taiwan	46.5	12%	6%	6%	9042	10%	El Salvador	0.5	4%	1%	-2%	3000	4%
							Guatemala	0.9	2%	1%	-2%	3010	3%
European Economies:							Honduras	0.4	0%	0%	-2%	1747	3%
Austria	32.4	6%	8%	-1%	15614	5%	Mexico	10.5	8%	12%	-2%	6296	3%
Belgium	98.5	6%	10%	-1%	15536	5%	Nicaragua	0.3	2%	0%	-6%	2007	-1%
Denmark	28.7	7%	6%	-1%	17487	5%	Panama	0.3	3%	1%	-2%	4481	4%
Finland	22.2	6%	5%	-1%	14956	5%	Paraguay	0.7	5%	1%	-1%	3756	4%
France	167.9	6%	30%	-1%	15434	5%	Peru	2.7	5%	2%	-3%	3599	3%
Germany	312.2	6%	34%	0%	15521	5%	Philippines	6.2	8%	3%	-2%	2464	4%
Italy	134.9	7%	30%	-1%	14967	6%	Portugal	12.2	10%	4%	0%	9210	7%
Netherlands	114.4	6%	12%	-2%	15625	5%	Spain	48.1	10%	17%	-1%	11218	6%
Norway	28.2	6%	6%	0%	17829	5%	Uruguay	1.5	6%	1%	-2%	6418	5%
Poland	14.2	3%	6%	-2%	5575	4%	Venezuela	8.1	-3%	4%	-4%	5773	2%
Sweden	46.2	6%	8%	-2%	16224	5%							
Switzerland	48.4	6%	9%	-3%	19258	5%							

Table 2.3: Reduced-Form Regressions of Bilateral Exports on US Ethnic Human-Capital Stocks

Dependent Variable is Log Bilateral Export Volume (US$)	Base Regression	Excluding Gravity Covariates	Full Non-Zero Sample	Including Zero-Valued Trade Series	Excluding Non-Ethnic Importers	Including Importer HC Stock	RF for Single IV Regressions	RF for Dual IV Regressions
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Exporter's US Ethnic HC Stock	0.569	1.349	1.370	0.685	0.536	0.682	0.454	0.376
	(0.201)	(0.241)	(0.298)	(0.252)	(0.254)	(0.255)	(0.188)	(0.217)
Log Importer's US Ethnic HC Stock						0.401		0.093
						(0.411)		(0.447)
Log Product of GDPs per Capita	1.032			1.823	0.333	0.499	0.942	0.159
	(0.451)			(0.247)	(0.536)	(0.578)	(0.438)	(0.539)
Log Product of GDPs	0.056			-1.020	0.725	0.442	0.156	0.906
	(0.447)			(0.159)	(0.461)	(0.540)	(0.434)	(0.493)
Observations	66358	66358	84741	88701	28850	28850	62839	25384

	Base Regression							
	Excluding Chinese	Excluding European	Excluding Hispanic	Excluding Indian	Excluding Japanese	Excluding Korean	Excluding Russian	Excluding Vietnamese
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Log Exporter's US Ethnic HC Stock	0.667	0.622	0.661	0.509	0.707	0.564	0.556	0.468
	(0.364)	(0.291)	(0.241)	(0.208)	(0.191)	(0.224)	(0.200)	(0.189)
Log Product of GDPs per Capita	1.008	0.509	1.474	1.151	1.132	1.038	1.013	0.998
	(0.509)	(0.301)	(0.576)	(0.458)	(0.449)	(0.470)	(0.456)	(0.451)
Log Product of GDPs	0.112	0.536	-0.367	-0.064	-0.083	0.057	0.082	0.103
	(0.501)	(0.334)	(0.574)	(0.453)	(0.441)	(0.463)	(0.450)	(0.449)
Observations	59351	42010	38436	64445	64277	64426	65747	65814

Notes: Regressions include bilateral export and year effects. Standard errors are clustered at the ethnicity level.

Table 2.4: Reduced-Form Regressions of Bilateral Exports on US Immigration Quotas Estimators

Dependent Variable is Log Bilateral Export Volume (US$)	Base Regression	Excluding Gravity Covariates	Full Non-Zero Sample	Including Zero-Valued Trade Series	Excluding Non-Ethnic Importers	Including Importer Estimator	RF for Single IV Regressions	RF for Dual IV Regressions
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Exporter's US Immigration Estimator	0.520 (0.200)	0.985 (0.296)	0.908 (0.305)	0.624 (0.042)	0.538 (0.157)	0.515 (0.205)	0.530 (0.206)	0.600 (0.206)
Log Importer's US Immigration Estimator						-0.148 (0.295)		-0.404 (0.237)
Log Product of GDPs per Capita	0.899 (0.351)			1.657 (0.155)	0.196 (0.458)	0.899 (0.352)	0.856 (0.354)	0.035 (0.518)
Log Product of GDPs	0.278 (0.315)			-0.766 (0.134)	0.932 (0.388)	0.284 (0.325)	0.300 (0.321)	1.118 (0.483)
Observations	66358	66358	84741	88701	28850	66358	62839	25384

	Base Regression							
	Excluding Chinese	Excluding European	Excluding Hispanic	Excluding Indian	Excluding Japanese	Excluding Korean	Excluding Russian	Excluding Vietnamese
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Log Exporter's US Immigration Estimator	0.663 (0.385)	0.452 (0.170)	0.646 (0.283)	0.352 (0.098)	0.497 (0.197)	0.533 (0.206)	0.521 (0.201)	0.542 (0.207)
Log Product of GDPs per Capita	0.909 (0.421)	0.547 (0.261)	1.237 (0.599)	1.018 (0.350)	0.930 (0.331)	0.904 (0.370)	0.883 (0.359)	0.901 (0.359)
Log Product of GDPs	0.279 (0.394)	0.592 (0.286)	0.022 (0.579)	0.161 (0.302)	0.246 (0.290)	0.259 (0.333)	0.301 (0.324)	0.258 (0.323)
Observations	59351	42010	38436	64445	64277	64426	65747	65814

Notes: Regressions include bilateral export and year effects. Standard errors are clustered at the ethnicity level.

Table 2.5: OLS Regressions of Bilateral Exports on TFP Indices

Dependent Variable is Log Bilateral Export Volume (US$)	Base Regression	Excluding Gravity Covariates	Full Non-Zero Sample	Including Zero-Valued Trade Series	Excluding Non-Ethnic Importers	Including Importer TFP Index	1980-1992 Sample	1985-1997 Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Exporter's TFP Index	0.164 (0.149)	0.938 (0.127)	0.887 (0.120)	0.289 (0.199)	0.411 (0.194)	0.319 (0.118)	0.012 (0.149)	0.343 (0.299)
Log Importer's TFP Index						-0.136 (0.164)		
Log Product of GDPs per Capita	0.769 (0.383)			1.492 (0.200)	-0.071 (0.599)	-0.046 (0.496)	0.957 (0.257)	0.340 (0.667)
Log Product of GDPs	0.357 (0.336)			-0.664 (0.124)	0.994 (0.403)	1.112 (0.497)	0.174 (0.251)	0.673 (0.627)
Observations	62839	62839	78794	82047	27182	25384	43351	46893

	Base Regression							
	Excluding Chinese	Excluding European	Excluding Hispanic	Excluding Indian	Excluding Japanese	Excluding Korean	Excluding Russian	Excluding Vietnamese
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Log Exporter's TFP Index	0.041 (0.193)	0.220 (0.147)	0.406 (0.083)	0.113 (0.175)	0.200 (0.140)	0.139 (0.187)	n.a.	n.a.
Log Product of GDPs per Capita	0.764 (0.456)	0.382 (0.343)	1.233 (0.689)	0.930 (0.364)	0.801 (0.370)	0.776 (0.399)		
Log Product of GDPs	0.404 (0.412)	0.649 (0.370)	-0.104 (0.668)	0.201 (0.316)	0.300 (0.309)	0.364 (0.351)		
Observations	55927	39607	36070	60926	60758	60907		

Notes: Regressions include bilateral export and year effects. Standard errors are clustered at the ethnicity level.

Table 2.6A: IV Regressions of Bilateral Exports on TFP Indices using US Ethnic Human-Capital Stocks Instruments

Dependent Variable is Log Bilateral Export Volume (US$)	OLS				US Ethnic Human-Capital Stocks IV			
	Base Regression	Including Gravity Covariates	Including Importer TFP Index	Including Gravity Covariates	Base Regression	Including Gravity Covariates	Including Importer TFP Index	Including Gravity Covariates
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. OLS and Second-Stage Regressions							
Log Exporter's TFP Index	0.887 (0.120)	0.164 (0.149)	1.047 (0.075)	0.319 (0.118)	0.946 (0.156)	0.623 (0.222)	1.118 (0.154)	0.705 (0.367)
Log Importer's TFP Index			0.591 (0.078)	-0.136 (0.164)			0.842 (0.429)	0.428 (0.826)
Log Product of GDPs per Capita		0.769 (0.383)		-0.046 (0.496)		1.023 (0.441)		-0.281 (0.602)
Log Product of GDPs		0.357 (0.336)		1.112 (0.497)		-0.302 (0.440)		0.731 (0.394)
	B.First-Stage Regressions for Exporter's TFP Index							
Log Exporter's US Ethnic HC Stock					1.096 (0.308)	1.009 (0.238)	1.120 (0.394)	0.764 (0.230)
Log Importer's US Ethnic HC Stock							0.021 (0.016)	-0.336 (0.159)
	C. First-Stage Regressions for Importer's TFP Index							
Log Exporter's US Ethnic HC Stock							0.007 (0.012)	-0.368 (0.151)
Log Importer's US Ethnic HC Stock							1.102 (0.386)	0.726 (0.237)
Observations	78794	62839	25384	25384	78794	62839	25384	25384

Notes: Regressions include bilateral export and year effects. Gravity covariates are included in first-stage regressions. Standard errors are clustered at the ethnicity level.

Table 2.6B: IV Regressions of Bilateral Exports on TFP Indices using US Immigration Quotas Instruments

Dependent Variable is Log Bilateral Export Volume (US$)	OLS				US Immigration Quotas IV			
	Base Regression	Including Gravity Covariates	Including Importer TFP Index	Including Gravity Covariates	Base Regression	Including Gravity Covariates	Including Importer TFP Index	Including Gravity Covariates
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. OLS and Second-Stage Regressions							
Log Exporter's TFP Index	0.887	0.164	1.047	0.319	1.342	1.130	1.685	1.227
	(0.120)	(0.149)	(0.075)	(0.118)	(0.455)	(0.610)	(0.685)	(1.418)
Log Importer's TFP Index			0.591	-0.136			0.050	-0.398
			(0.078)	(0.164)			(0.304)	(0.983)
Log Product of GDPs per Capita		0.769		-0.046		0.820		-0.126
		(0.383)		(0.496)		(0.693)		(1.088)
Log Product of GDPs		0.357		1.112		-0.293		0.753
		(0.336)		(0.497)		(0.546)		(0.688)
	B.First-Stage Regressions for Exporter's TFP Index							
Log Exporter's US Immigration Estimator					0.615	0.499	0.644	0.394
					(0.255)	(0.176)	(0.250)	(0.137)
Log Importer's US Immigration Estimator							0.008	-0.248
							(0.018)	(0.104)
	C. First-Stage Regressions for Importer's TFP Index							
Log Exporter's US Immigration Estimator							-0.002	-0.249
							(0.017)	(0.103)
Log Importer's US Immigration Estimator							0.646	0.392
							(0.253)	(0.139)
Observations	78794	62839	25384	25384	78794	62839	25384	25384

Notes: Regressions include bilateral export and year effects. Gravity covariates are included in first-stage regressions. Standard errors are clustered at the ethnicity level.

Table 2.7: OLS Regressions of Industry-Level Multilateral Trade on US Ethnic Human-Capital Stocks

	Levels Regressions				First-Differences Regressions			
	Multilateral Exports		Multilateral Imports		Multilateral Exports		Multilateral Imports	
Dependent Variable is Log Industry Trade Volume (US$)	No Weights	Industry Patent Weights	No Weights	Industry Patent Weights	No Weights	Industry Patent Weights	No Weights	Industry Patent Weights
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Base Industry-Level Regressions							
Log Exporter's US Ethnic HC Stock	0.719	0.809	0.206	0.545	0.214	0.325	-0.014	0.086
	(0.115)	(0.131)	(0.111)	(0.107)	(0.098)	(0.101)	(0.084)	(0.086)
Observations	23948	23948	23948	23948	22437	22437	22437	22437
	B. Coefficients for Exporter's US Ethnic HC Stock from Variations on Base Industry-Level Regressions							
Excluding Computer and Drug Industries	0.688	0.766	0.172	0.494	0.202	0.303	-0.019	0.075
	(0.118)	(0.133)	(0.112)	(0.103)	(0.100)	(0.108)	(0.086)	(0.090)
Including Ethnicity Time Trends	0.699	0.967	0.064	0.278	0.141	0.324	0.092	0.182
	(0.147)	(0.178)	(0.122)	(0.147)	(0.117)	(0.133)	(0.095)	(0.099)
Including Country Time Trends	0.553	0.882	-0.065	0.195	0.127	0.309	0.087	0.177
	(0.141)	(0.179)	(0.122)	(0.142)	(0.118)	(0.134)	(0.095)	(0.099)
Including Ethnic-Year Effects	0.602	0.954	-0.111	0.137	0.043	0.226	-0.018	0.032
	(0.153)	(0.200)	(0.136)	(0.156)	(0.117)	(0.129)	(0.098)	(0.086)
Including Country-Year Effects	0.435	0.847	-0.258	0.067	-0.015	0.212	-0.070	-0.017
	(0.147)	(0.209)	(0.141)	(0.158)	(0.120)	(0.134)	(0.095)	(0.089)

Notes: Regressions include country-industry effects, industry-year effects, and gravity covariates. Standard errors are clustered at the ethnicity-industry level.

Table 2.8: Industry-Share Regressions to Test Rybczynski Effect

Dependent Variable is Industry Export Share of Country Exports	No Weights	Including Interaction of Ethnic Time Trends With Industry Quintiles			Industry Patent Weights	Including Interaction of Ethnic Time Trends With Industry Quintiles		
		Wage	Capital-Labor	Non-Product.		Wage	Capital-Labor	Non-Product.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of Exporter's US Patenting in Industry	0.694 (0.382)	0.738 (0.379)	0.689 (0.361)	0.699 (0.360)	0.801 (0.412)	0.806 (0.348)	0.732 (0.332)	0.764 (0.349)
Observations	56861	56861	56861	56861	56861	56861	56861	56861

Notes: Regressions include country-industry and industry-year effects. Standard errors are clustered at the ethnicity-industry level.

Table 2.9: Geographic Distance Regressions

Dependent Variable is Log Bilateral Export Volume (US$)	Observations with Distance Data	Exports to Border Countries	Exports to Countries ≤ 1500 km	Exports to Countries 1501-3000	Exports to Countries 3001-6000	Exports to Countries 6001-9000	Exports to Countries > 9000 km	Exports to Non-Border Countries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Exporter's US Ethnic HC Stock	0.586 (0.208)	1.772 (0.801)	0.977 (0.475)	0.564 (0.614)	0.895 (0.396)	0.614 (0.301)	0.506 (0.305)	0.570 (0.206)
Log Product of GDPs per Capita	1.055 (0.459)	1.011 (0.327)	1.461 (0.238)	0.380 (0.655)	-0.171 (1.119)	1.760 (0.423)	0.859 (0.279)	1.083 (0.469)
Log Product of GDPs	0.022 (0.451)	-0.503 (0.416)	-0.907 (0.317)	0.322 (0.714)	0.985 (1.141)	-0.580 (0.405)	0.419 (0.317)	0.012 (0.468)
Observations	63826	1933	3923	5575	10785	15110	26500	61893

Notes: Regressions include bilateral export and year effects. Standard errors are clustered at the ethnicity level.

Table 2.A1: ISIC Revision 2 Industry Codes

ISIC	Industry Title	US Quintiles (5 = Highest) K/L	Wage	Skill
3111	Slaughtering, preparing and preserving meat	2	1	1
3112	Man. of dairy products	4	3	4
3113	Canning and preserving of fruits and vegetables	4	2	1
3114	Canning, preserving and processing of fish and crustaceans		n.a.	
3115	Man. of vegetable and animal oils and fats	5	3	4
3116	Grain mill products	5	4	4
3117	Man. of bakery products	3	2	5
3118	Sugar factories and refineries	4	2	2
3119	Man. of cocoa, chocolate and sugar confectionery		n.a.	
3121	Man. of food products n.e.c.	3	2	3
3122	Man. of prepared animal feeds		n.a.	
3131	Distilling, rectifying and blending spirits	5	4	5
3132	Wine industries		n.a.	
3133	Malt liquors and malt		n.a.	
3134	Soft drinks and carbonated waters industries		n.a.	
3140	Tobacco manufactures	5	5	3
3211	Spinning, weaving and finishing textiles	3	1	1
3212	Man. of made-up textile goods except wearing apparel		n.a.	
3213	Knitting mills	1	1	1
3214	Man. of carpets and rugs	2	2	2
3215	Cordage, rope and twine industries		n.a.	
3219	Man. of textiles n.e.c.	1	1	2
3220	Man. of wearing apparel, except footwear	1	1	1
3231	Tanneries and leather finishing	2	2	1
3232	Fur dressing and dyeing industries	1	1	4
3233	Man. of products of leather, except footwear and wearing apparel	1	1	2
3240	Man. of footwear, except vulcanized or moulded rubber or plastic	1	1	1
3311	Sawmills, planing and other wood mills	2	1	1
3312	Man. of wooden and cane containers and small cane ware	1	1	1
3319	Man. of wood and cork products n.e.c.	1	1	1
3320	Man. of furniture and fixtures, except primarily of metal	1	1	1
3411	Man. of pulp, paper and paperboard	5	5	2
3412	Man. of containers and boxes of paper and paperboard	3	3	2
3419	Man. of pulp, paper and paperboard articles n.e.c.	3	3	2
3420	Printing, publishing and allied industries	1	3	5
3511	Man. of basic industrial chemicals except fertilizers	5	5	5
3512	Man. of fertilizers and pesticides	5	5	4
3513	Man. of synthetic resins, plastic and man-made fibres except glass	5	5	4
3521	Man. of paints, varnishes and lacquers	4	4	5
3522	Man. of drugs and medicines	5	5	5

Table 2.A1: ISIC Revision 2 Industry Codes (continued)

ISIC	Industry Title	US Quintiles (5 = Highest) K/L	Wage	Skill
3523	Man. of soap and cleaning, preparations, perfumes, cosmetics, etc.	4	4	5
3529	Man. of chemical products n.e.c.	4	4	5
3530	Petroleum refineries	5	5	4
3540	Man. of miscellaneous products of petroleum and coal	5	4	4
3551	Tyre and tube industries	5	5	1
3559	Man. of rubber products n.e.c.	2	2	3
3560	Man. of plastic products n.e.c.	2	2	2
3610	Man. of pottery, china and earthenware	1	2	2
3620	Man. of glass and glass products	4	3	1
3691	Man. of structural clay products	3	2	2
3692	Man. of cement, lime and plaster	4	3	3
3699	Man. of non-metallic mineral products n.e.c.	4	3	3
3710	Iron and steel basic industries	5	5	2
3720	Non-ferrous metal basic industries	5	4	3
3811	Man. of cutlery, hand tools and general hardware	3	3	3
3812	Man. of furniture and fixtures primarily of metal		n.a.	
3813	Man. of structural metal products	2	3	3
3819	Man. of fabricated metal products except mach. and equip. n.e.c.	3	3	3
3821	Man. of engines and turbines	5	5	4
3822	Man. of agricultural mach. and equip.	4	3	3
3823	Man. of metal and wood-working mach.	3	4	3
3824	Man. of special ind. mach./equip. except metal and wood-working	2	4	5
3825	Man. of office, computing and accounting mach.	4	5	5
3829	Mach. and equip. except electrical n.e.c.	3	4	4
3831	Man. of electrical industrial mach. and apparatus	2	3	4
3832	Man. of radio, television and communication equip. and apparatus	3	5	5
3833	Man. of electrical appliances and household goods	3	2	3
3839	Man. of electrical apparatus and supplies n.e.c.	4	4	4
3841	Shipbuilding and repairing	2	3	2
3842	Man. of railroad equip.	3	4	3
3843	Man. of motor vehicles	4	5	1
3844	Man. of motorcycles and bicycles	2	3	2
3845	Man. of aircraft	3	5	5
3849	Man. of transport equip. n.e.c	3	5	5
3851	Man. of prof. and scientific, measuring/controlling equip., n.e.c	2	4	5
3852	Man. of photographic and optical goods	4	4	5
3853	Man. of watches and clocks	2	2	3
3901	Man. of jewellery and related articles	1	2	4
3902	Man. of musical instruments	1	1	2
3903	Man. of sporting and athletic goods	2	1	3
3909	Manufacturing industries n.e.c.		n.a.	

Table 2.A2: IV Regressions of Bilateral Exports on TFP Indices - Sample Decomposition

Dependent Variable is Log Bilateral Export Volume (US$)	Base Regression							
	Excluding Chinese	Excluding European	Excluding Hispanic	Excluding Indian	Excluding Japanese	Excluding Korean	1980-1992 Sample	1985-1997 Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Second-Stage Coefficients using US Ethnic Human-Capital Stocks Instruments							
Log Exporter's TFP Index	0.871	0.444	0.708	0.530	0.728	0.630	0.372	0.893
	(0.634)	(0.139)	(0.159)	(0.200)	(0.259)	(0.281)	(0.203)	(0.239)
	B. First-Stage Coefficients using US Ethnic Human-Capital Stocks Instruments							
Log Exporter's US Ethnic HC Stock	0.700	1.597	1.087	1.006	1.026	0.938	0.935	1.004
	(0.380)	(0.036)	(0.080)	(0.260)	(0.242)	(0.304)	(0.229)	(0.257)
Observations	55927	39607	36070	60926	60758	60907	43351	46893
	C. Second-Stage Coefficients using US Immigration Quotas Instruments							
Log Exporter's TFP Index	3.635	1.150	1.227	0.717	1.074	1.064	n.a.	1.408
	(2.007)	(0.671)	(0.766)	(0.246)	(0.601)	(0.544)		(0.624)
	D. First-Stage Coefficients using US Immigration Quotas Instruments							
Log Exporter's US Immigration Estimator	0.209	0.420	0.597	0.513	0.502	0.549	n.a.	0.457
	(0.221)	(0.110)	(0.215)	(0.246)	(0.178)	(0.185)		(0.168)
Observations	55927	39607	36070	60926	60758	60907		46893

Notes: Regressions include bilateral export effects, year effects, and gravity covariates. Standard errors are clustered at the ethnicity level.

Chapter 3

The Ethnic Composition of US Inventors

Summary 3 *The ethnic composition of US scientists and engineers is undergoing a signficant transformation. This study applies an ethnic-name database to individual patent records granted by the United States Patent and Trademark Office to document these trends with greater detail than previously available. Most notably, the contribution of Chinese and Indian scientists to US technology formation increased dramatically in the 1990s. Growth in ethnic innovation is concentrated in high-tech sectors; the institutional and geographic dimensions are further characterized.*

3.1 Introduction

The contributions of immigrants to US technology formation are staggering: while foreign-born account for just over 10% of the US working population, they represent 25% of the US science and engineering (SE) workforce and nearly 50% of those with doctorates. Even looking within the Ph.D. level, ethnic 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.[1] Moreover, ethnic entrepreneurs are very active in commercializing new technologies, especially

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

in high-tech sectors (e.g., Saxenian 2002a). The magnitude of these ethnic contributions raises many research and policy questions: debates regarding the appropriate quota for H1-B temporary visas, the possible crowding out of native students from SE fields, and the brain drain or brain circulation effect on sending countries are just three examples.[2]

Econometric studies quantifying the role of ethnic scientists and engineers for technology formation and diffusion are often hampered, however, by data constraints. It is very difficult to assemble sufficient cross-sectional and longitudinal variation for large-scale panel exercises.[3] This paper describes a new approach for quantifying the ethnic composition of US inventors with previously unavailable detail. The technique exploits the rarely used inventor names contained in the NBER Patent Data File, originally compiled by Hall, Jaffe, and Trajtenberg (2001). This dataset provides micro-records for all patents granted by the USPTO from January 1975 to December 1999. Each patent record lists one or more inventors, with 4.3 million inventor names associated with the 2.9 million patents. The USPTO grants patents to inventors living within and outside of the US, with the latter accounting for just under half of the total.

This study maps into these inventor names an ethnic-name database typically used for commercial applications.[4] This approach exploits the idea that inventors with the surnames Chang or Wang are likely of Chinese ethnicity, those with surnames Rodriguez or Martinez of Hispanic ethnicity, and so on. The match rates range from 93%-99% for US domestic inventor records, depending upon the procedure employed, and the process affords the distinction of nine ethnicities: Chinese, English, European, Hispanic/Filipino, Indian/Hindi, Japanese, Korean, Russian, and Vietnamese. Moreover, because the matching is done at the micro-level, greater ethnic composition detail is available annually on multiple dimensions: detailed technology categories, MSAs, companies, and so on.[5]

The next section details the ethnic-name matching strategy, outlines the strengths and

[2] Representative papers are Lowell (2000), Borjas (2005), and Saxenian (2002b) respectively.

[3] While the decennial Census provides detailed cross-sectional descriptions, its longitudinal variation is necessarily limited. On the other hand, the annual Current Population Survey provides poor cross-sectional detail and does not ask immigrant status until 1994. The SESTAT database offers a better trade-off between the two dimensions, but suffers important sampling biases with respect to immigrants (Kannankutty and Wilkinson 1999).

[4] The database is constructed by the Melissa Data Corporation for the design of direct-mail advertisements. I am grateful to the George Schultz Fund for financial assistance in its purchase.

[5] This ethnic patenting database is employed by Kerr (2005) and associated papers to study the role of ethnic scientists and entrepreneurs in technology formation and diffusion.

weaknesses of the database selected, and offers some validation exercises using patent records filed by foreign inventors with the USPTO. Section 3.3 then documents the growing contribution of ethnic inventors to US technology formation. The rapid increase during the 1990s in the percentage of high-tech patents going to Chinese and Indian inventors is particularly striking. The relative contributions from scientists of European ethnicity, however, decline somewhat from their levels in 1980. The institutional and geographic dimensions of ethnic innovation are further delineated. Section 3.4 concludes.

3.2 Ethnic-Name Matching Strategy

This section describes the ethnic-name matching strategy employed with the inventor names contained in the NBER Patent Data File. To begin, two common liabilities associated with using ethnic-name databases are identified. Addressing these limitations guides the selection of the Melissa database and the design of the matching strategy, which is described in detail. The section concludes with descriptive statistics from a quality-assurance exercise of applying the ethnic-name strategy to inventors residing outside of the US who file patent applications with the USPTO.

Ethnic-name databases suffer from two inherent limitations — not all ethnicities are covered, and included ethnicities usually receive unequal treatment. The Melissa database's strength is the identification of Asian ethnicities, especially Chinese, Indian/Hindi, Japanese, Korean, Russian, and Vietnamese names. The database is comparatively weaker for looking within continental Europe. For example, Dutch surnames are collected without first names, while the opposite is true for French names. The Asian comparative advantage and overall cost effectiveness led to the selection of the Melissa database, as well as the European amalgamation employed in this study. In total, nine ethnicities are distinguished: Chinese, English, European, Hispanic/Filipino, Indian/Hindi, Japanese, Korean, Russian, and Vietnamese.[6,7] The largest

[6]The Melissa database provides lists for the Hispanic/Latino and Filipino/Tagalog ethnicities. However, extensive overlap exists between these two groupings (e.g., the common surnames Martinez and Ramirez are in both ethnic lists). The final matching procedure combines these groups into a joint Hispanic/Filipino ethnicity, while in earlier work they are kept separate (Kerr 2004). This choice is not a first-order concern.

[7]The ethnic groups employed: Chinese, English, European (including Dutch, French, German, Italian, and Polish names), Hispanic/Filipino (including Latino and Filipino/Tagalog names), Indian/Hindi (including Bangladeshi and Pakistani names), Japanese, Korean, Russian (including Armenian and Carpatho-Rusyns

ethnicity in the US SE workforce absent from the ethnic-name database is Iranian, which accounted for 0.7% of bachelor-level SEs in the 1990 Census.

The second limitation is that commercial databases vary in the number of names they contain for each ethnicity. These differences reflect both uneven coverage and that some ethnicities are more homogeneous in their naming conventions.[8] Two polar matching strategies are employed to ensure coverage differences do not overly influence ethnicity assignments.

> *Full Matching:* This procedure utilizes all of the name assignments in the Melissa database and manually codes any unmatched surname or first name associated with 100 or more inventor records. This technique further exploits the international distribution of inventor names within the patent database to provide superior results.[9] The match rate for this restricted procedure is 98% (98% US, 97% foreign).[10]
>
> *Restricted Matching:* A second strategy employs a uniform name database using only the 3000 and 200 most common surnames and first names, respectively, for each ethnicity. These numerical bars are the lowest common denominators across the major ethnicities studied. The match rate for this restricted procedure is 89% (93% US, 85% foreign).

For matching, names in both the patent and ethnic-name databases are capitalized and truncated to ten characters. Approximately 89% of the patent name records have a unique surname, first name, or middle name match in the Full Matching procedure (77% in the Restricted Matching), affording a single ethnicity determination with priority given to surname matches. For inventors residing in the US, representative probabilities are assigned to non-unique matches using the masters-level SE communities by MSA. Ethnic probabilities for the

---

names), and Vietnamese. Jewish ethnic names overlapped extensively with other ethnic groupings and are excluded. The Bangladeshi and Pakistani name counts are extremely small (8 and 15 respectively) and do not influence the Indian/Hindi outcome. A handful of names classified as Arab, Burmese, and Malay are discarded.

[8]For example, the Herfindahl indices for Korean (470) and Vietnamese (1121) surnames are significantly higher than Japanese (132) and English (164) due to frequent Korean surnames like Kim (16%) and Park (12%) and Vietnamese surnames like Nguyen (29%) and Tran (12%).

[9]A simple rule is applied to take advantage of the information embedded in the patent database itself. If over 90% of USPTO records associated with a name are concentrated in a non-English ethnicity country or region, the name is assigned that ethnicity. As the test includes the domestic US inventors, comprising over 50% of all inventors, this technique is very stringent and mainly bolsters European ethnic matching (the comparative weakness of the Melissa database). The rule is not applied to names with fewer than ten occurrences.

[10]The matching rate should be less than 100% with the Melissa database as not all ethnicities are included.

remaining 3% of records (mostly foreign) are calculated as equal shares.[11]

The application of the ethnic-name database to the inventors residing abroad provides a natural quality-assurance exercise for the technique. The top panel of Table 3.1 summarizes the results, with the rows presenting the matched characteristics for countries and regions grouped to the ethnicities identifiable with the database. The results are very encouraging. First, the Full Matching procedure assigns ethnicities to a large percentage of foreign records, with the match rates greater than 94% for all ethnicities but Indian (82%). In the Restricted Matching procedure, a success rate of greater than 74% holds for all ethnicities.

Second, the estimated inventor compositions are reasonable. The own-ethnicity shares are summarized in the fourth and fifth columns. The weighted average is 88% in the Full Matching procedure, and own-ethnicity contributions are greater than 80% in the UK, China, India, Japan, Korea, and Russia regardless of the matching procedure employed. Like the US, own-ethnicity contributions should be less than 100% due to foreign researchers. The high success rate using the Restricted Matching procedure indicates that the ethnic-name database performs well without exploiting the international distribution of names, although power is lost with Europe. Likewise, uneven coverage in the Melissa database is not driving the ethnic composition trends.[12]

The bottom panel of Table 3.1 presents the complete ethnic compositions estimated for the foreign countries. Many of the positive off-diagonals are to be expected, either due to foreign expatriates (UK), small sample sizes (Vietnam), or overlaps of common names.[13] One advantage the matching technique possesses for inventors residing in the US is the ability to use the Census to assign probabilistic estimates for overlapping names; foreign records are only

[11] MSA ethnic compositions are averages of the 1980 and 1990 5% Census files. The sample considers civilians aged 22-54 listing Engineers, Mathematical and Computer Scientists, or Natural Scientists as their occupations. The master's degree cut-off reflects the higher average education level of patenting scientists within the scientific community (Kannankutty and Wilkinson 1999). Country-of-birth is used to assign ethnicities into broad categories that match the name records. To illustrate, take the San Francisco scientific community to be 12.1% Chinese, 66.1% English, and 4.6% European (with other ethnicities omitted). A San Francisco-based record matching to Chinese, English, and European surnames would be assigned a probabilistic ethnicity of 14.6% Chinese, 79.8% English, and 5.6% European (summing to 100%). A China-based record matching all three ethnicities would be assigned a 33.3% probability for each.

[12] The main US SE ethnicity missing from the database is Iranian. Running the ethnic-name database on the few patents from Iran yields a 42%-65% match rate. Iran's predicted composition does not favor any of the nine ethnicities studied, with the largest overlap being the Russian ethnicity at 31%.

[13] Two prominent examples are the surname Lee (Chinese, English, and Korean) and the first name Igor (Hispanic and Russian). The most overlap occurs between the European and Hispanic ethnicities.

assigned as equal shares. The last two columns of Table 3.1's top panel indicate the percentage of the foreign inventors assigned at least partially to their own-ethnicity. While this study does not make the strong assumption that ties should go to the country's own-ethnicity, the additional power provided by using the US Census for breaking domestic ties is illustrated.

Finally, visual confirmation of the top 1000 surnames and first names in the USPTO records confirms the matching technique works well.[14] While some inventors are certainly misclassified, the measurement error in aggregate trends building from the micro-data is minor. The Full Matching procedure is the preferred technique, and underlies the trends presented in the next section, but most applications find little to no difference when the Restricted Matching dataset is employed instead.

3.3 Ethnic Composition of US Inventors

Table 3.2 describes the ethnic composition of US inventors for 1975-1997.[15] The trends demonstrate a growing ethnic contribution to US technology development, especially among Chinese and Indian scientists. Also matching popular perceptions, 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 SE workforce in the 1990 Census.[16]

Figure 3.1 illustrates the evolving ethnic composition of US inventors from 1980-1997. Looking across all technology categories, the European ethnicity is the largest foreign contributor

[14]Appendix Table 3.A1 lists the fifty most common surnames for each ethnicity, along with their relative contribution, for US inventors. The counts sum the ethnic contribution from inventors with each surname, including partial or split assignments, and are not necessarily direct or exclusive matches (e.g., the ethnic match may have occurred through the first name).

[15]The NBER database contains records for patents granted by the USPTO through December 1999. The application years of patents, however, provide the best description of when innovative research is being undertaken, due to the substantial and uneven lags in the USPTO reviews. Accordingly, the annual descriptions employed in this study are undertaken by application years. Unfortunately, this approach does lead to significant attrition in the last two years of the 1990s — patents are only included in the database if they have been granted, but a smaller number of applications close to the December 1999 cut-off have completed the review cycle.

[16]The name-matching procedure has limited power for distinguishing first-generation versus later-generation immigrants. This is most relevant for the European ethnicity, and the estimated European contribution is naturally higher than the immigrant contribution measured by foreign born.

to US technology development. Like the English ethnicity, however, the European share of US domestic inventors declines steadily from 8% in 1981 to 7% in 1997. This declining share is partly due to the exceptional growth of the Chinese and Indian ethnicities, which increase from 3% to 6% and 2% to 4%, respectively, over the seventeen years. As shown below, this Chinese and Indian growth is concentrated in high-tech sectors, where Chinese inventors supplant European researchers as the largest ethnic contributor to US technology formation.

Among the other ethnicities, the Hispanic contribution grows from 3% in 1980 to 4% in 1997. The level of this series is likely mismeasured due to the extensive overlap of Hispanic and European names, but the positive growth is consistent with stronger Latino and Filipino scientific contributions in Florida and California. While small, the Korean and Russian shares almost double from 1980 to 1997 (0.5% to 0.9% and 1.2% to 2.0%, respectively). Although difficult to see with Figure 3.1's scaling, much of the Russian increase occurs in the 1990s following the dissolution of the Soviet Union. The Japanese share also increases from 0.6% to 0.9%. Finally, while the Vietnamese contribution is the lowest throughout the sample, it does exhibit the strongest relative growth from 0.1% to 0.5%.

3.3.1 Contributions by Technology

Figure 3.2 documents the total ethnic contribution by the six broad technology groups into which patents are often classified: Chemicals, Computers and Communications, Drugs and Medical, Electrical and Electronic, Mechanical, and Others. The miscellaneous group includes patents for agriculture, textiles, furniture, and the like. Growth in ethnic patenting is clearly stronger in high-tech sectors than in more traditional industries. Figures 3.3 through 3.8 provide the ethnic contributions within each technology category. The growing ethnic contribution in high-tech sectors is easily traced to the Chinese and Indian ethnicities. Moreover, these two ethnicities exhibit the most interesting and economically meaningful variation across technologies, as summarized in Figures 3.9 and 3.10.

3.3.2 Contributions by Institution

Figure 3.11 demonstrates that intriguing differences in ethnic scientific contributions also exist by institution type. Over the 1980-1997 period, ethnic inventors are more concentrated in government and university research labs and in publicly listed companies than in private companies or as unaffiliated inventors.[17] Part of this levels difference is certainly due to visa sponsorships by larger institutions. Growth in ethnic shares are initially stronger in the government and university labs, but publicly listed companies close the gap by 1997. The other interesting trend in Figure 3.11 is for private companies, where the ethnic contribution sharply increases in the 1990s. This rise coincides with the strong growth in ethnic entrepreneurship in high-tech sectors. For example, the Chinese share of computer patenting in private firms grows from 1.5% in 1980 to 3.7% in 1990, before exploding to 9.0% in 1997.

3.3.3 Contributions by Geography

This paper closes its descriptive statistics with an examination of the 1985-1997 ethnic inventor contributions by major MSAs.[18] The first three columns of Table 3.3 document each city's share of US ethnic patenting. Not surprisingly, these shares are highly correlated with city size, with the three largest ethnic centers for 1990 found in New York (13%), San Francisco (8%), and Los Angeles (7%). Comparing these ethnic patenting percentages with the total patenting shares, listed in the second set of three columns, reveals the more interesting fact that ethnic patenting concentration is higher than that of general innovation. The 1990 patent shares of New York, San Francisco, and Los Angeles are lower at 9%, 6%, and 6%, respectively. Similarly, 75% of ethnic research occurs in the major MSAs listed in Table 3.3, compared to 68% of total patenting. The final three columns exhibit the raw patent counts by city.[19]

Not only are ethnic scientists disproportionately concentrated in major cities, but growth

[17] Industry patents account for 72% of patents granted from 1980-1997. Public companies account for 59% of industry patents during the period and are identified through Compustat records. Government and university institutions are identified through institution names and account for about 4% of patents granted. Federally funded research and development centers (FFRDCs) are included in both industry and government groups. Unaffiliated applications account for about 26% of patents granted.

[18] 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.

[19] Raw patent counts should be treated with caution. Changes in the personnel resources and review policies of the USPTO influence the number of patents granted over time (Griliches 1990), and the explosive climb in patent grants over the last two decades is difficult to interpret (e.g., Kortum and Lerner 2000, Kim and Marshcke 2004, and Hall 2004). Accordingly, this study considers patent shares, which avoids these interpretation concerns.

Studies seeking to quantify the number of ethnic researchers in the US should supplement this data with immigration records or demographic surveys (with an unfortunate loss of detail). Trajtenberg (2005) also employs the USPTO inventor names to identify individual scientists.

in a city's share of ethnic patenting is highly correlated with growth in its share of total US patenting. San Francisco is the most prominent example. Its share of ethnic invention rose from 8% in 1980 to 17% in 1997, while its patenting share also increased from 6% to 10%. Across the whole sample and including all of the intervening years, an increase of 1% in an MSA's ethnic patenting share correlates with a 0.6% increase in the MSA's total invention share. This coefficient is remarkably high, as the ethnic share of total invention during this period is around 20%. Shifts in the concentration of ethnic inventors appear to facilitate changes in the geographic composition of US innovation.[20]

3.4 Conclusion

Ethnic scientists and engineers are an important and growing contributor to US technology development. The Chinese and Indian ethnicities, in particular, are now an integral part of US invention in high-tech sectors. This paper describes how the probable ethnicities of US researchers can be determined at the micro-level through their names available with USPTO patent records. The ethnic-name database this study employs distinguishes nine ethnic groups, and the matched database describes the ethnic composition of US inventors with previously unavailable cross-sectional and longitudinal detail. This richer variation can support more detailed and informative empirical analyses than feasible otherwise.

[20] The ethnic-name approach does not distinguish ethnic inventor shifts due to new immigration, domestic migration, or occupational changes. It is likewise beyond the scope of this descriptive note to explore issues of causality or effects on native workers.

Bibliography

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

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

[3] Hall, Bronwyn, "Exploring the Patent Explosion", NBER Working Paper 10605 (2004).

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

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

[6] Kannankutty, Nirmala, and R. Keith Wilkinson, "SESTAT: A Tool for Studying Scientists and Engineers in the United States", National Science Foundation SRS Special Report (1999).

[7] Kerr, William, "Ethnic Scientific Communities and International Technology Diffusion", Job Market Paper (2004).

[8] Kerr, William, "The Role of Immigrant Scientists and Entrepreneurs in International Technology Transfer", MIT Ph.D. Dissertation (2005).

[9] Kim, Jinyoung, and Gerald Marschke, "Accounting for the Recent Surge in U.S. Patenting: Changes in R&D Expenditures, Patent Yields, and the High Tech Sector", *Economics of Innovation and New Technologies*, 13:6 (2004), 543-558.

[10] Kortum, Samuel, and Joshua Lerner, "Assessing the Contribution of Venture Capital to Innovation", *RAND Journal of Economics*, 31:4 (2000), 674-692.

[11] Lowell, B. Lindsay, "H1-B Temporary Workers: Estimating the Population", The Center for Comparative Immigration Studies Working Paper 12 (2000).

[12] 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).

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

[14] 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.

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

[16] Trajtenberg, Manuel, "The Mobility of Inventors and the Productivity of Research", Working Paper (2005).

Figure 3.1: US Ethnic Patenting

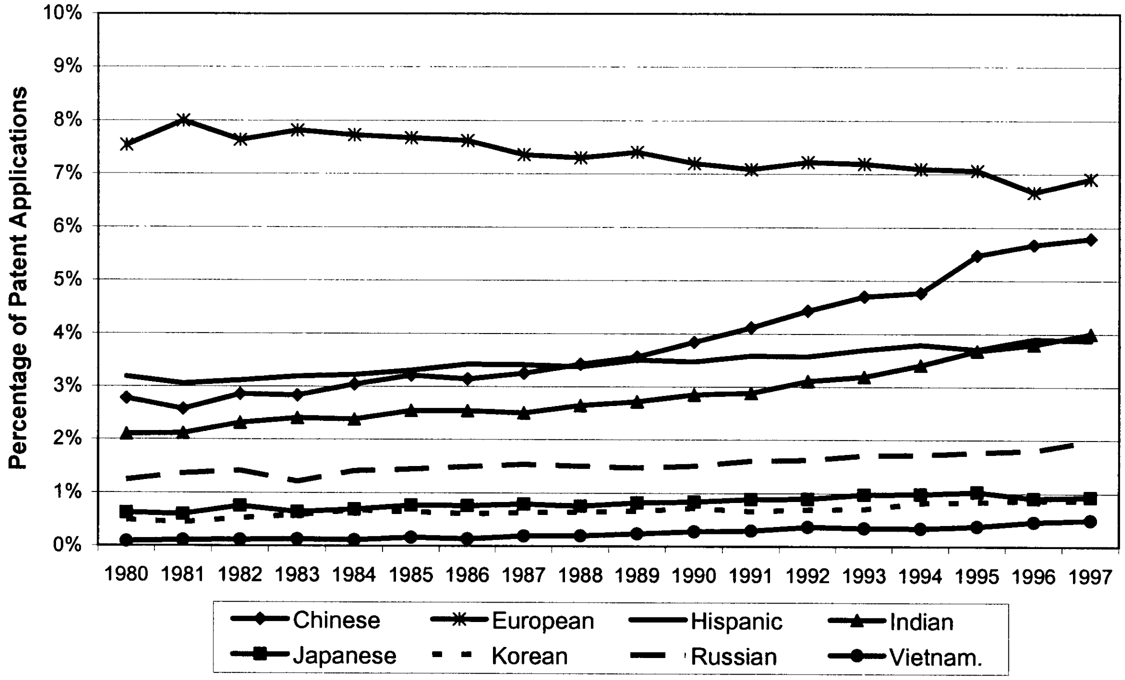


Figure 3.2: Ethnic Share by Technology

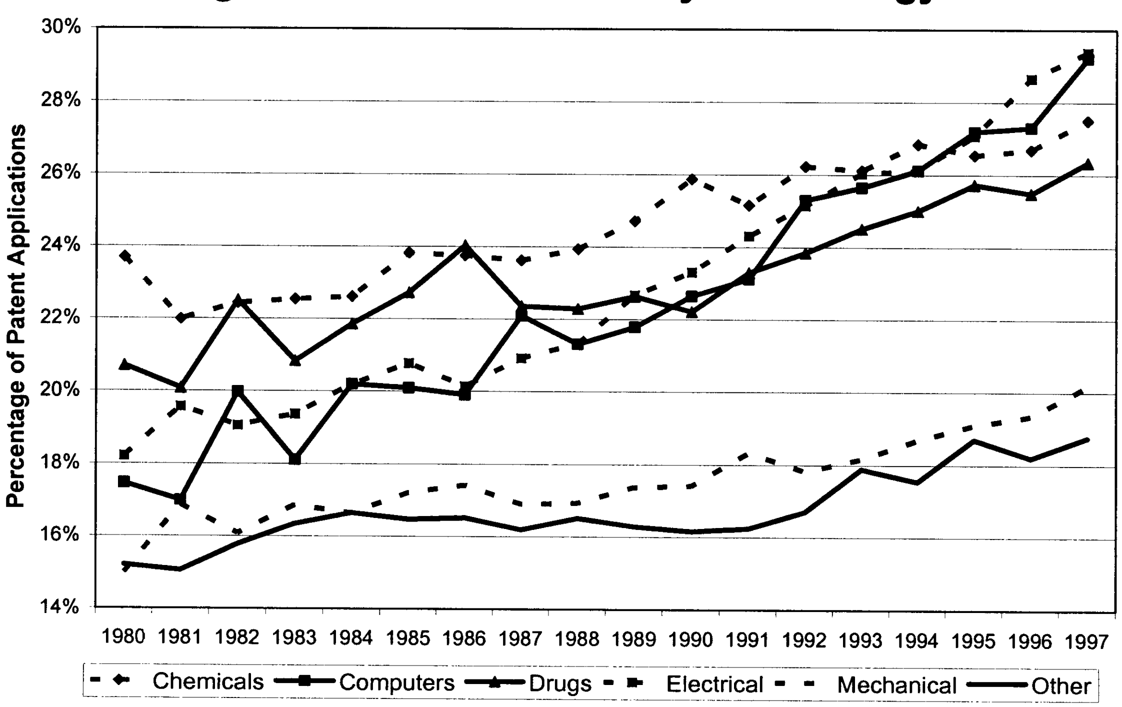


Fig. 3.3: US Ethnic Patenting - Chemicals

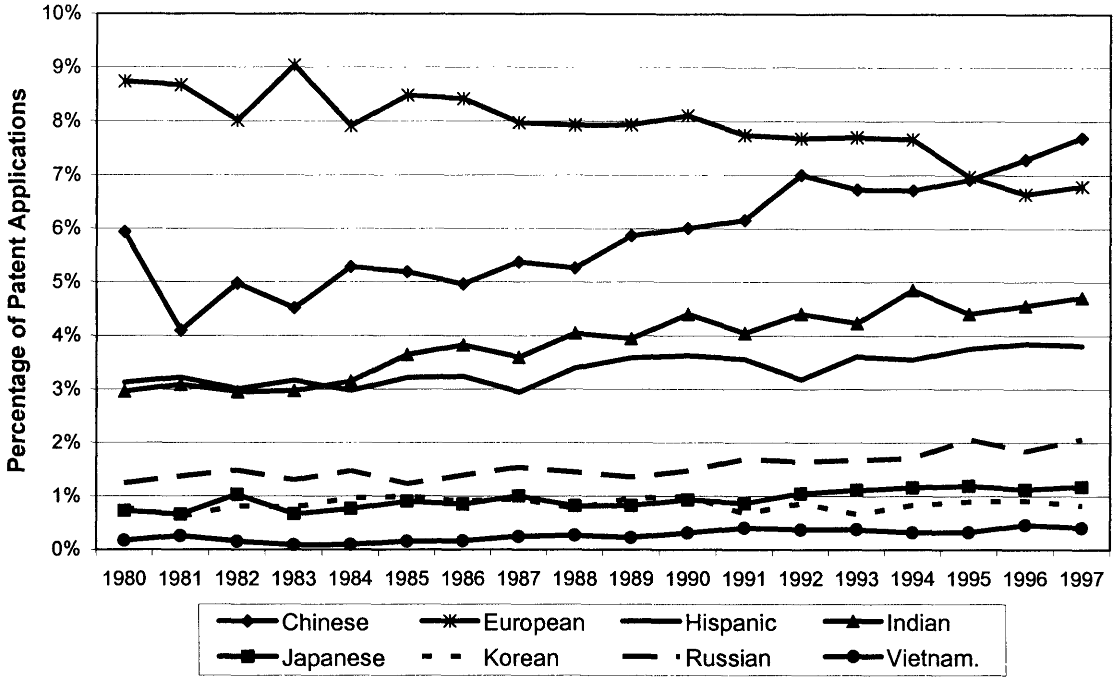


Fig. 3.4: US Ethnic Patenting - Computers

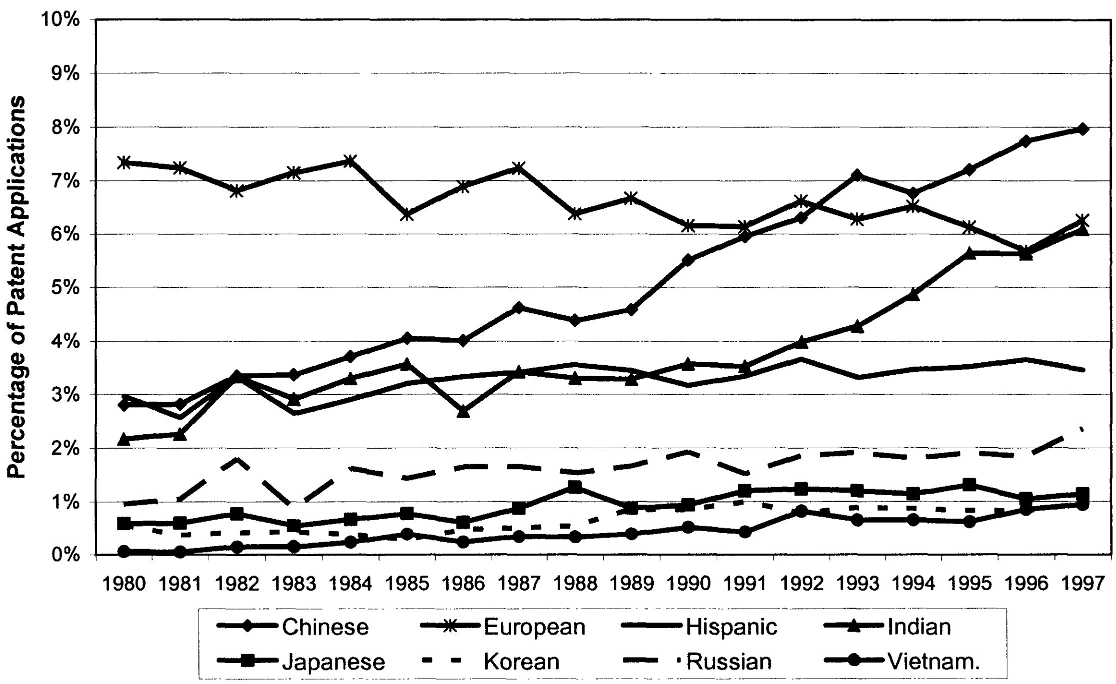


Fig. 3.5: US Ethnic Patenting - Drugs

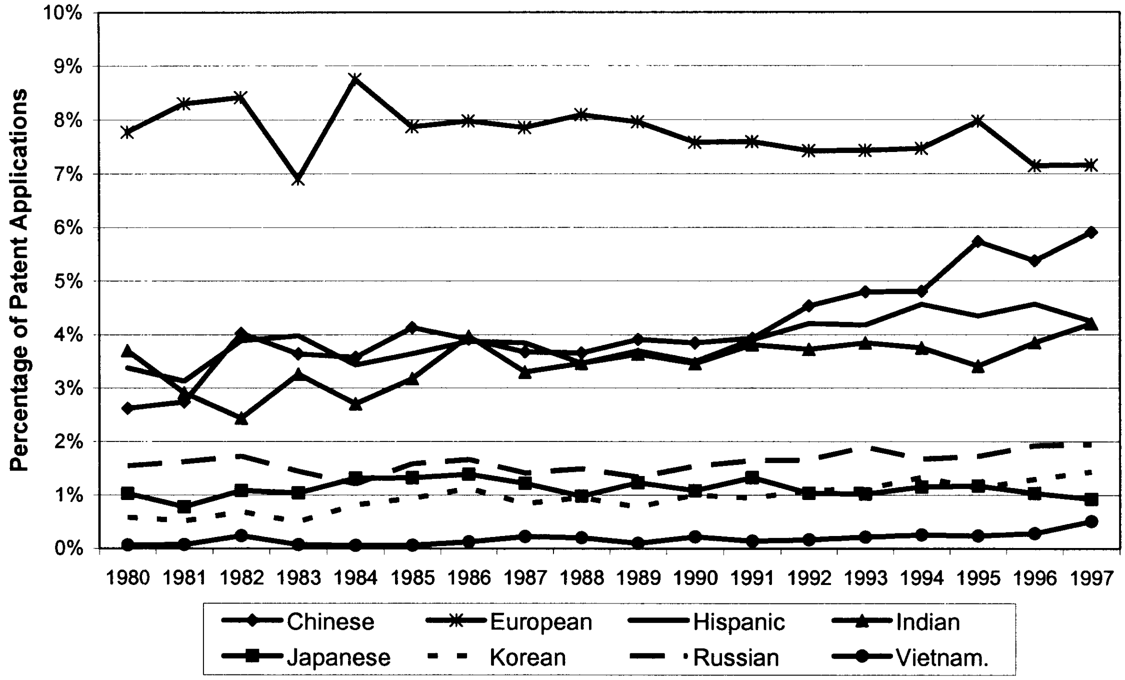


Fig. 3.6: US Ethnic Patenting - Electrical

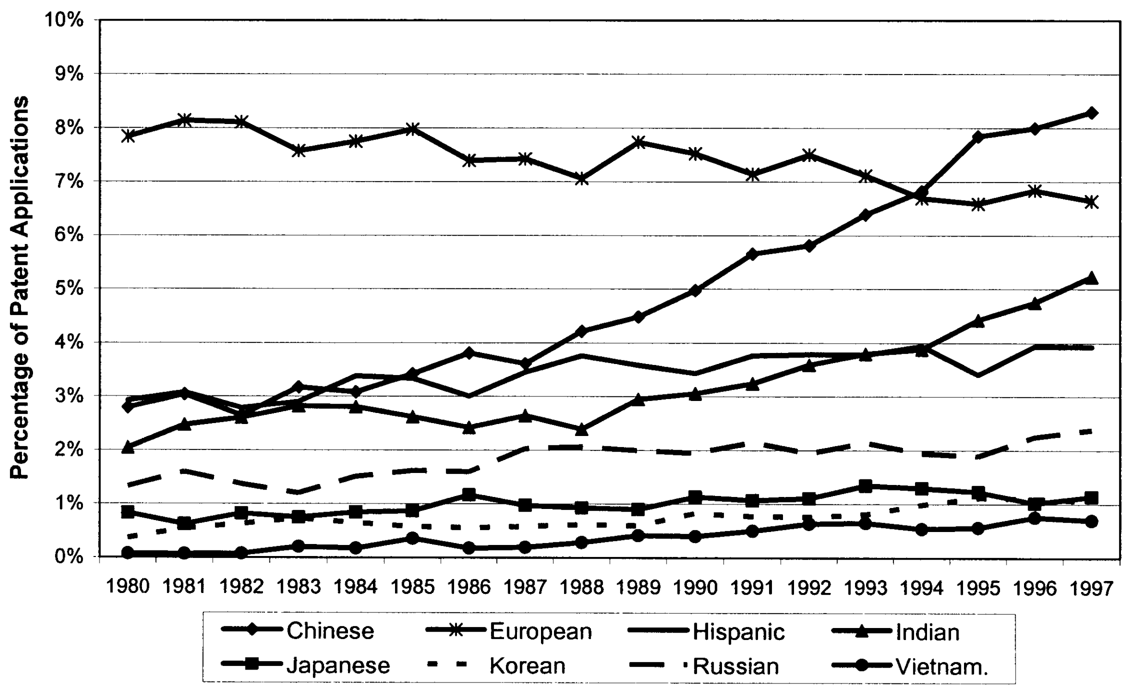


Fig. 3.7: US Ethnic Patenting - Mechanical

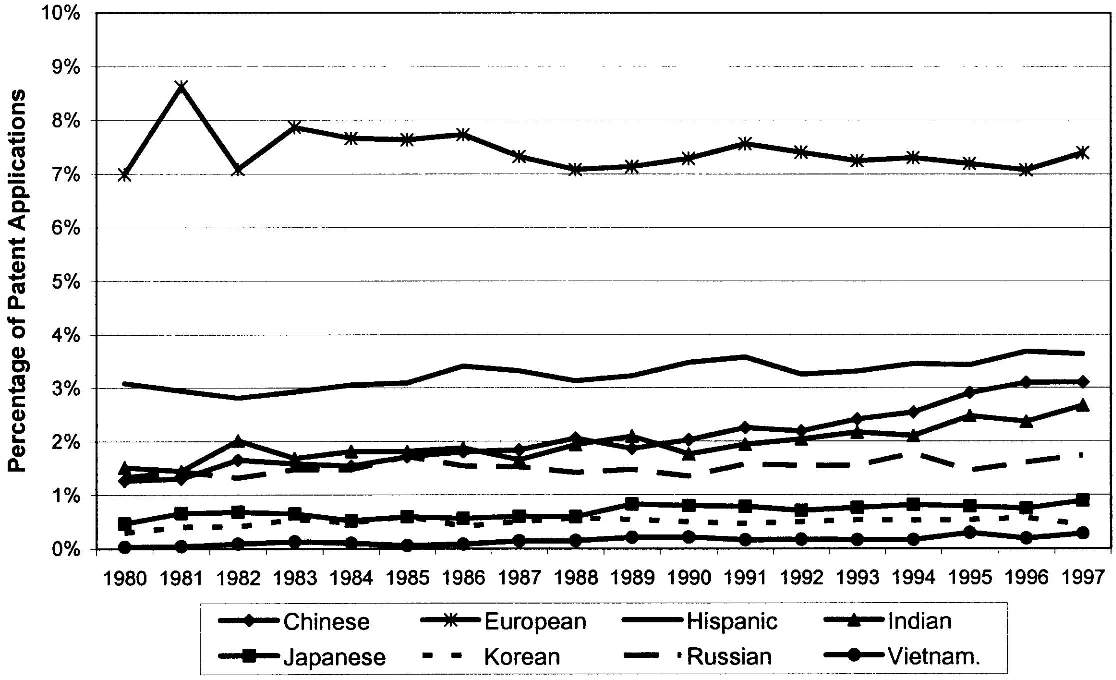


Fig. 3.8: US Ethnic Patenting - Other

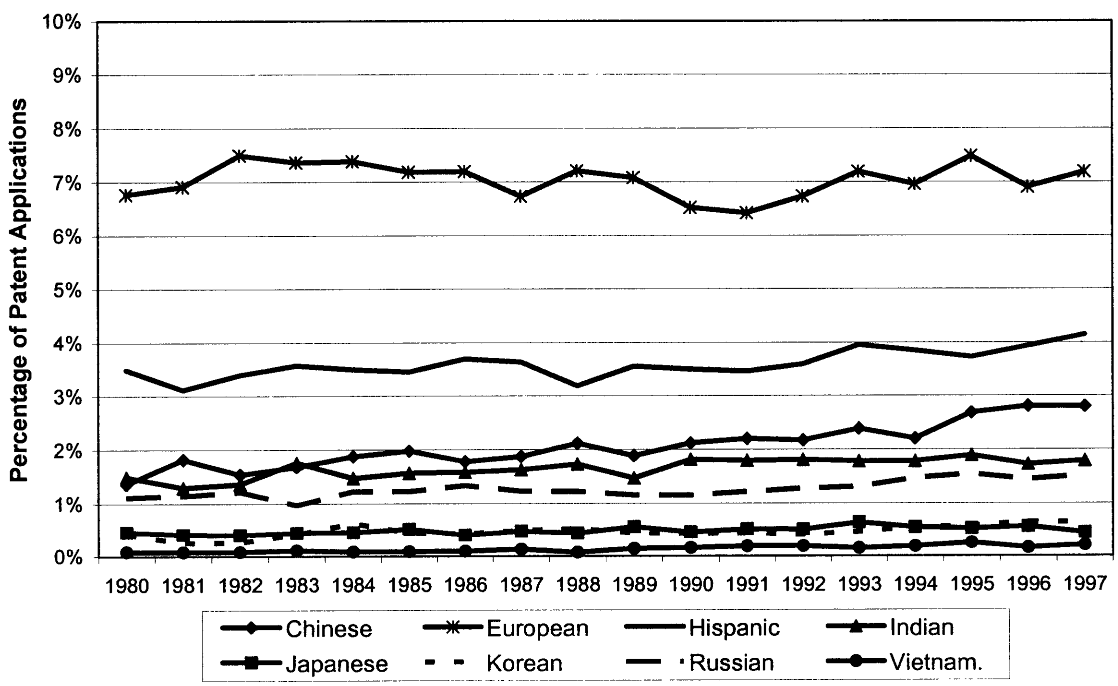


Fig. 3.9: Chinese Contribution by Technology

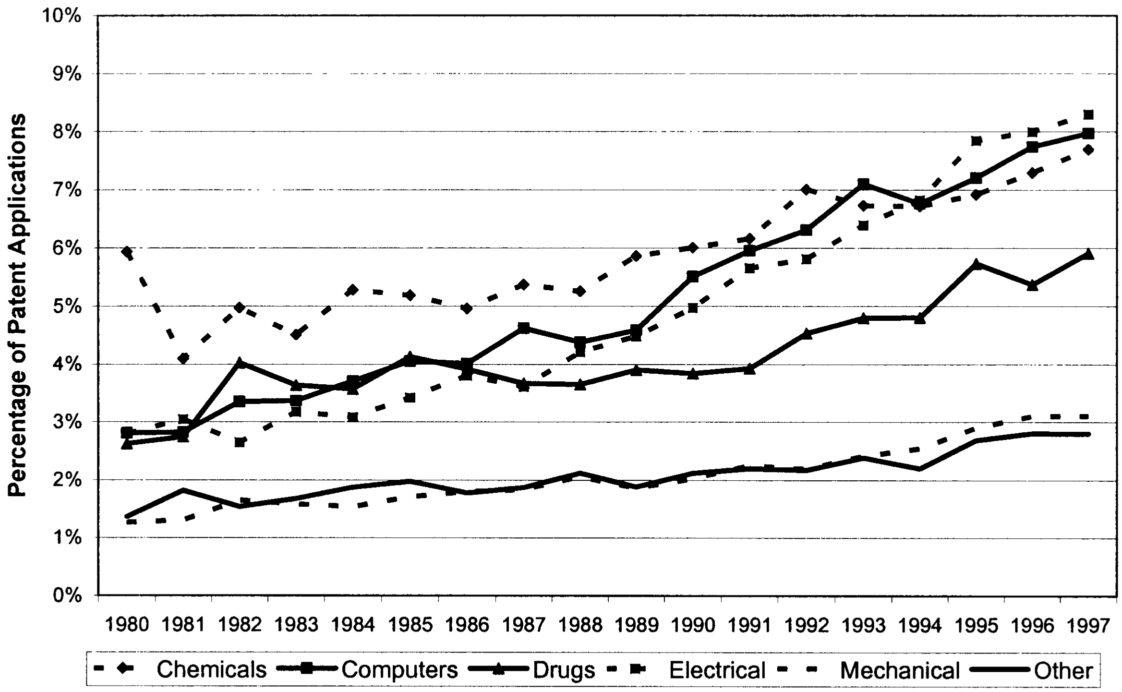
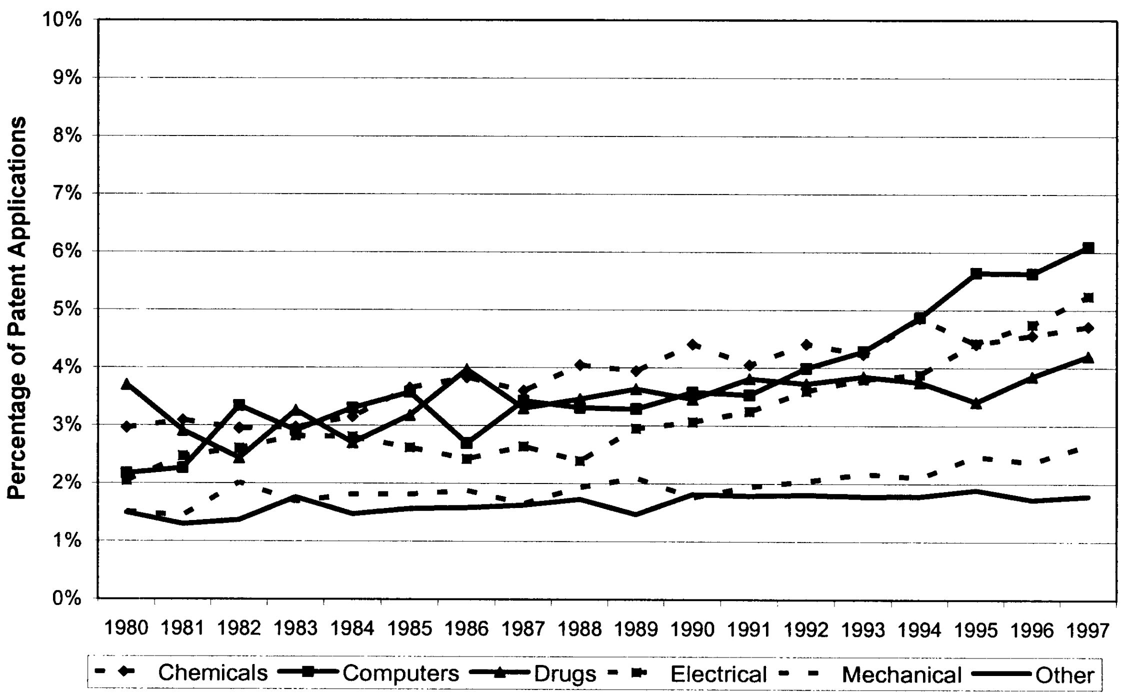


Fig. 3.10: Indian Contribution by Technology



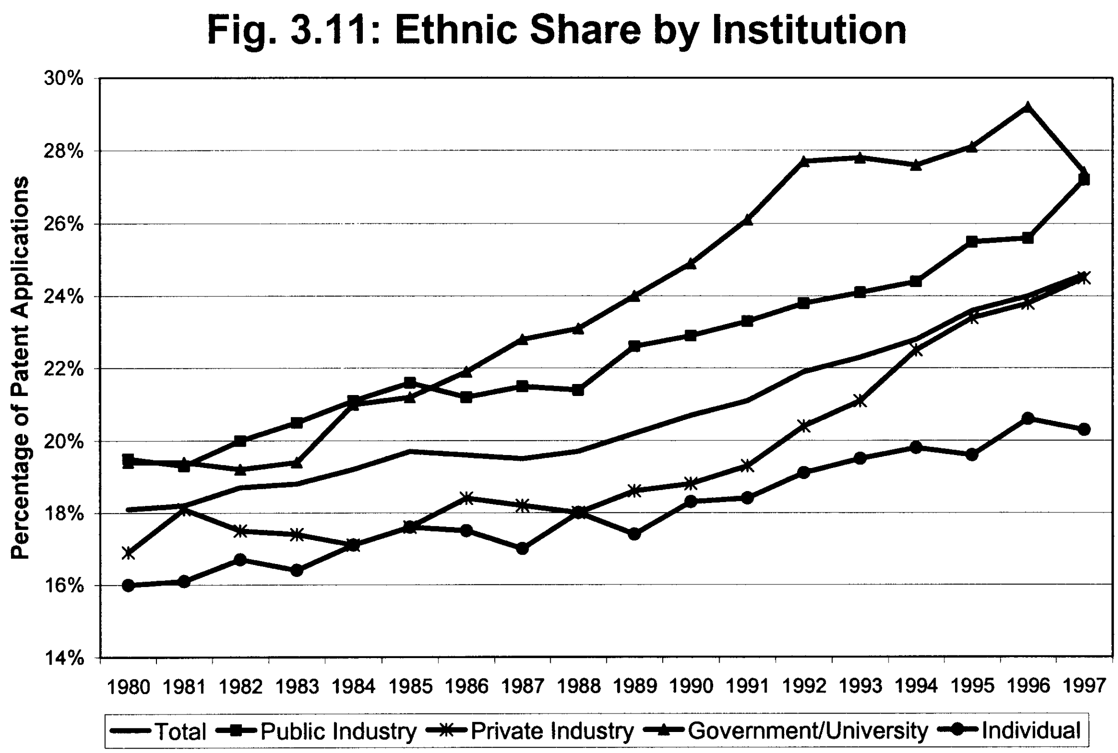


Fig. 3.11: Ethnic Share by Institution
30%
28%
26%
24%
22%
20%
18%
16%
14%
Percentage of Patent Applications
1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997
Total
Public Industry
Private Industry
Government/University
Individual

Table 3.1: Descriptive Statistics for Foreign Country Patent Records

		Match Rate %		Own-Ethnicity %		Own-Match %	
	Obs.	Full	Rest.	Full	Rest.	Full	Rest.
United Kingdom	118,475	99	96	87	84	94	93
China, Singapore	33,309	99	96	89	87	91	90
Western Europe	757,185	96	79	73	47	78	59
Hispanic Nations	12,985	98	74	72	67	94	94
India	2,083	82	76	80	83	82	85
Japan	1,043,105	98	89	100	96	100	96
South Korea	26,553	100	99	83	82	90	90
Russia	29,870	94	84	83	85	93	94
Vietnam	9	100	100	62	89	89	89

	Estimated Ethnic Composition of Country's or Region's Inventors (Full Matching)								
	ENG	CHN	EUR	HIS	IND	JAP	KOR	RUS	VNM
United Kingdom	**87**	1	4	3	2	0	0	2	0
China, Singapore	3	**89**	1	1	0	1	5	0	1
Western Europe	18	1	**73**	6	1	0	0	2	0
Hispanic Nations	14	1	9	**72**	1	1	0	2	0
India	8	1	4	5	**80**	0	0	2	0
Japan	0	0	0	0	0	**100**	0	0	0
South Korea	4	11	0	0	0	1	**83**	0	0
Russia	5	1	2	9	0	0	0	**83**	0
Vietnam	5	18	16	0	0	0	0	0	**62**

Notes: Greater China includes Mainland China, Hong Kong, Macao, and Taiwan. Western Europe includes Austria, Belgium, Denmark, Finland, France, Germany, Italy, Luxembourg, Nethlerlands, Norway, Poland, Sweden, and Switzerland. Hispanic Nations includes Argentina, Belize, Brazil, Chile, Columbia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Philippines, Portugal, Spain, Uruguay, and Venezuela. Russia includes former Soviet Union countries.

Table 3.2: Descriptive Statistics for US Ethnic Patents

	Ethnicity of Inventor								
	English	Chinese	European	Hispanic	Indian	Japanese	Korean	Russian	Vietnam.
1975-1986 Share	81.6%	2.7%	7.8%	3.1%	2.2%	0.6%	0.5%	1.3%	0.1%
1987-1997 Share	77.3%	5.0%	7.0%	3.6%	3.5%	0.8%	0.8%	1.6%	0.4%
Chemicals	75.8%	5.7%	7.9%	3.4%	3.7%	0.9%	0.8%	1.5%	0.3%
Computers	76.5%	5.7%	6.4%	3.3%	4.2%	0.9%	0.7%	1.6%	0.6%
Pharmaceuticals	76.1%	4.8%	7.6%	4.0%	3.6%	1.1%	0.9%	1.6%	0.2%
Electrical	77.0%	5.2%	7.4%	3.4%	3.3%	0.9%	0.8%	1.7%	0.4%
Mechanical	82.8%	2.1%	7.3%	3.1%	2.1%	0.6%	0.5%	1.4%	0.1%
Miscellaneous	83.2%	2.1%	7.1%	3.5%	1.8%	0.5%	0.4%	1.2%	0.1%
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)
1990 Bachelors %	87.6%	2.7%	2.3%	2.4%	2.3%	0.6%	0.5%	0.4%	1.2%
1990 Masters %	78.9%	6.7%	3.4%	2.2%	5.4%	0.9%	0.7%	0.8%	1.0%
1990 Doctorate %	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). MSA percentages are for 1985-1997. 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 recoding 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 groupings listed in Table 4; English provides a residual in the Census statistics.

Table 3.3: Ethnic Inventor Contributions by MSA

	Ethnic Patenting Share			Total Patenting Share			Total Patent Counts		
	1980	1990	1997	1980	1990	1997	1980	1990	1997
Atlanta, GA	0.4%	0.7%	1.0%	0.8%	0.9%	1.3%	287	489	808
Austin, TX	0.6%	1.2%	2.1%	0.5%	0.9%	1.9%	192	470	1196
Baltimore, MD	0.6%	0.8%	0.8%	0.7%	0.8%	0.8%	279	423	474
Boston, MA	3.3%	3.0%	2.5%	2.7%	2.7%	2.3%	1003	1440	1438
Buffalo, NY	0.5%	0.7%	0.5%	0.5%	0.6%	0.4%	187	297	263
Charlotte, SC	0.2%	0.3%	0.2%	0.3%	0.3%	0.3%	126	172	213
Chicago, IL	5.9%	5.1%	3.9%	5.2%	4.6%	3.7%	1948	2451	2317
Cincinnati, OH	0.8%	0.8%	0.9%	0.9%	1.1%	1.2%	344	583	766
Cleveland, OH	2.1%	1.5%	1.2%	2.1%	1.7%	1.4%	811	920	863
Columbus, OH	1.1%	0.4%	0.4%	0.8%	0.5%	0.4%	287	276	276
Dallas-Fort Worth, TX	1.2%	1.9%	2.2%	1.5%	2.0%	2.1%	574	1068	1302
Denver, CO	0.9%	0.9%	0.9%	1.0%	1.2%	1.3%	378	620	777
Detroit, MI	2.8%	3.2%	3.0%	3.3%	3.2%	3.2%	1228	1719	1999
Greensboro-W. Salem, NC	0.2%	0.1%	0.2%	0.3%	0.3%	0.3%	99	163	186
Hartford, CT	0.8%	0.7%	0.5%	0.9%	0.8%	0.6%	323	410	354
Houston, TX	2.6%	2.8%	1.9%	2.9%	2.7%	2.1%	1096	1449	1269
Indianapolis, IN	0.6%	0.4%	0.4%	0.8%	0.6%	0.7%	286	323	446
Jacksonville, NC	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%	44	70	103
Kansas City, MO	0.1%	0.2%	0.2%	0.4%	0.3%	0.4%	133	178	219
Las Vegas, NV	0.0%	0.1%	0.1%	0.1%	0.1%	0.2%	21	67	132
Los Angeles, CA	6.7%	6.7%	7.1%	6.4%	5.8%	5.5%	2410	3086	3383
Memphis, TN	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	70	92	126
Miami, FL	1.3%	1.1%	1.0%	0.9%	0.8%	0.7%	342	421	460
Milwaukee, WI	0.8%	0.9%	0.7%	0.9%	0.9%	0.8%	326	503	519
Minneapolis-St. Paul, MN	2.0%	1.8%	2.0%	2.2%	2.3%	2.7%	847	1231	1674

Table 3.3: Ethnic Inventor Contributions by MSA (continued)

	Ethnic Patenting Share			Total Patenting Share			Total Patent Counts		
	1980	1990	1997	1980	1990	1997	1980	1990	1997
Nashville, TN	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	58	101	93
New Orleans, LA	0.4%	0.3%	0.1%	0.2%	0.2%	0.1%	90	131	92
New York, NY	15.6%	13.3%	10.6%	10.7%	8.6%	7.7%	4028	4573	4731
Norfolk-VA Beach, VA	0.1%	0.2%	0.1%	0.2%	0.3%	0.2%	74	140	127
Orlando, FL	0.2%	0.3%	0.3%	0.3%	0.3%	0.3%	108	176	158
Philadelphia, PA	5.1%	4.8%	3.4%	4.2%	4.0%	3.2%	1599	2136	2000
Phoenix, AZ	0.8%	1.0%	1.3%	1.1%	1.3%	1.4%	404	676	844
Pittsburgh, PA	2.2%	1.4%	0.8%	2.0%	1.3%	0.9%	771	677	585
Portland, OR	0.5%	0.5%	1.0%	0.8%	0.7%	1.0%	301	367	641
Providence, RI	0.9%	1.0%	1.1%	0.8%	0.9%	0.9%	291	493	546
Raleigh-Durham, NC	0.4%	0.5%	0.9%	0.5%	0.5%	1.0%	172	285	599
Richmond, VA	0.2%	0.3%	0.1%	0.3%	0.3%	0.2%	96	176	125
Sacramento, CA	0.3%	0.4%	0.4%	0.2%	0.4%	0.5%	88	193	280
St. Louis, MO	0.7%	0.9%	0.9%	0.9%	1.0%	0.9%	323	509	538
Salt Lake City, UT	0.3%	0.3%	0.4%	0.5%	0.5%	0.6%	170	286	380
San Antonio, TX	0.3%	0.2%	0.4%	0.2%	0.2%	0.3%	60	106	191
San Diego, CA	1.1%	1.6%	1.9%	1.2%	1.6%	1.7%	445	877	1042
San Francisco, CA	7.7%	8.1%	16.5%	5.6%	6.0%	10.3%	2108	3213	6398
Seattle, WA	0.8%	1.2%	1.3%	1.1%	1.3%	1.4%	425	687	866
Tallahassee, FL	0.2%	0.4%	0.4%	0.5%	0.5%	0.5%	174	269	331
Washington, DC	1.4%	1.6%	1.4%	1.3%	1.5%	1.3%	488	790	793
West Palm Beach, FL	0.2%	0.6%	0.4%	0.3%	0.5%	0.5%	130	265	296
Not in a MSA	24.8%	25.7%	22.4%	31.0%	32.4%	30.1%	11677	17255	18611
Total	100%	100%	100%	100%	100%	100%	39701	55292	63827

Table 3.A1: Most Common Ethnic Surnames for Inventors Residing in the US

Chinese		English		European		Hispanic / Filipino		Indian / Hindi	
Chan	1335	Adams	2545	Abel	180	Acosta	65	Adler	319
Chang	3214	Allen	3019	Albrecht	327	Acquaviva	58	Agarwal	184
Chao	427	Anderson	6271	Antos	220	Adell	86	Aggarwal	94
Chau	163	Bailey	1559	Auerbach	138	Alvarez	235	Agrawal	375
Chen	4306	Baker	2883	Baer	286	Arroyo	68	Ahmad	148
Cheng	1057	Bell	1677	Bauer	931	Ayer	134	Ahmed	337
Cheung	351	Bennett	1522	Beck	1094	Ayres	151	Akram	98
Chiang	584	Brown	6818	Bender	332	Bales	173	Ali	193
Chien	164	Burns	1145	Berg	933	Bartos	58	Arora	95
Chiu	427	Butler	1131	Berger	784	Blanco	71	Ash	183
Chou	513	Campbell	2339	Bodor	135	Bolanos	116	Aslam	90
Chow	468	Carlson	1542	Budzich	112	Boles	60	Badesha	90
Chu	1184	Carter	1522	Caron	188	Cabrera	62	Baliga	111
Chuang	160	Clark	3273	Cerami	117	Calderon	77	Banerjee	137
Fan	363	Cohen	1513	Chandraratna	126	Camacho	59	Basu	101
Fang	256	Cole	1228	Collette	116	Cardenas	58	Bhat	94
Feng	184	Collins	1681	Crivello	119	Carnes	69	Bhatia	172
Fong	298	Cook	1994	D'Amico	126	Castillo	61	Bhatt	105
Fu	220	Cooper	1788	Dietrich	200	Chavez	76	Bhattacharya	105
Fung	248	Cox	1370	Dietz	298	Contreras	61	Bhattacharyya	90
Guo	182	Davis	5229	Eberhardt	136	Cruz	118	Bose	159
Han	379	Edwards	1962	Eckenhoff	118	D'Alelio	69	Brunelle	116
He	194	Erickson	1191	Effland	133	D'Silva	87	Chandra	91
Ho	817	Evans	2494	Ehrlich	187	Das	409	Chatterjee	293
Hou	161	Fischer	1126	Ferrari	124	Delgado	102	Chattha	90
Hsieh	517	Fisher	1585	Fischell	161	Dias	101	Cherukuri	134
Hsu	1153	Foster	1650	Fuchs	219	Diaz	303	Chubb	90
Hu	494	Fox	1230	Gelardi	127	Dominguez	111	Datta	202
Huang	1545	Gardner	1257	Grabbe	136	Duran	87	Desai	442
Hung	317	Gordon	1500	Grasselli	135	Elias	163	Dixit	132
Jiang	281	Graham	1284	Gunther	173	Fernandes	81	Dutta	103
Kao	350	Gray	1521	Guttag	127	Fernandez	285	Fazan	107
Kung	225	Green	2051	Haas	514	Francisco	64	Gaffar	150
Kuo	600	Hall	2928	Hansen	1730	Freitas	78	Gandhi	105
Lai	466	Hanson	1289	Hartman	757	Gagnon	157	Ganguly	110
Lam	491	Harris	2838	Hartmann	220	Garcia	612	Garg	138
Lau	578	Hayes	1200	Hause	134	Garza	76	Ghosh	237
Lee	1325	Hill	2061	Hecht	142	Gomes	89	Goel	208
Leung	500	Hoffman	1433	Heinz	116	Gomez	179	Goli	100
Lew	403	Howard	1158	Henrick	123	Gonsalves	60	Gupta	851
Li	1652	Hughes	1340	Horodysky	232	Gonzales	131	Harandi	159
Liang	418	Jackson	2319	Horvath	221	Gonzalez	441	Hassan	110
Liao	194	Jensen	1227	Jacobs	1122	Gutierrez	387	Hussain	98
Lien	202	Johnson	10718	Kanner	118	Halasa	147	Imran	118
Lim	178	Johnston	1167	Kasper	155	Hernandez	324	Iyer	219
Lin	2348	Jones	6068	Kempf	144	Herrera	71	Jain	397
Ling	211	Keller	1132	Knapp	529	Herron	220	Joshi	319
Liu	1981	Kelly	1685	Knifton	201	Jimenez	90	Kamath	111
Lo	503	Kennedy	1303	Koenig	307	Konopka	62	Kapoor	145
Lu	650	King	2591	Kresge	125	Kulprathipanja	76	Khanna	210
Ma	437	Klein	1372	Kukes	123	Lee	126	Krishnakumar	97
Mao	178	Larson	1561	Lange	443	Lieb	62	Krishnamurthy	119

Table 3.A1: Most Common US Ethnic Surnames (continued)

Chinese		English		European		Hispanic / Filipino		Indian / Hindi	
Ng	451	Lee	5438	Lapeyre	161	Lomas	63	Krishnan	167
Ong	232	Lewis	2788	Laskaris	120	Lopez	377	Kulkarni	119
Pai	198	Long	1446	Lemelson	299	Machado	79	Kumar	777
Pan	444	Marshall	1213	Lorenz	198	Mares	82	Lal	175
Peng	165	Martin	4214	Ludwig	304	Marin	103	Malik	179
Shen	669	Miller	9011	Lutz	402	Marquez	75	Mathur	112
Shi	194	Mitchell	1862	Maier	319	Martinez	534	Mehra	102
Shieh	151	Moore	3572	Mayer	704	Medina	92	Mehrotra	126
Shih	513	Morgan	1663	Meyer	1815	Menard	89	Mehta	436
Shu	264	Morris	1908	Milberger	114	Mendoza	79	Menon	125
Shum	152	Murphy	1968	Mitra	140	Molina	85	Mishra	114
Sih	318	Murray	1246	Molnar	162	Molitor	71	Misra	113
Song	286	Myers	1573	Morin	170	Munoz	62	Mookherjee	271
Su	443	Nelson	3854	Mueller	1349	Nestor	96	Nair	203
Sun	691	Olson	1722	Muller	546	Nunez	66	Narang	96
Tai	178	Palmer	1145	Nagel	263	Ondetti	104	Narayanan	231
Tam	283	Parker	1976	Nilssen	213	Ortega	71	Natarajan	144
Tan	366	Peters	1200	Novak	436	Ortiz	168	Nath	102
Tang	769	Peterson	2769	Pagano	112	Padilla	66	Parekh	107
Teng	242	Phillips	2299	Pastor	204	Pallos	92	Parikh	123
Ting	213	Price	1148	Pittet	119	Pereira	87	Patel	1819
Tong	270	Reed	1625	Ponticello	126	Perez	269	Patil	188
Trinh	178	Richardson	1224	Rao	241	Pfiester	69	Prasad	240
Tsai	441	Roberts	2524	Reitz	138	Quintana	77	Puri	108
Tsang	255	Robinson	2112	Rivier	125	Ramirez	168	Qureshi	102
Tsao	218	Rogers	1770	Roman	226	Ramos	114	Rahman	133
Tseng	281	Ross	1499	Rostoker	201	Regnier	70	Raj	97
Tung	302	Russell	1476	Schmidt	2025	Reis	86	Rajagopalan	108
Wan	173	Ryan	1245	Schneider	1377	Reno	73	Ramachandran	175
Wang	3381	Scott	2191	Schultz	1230	Reyes	69	Ramakrishnan	94
Wei	428	Shaw	1535	Schulz	518	Rivera	174	Raman	95
Wong	2210	Smith	13623	Schwartz	1493	Robeson	96	Ramesh	96
Woo	354	Snyder	1402	Schwarz	418	Rodrigues	74	Rao	526
Wu	1956	Stevens	1317	Speranza	188	Rodriguez	520	Ravichandran	91
Xu	368	Stewart	1678	Spitz	119	Romero	103	Saari	93
Yan	297	Sullivan	1473	Straeter	253	Ruiz	159	Sandhu	252
Yang	1315	Taylor	4081	Theeuwes	224	Salazar	77	Shah	1115
Yao	208	Thomas	2923	Trokhan	111	Sanchez	327	Sharma	408
Yee	335	Thompson	3736	Uskokovic	124	Silva	217	Singh	914
Yeh	482	Turner	1622	Van Scott	115	Solar	70	Singhal	97
Yen	304	Walker	2758	Vock	407	Soled	59	Sinha	149
Yin	159	Ward	1679	Wachter	124	Soto	62	Sircar	171
Yu	1207	Watson	1289	Wagner	1512	Souza	95	Srinivasan	271
Yuan	236	White	3792	Weber	1646	Suarez	99	Srivastava	177
Zhang	629	Williams	5982	Weder	530	Torres	172	Subramanian	173
Zhao	223	Wilson	4650	Weiss	935	Varga	70	Thakur	118
Zheng	162	Wood	2257	Wolf	961	Vasquez	64	Varma	117
Zhou	269	Wright	2798	Zimmerman	931	Vazquez	73	Venkatesan	116
Zhu	196	Young	3593	Zimmermann	119	Vinals	231	Vora	176

Table 3.A1: Most Common US Ethnic Surnames (continued)

Japanese		Korean		Russian		Vietnamese	
Arakawa	46	Ahn	94	Aghajanian	64	Bahn	7
Asato	73	Bae	65	Anscher	44	Banh	6
Chen	36	Baek	25	Askin	39	Be	5
Doi	51	Bak	34	Avakian	35	Bearce	7
Fujii	40	Bang	34	Babler	58	Bi	35
Fujimoto	55	Bark	23	Banko	34	Bich	15
Fujioka	54	Cha	20	Barna	46	Bien	59
Fukuda	64	Chai	77	Benko	33	Bihn	7
Furukawa	35	Chin	541	Blonder	66	Bui	109
Hasegawa	96	Cho	448	Borsuk	42	Can	6
Hashimoto	72	Choe	100	Danko	52	Chich	5
Hayashi	103	Choi	322	Dombroski	37	Diem	17
Hey	33	Chon	16	Duvdevani	42	Dien	6
Higham	35	Chong	99	Elko	36	Diep	26
Higuchi	76	Choo	37	Favstritsky	44	Dinh	60
Honda	40	Chun	155	Frenkel	50	DoMinh	16
Hori	33	Chung	688	Garabedian	60	Doan	204
Hornak	53	Drozd	22	Gelfand	81	Dominh	5
Ide	111	Ewbank	21	Georgiev	41	Donlan	17
Imai	92	Eyuboglu	27	Ginzburg	62	Dotrong	8
Inoue	33	Gang	20	Gitlin	50	Dovan	26
Irick	84	Gu	118	Godlewski	38	Duan	33
Ishida	34	Hahm	18	Goralski	57	Due	6
Ishii	37	Hahn	620	Gordin	42	Duong	52
Ishikawa	59	Hansell	29	Gorin	58	Eskew	7
Ito	140	Hogle	17	Gregorian	34	Gran	11
Iwamoto	32	Hohn	19	Grinberg	64	Hoang	103
Iwasaki	48	Hone	16	Grushkin	37	Hopping	8
Izu	45	Hong	319	Grzybowski	36	Huynh	101
Kaneko	72	Hosking	24	Gurevich	45	Huynh-Ba	8
Kato	59	Hwang	517	Guzik	48	Khau	5
Kaun	32	Hyun	32	Hrib	37	Khaw	9
Kautz	64	Ih	16	Hynecek	58	Khieu	13
Kawakami	33	Im	37	Ibrahim	103	Khu	5
Kawasaki	56	Jang	94	Iranmanesh	44	Kiem	5
Kaya	44	Jeong	34	Ivanov	37	Lahue	10
Kimura	63	Ji	42	Janko	34	Laursen	19
Kino	37	Jin	175	Jastrzebski	37	Lavan	11
Kirihata	34	Joo	19	Juhasz	39	Le	415
Kiwala	132	Ju	100	Kahle	89	Le Duc	6
Kobayashi	125	Jung	205	Kaminski	254	Le Van	7
Maki	81	Kahng	17	Kaminsky	62	Leen	10
Maruyama	32	Kang	275	Kaplinsky	49	Loan	5
Matsuda	36	Kim	1987	Keritsis	35	Luong	30
Matsumoto	78	Ko	217	Khan	62	Ly	31
Matsunaga	32	Koh	40	Khandros	55	Minh	17
Miyano	54	Koo	90	Kneller	41	Nellums	12
Mizuhara	83	Kun	54	Korsunsky	80	Nghiem	5
Mori	39	Kwak	46	Kowal	57	Ngo	196
Morita	40	Kwon	156	Kozel	33	Nguyen	1514
Moslehi	103	Lee	325	Kulka	35	Nguyen-Dinh	7
Motoyama	49	Lim	82	Kurkov	35	Nguyenphu	7

Table 3.A1: Most Common US Ethnic Surnames (continued)

Japanese		Korean		Russian		Vietnamese	
Najjar	76	Mennie	33	Lapidus	34	Nho	7
Nakagawa	74	Min	71	Lee	48	Nhu	6
Nakajima	32	Minshall	18	Lisak	36	Nieh	53
Nakamura	74	Nam	18	Lopata	50	Nim	12
Nakanishi	46	Nevins	24	Lukacs	37	Ninh	8
Nakano	53	Nyce	18	Lysenko	39	Pham	286
Nakao	41	Oh	151	Magnotta	35	Phy	19
Nemoto	50	Paek	25	Mankovitz	34	Postman	8
Nishimura	32	Paik	82	Messing	47	Quach	24
Nishioka	43	Pak	64	Metlitsky	81	Quy	6
Noda	48	Park	912	Mikhail	70	Roch	26
Ogawa	39	Quay	58	Milkovic	46	Sien	6
Ogura	57	Rhee	120	Minaskanian	39	Sinh	7
Ohkawa	48	Rhim	17	Mooradian	50	Ta	39
Okada	37	Rim	30	Nadelson	92	Takach	11
Okamoto	62	Ronen	19	Nappholz	38	Tau	7
Okumura	45	Ryang	24	Narayan	203	Thach	11
Ono	34	Ryu	46	Neuwirth	42	Thai	16
Ovshinsky	194	Sahm	24	Onopchenko	59	Thiem	10
Saito	49	Sahoo	22	Orloff	36	Thut	16
Sasaki	70	Sellstrom	23	Papadopoulos	47	Tiedt	6
Sato	134	Seo	18	Pinchuk	62	Tiep	11
Seto	37	Sheem	21	Pinsky	34	Tietjen	32
Shibata	52	Shim	101	Raber	45	To	7
Shida	45	Shin	149	Rabii	34	Ton-That	6
Shimizu	32	Shinn	64	Rabinovich	52	Tran	631
Shinkai	48	Sim	43	Rubsamen	47	Trandai	7
Shoji	45	Sjostrom	18	Sahatjian	40	Trang	12
Sigmund	35	So	149	Sarkisian	35	Trank	7
Suto	33	Sohn	42	Sarraf	38	Tri	7
Suzuki	152	Son	72	Schwan	77	Trieu	8
Takahashi	81	Sue	36	Simko	70	Trong	7
Takekoshi	50	Suh	188	Sipos	38	Truc	8
Takeuchi	61	Suk	23	Skowronski	44	Tu	190
Tamura	50	Sung	255	Smetana	42	Tuten	19
Tanaka	191	Uhm	16	Sofranko	61	Tuy	14
Ueda	34	Um	22	Sorkin	52	Ty	21
Wada	47	Whang	40	Stanko	37	Van	18
Watanabe	140	Won	48	Tabak	85	Van Cleve	30
Yamada	62	Yi	56	Tepman	41	Van Dam	8
Yamaguchi	42	Yim	55	Terzian	75	Van Le	18
Yamamoto	178	Yohn	16	Tsinberg	38	Van Nguyen	11
Yamasaki	42	Yoo	133	Tults	34	Van Pham	8
Yamashita	32	Yoon	405	Uram	43	Van Phan	27
Yasuda	50	You	58	Vartanian	42	Van Tran	13
Yasui	51	Yuh	40	Veltman	39	Vo	95
Yokoyama	52	Yum	69	Warchol	34	Vo-Dinh	19
Yoshida	127	Yun	68	Wasilewski	34	Vu	141
Yuan	40	Zhu	24	Welsch	44	Vuong	33

Chapter 4

Income Inequality and Social Norms for Compensation Differentials and Government-Led Redistribution

Summary 4 *In cross-sectional studies, countries with greater income inequality typically exhibit less support for government-led redistribution and greater acceptance of wage inequality (e.g., United States versus Western Europe). If individual nations evolve along this pattern, a vicious cycle could form with reduced social concern amplifying primal increases in inequality due to forces like skill-biased technical change. Exploring movements around these long-term levels, however, this study finds increases in inequality are met with greater, not less, support for redistribution. Larger compensation differentials are accepted as inequality grows, but of a smaller magnitude than the actual increase. These findings suggest short-run responses in social norms do not amplify inequality shocks.*

4.1 Introduction

Accounting for the substantial increase in wage and income inequality over the last three decades is a central theme of recent economic research. The bulk of the literature focuses on forces operating within the labor market on the supply and demand for skilled workers. These include the slower growth rate in the supply of educated workers, the introduction of labor-saving production and computing technologies, and capital deepening.[1] Others researchers consider structural changes of the labor market itself, like the decline of institutions and policies that have historically compressed the wage structure (e.g., unions, minimum wages)[2] and the proliferation of "superstar" labor markets where top performers earn disproportionate sums to those just behind them.[3] The potential erosion of social norms regarding compensation inequality and redistribution is also widely discussed. For the United States, particular emphasis is placed on the explosion in executive pay and deepening within-establishment inequality.[4]

While the early work considers each of these determinants in isolation, it is increasingly clear that the interactions among the factors bear significant responsibility. Moreover, a greater potential for the entrenchment or amplification of inequality exists in this general-equilibrium setting.[5] Taking skill-biased technical change as an example, its individual effect on inequality will be checked in the long-run as firms substitute towards cheaper factors of production or labor supplies adjust. If the bias is sufficient, however, the technical change and its concomitant increase in inequality may also prompt lasting changes in the structure of the labor market (e.g., deunionization, increased segregation of skilled workers) that magnify its solitary effect. Of course, interactions can alternatively dampen inequality shocks.

This potential for amplification is particularly strong for social norms regarding income equalization. First, if changes in inequality directly influence ideology, then social norms are a propagation channel for any shock to the income distribution, regardless of the source. Second, of all the factors discussed, social attitudes are the least governed (if at all) by market-like mechanisms that can retard excessive changes. The potential thus exists for the formation of a "vicious cycle" where increases in disparity weaken concern for wage equality or redistribution. This weakened concern affords greater future compensation differentials, a shrinking of the welfare state, and so on that further increase inequality and again shift norms. Alternatively, changes in social norms can counteract inequality increases.

Support for the vicious-cycle hypothesis can be taken from the cross-sectional distributions

[1] Berman, Bound, and Griliches (1994); Katz and Murphy (1995); Autor, Katz, and Krueger (1998); Krusell *et. al.* (2000); and Card and Lemieux (2001).

[2] DiNardo, Fortin, and Lemieux (1996); Lee (1999); Card (2001); and Golan, Perloff, and Wu (2001).

[3] Rosen (1981); Frank and Cook (1995); and Economist (1999).

[4] Bok (1993); Economist (1999); Piketty and Saez (2001); and Krugman (2002).

[5] Acemoglu, Aghion, and Violante (2001); Benabou (2002); and Hasser *et. al.* (2003).

of countries (particularly long-term OECD members) and regions of the United States. Nations with greater income inequality typically demonstrate less support for redistribution and greater acceptance of wage inequality than their more-equal counterparts. While the evolution of countries or regions along this pattern would be consistent with hypotheses of reduced social concern, this response is not guaranteed as many primal factors determining these long-term ideology positions (e.g., beliefs regarding social mobility) may be stable.[6] The empirical response of social norms to changes in inequality has yet to be explored systematically.

This paper investigates this question by focusing on short-term movements in inequality and social attitudes around the long-term level of each country or United States region. A fixed-effect estimation strategy removes permanent differences in inequality and redistribution philosophies, as well as common time trends. The contribution of this study is to characterize how the resulting longitudinal responses resemble and differ from the cross-sectional pattern. How responses differ by income class and neighborhood racial heterogeneity is also considered.[7] The primary results are drawn from a panel of countries repeatedly surveyed by the International Social Survey Programme (ISSP) and the World Value Survey (WVS). Additional support and extensions are developed through regional variation in the United States captured by the General Social Survey (GSS). To establish causality, an instrument-variable specification that exploits exogenous changes in the real federal minimum-wage rate interacted with predetermined regional characteristics is also employed.

The results of this study suggest that increases in income inequality are met with greater, not less, concern for inequality. Moreover, the greater concern translates into increases in support for government-led redistribution and more-progressive taxation. In line with changing factors of production, norms for compensation differentials do increase with greater inequality, but the response is significantly less than one-for-one. While greater class conflict is perceived along income dimensions, the increases in support for redistribution among wealthy individuals are as

[6] The determinants of this cross-sectional pattern have been a frequent and lively political-economy topic since at least de Toqueville. Alesina, Glaeser, and Sacerdote (2001) offer a broad study of why the United States has both higher inequality and a smaller welfare state than Western Europe, including appropriate references.

[7] Political-economy models differ in their predictions of how responses to inequality changes vary by income class. Piketty (1995) constructs a Rawlsian model where increases in the inequality of opportunity, holding fixed beliefs regarding the incentive costs of effort, promote greater support for redistribution independent of current income. On the other hand, the standard median-voter model (e.g., Meltzer and Richard 1981) suggests increases in inequality lead to a divergence in preferences for redistribution as gaps to the median income widen.

strong as those of poorer individuals. Taken together, these findings suggest localized increases in inequality alone are unlikely to prompt a vicious cycle of changing social norms amplifying primal inequality changes.

Before proceeding to the analysis, it is worthwhile to place these findings in the context of several other research strands. First, it was earlier noted the decisions of skilled workers to take higher-wage jobs may lead to structural changes in the labor market that promote a further expansion of inequality. For example, Acemoglu, Aghion, and Violante (2001) argue biased technical change increases the outside options of skilled workers and thereby prompts the decline of unions, an institution that often compresses the wage structure. The rational decisions by skilled workers to take the higher-paid, non-union jobs are not at odds with this study's findings; in fact, increases in inequality are found to be associated with modest increases in support for wage differentials. The important point this study makes, however, is that this segregation is not accompanied by a reduced concern over distributive equality.

It is also important to distinguish ideology regarding inequality from other norms that influence perceptions of distributive justice. Political economists have long considered how beliefs regarding the determinants of success affect attitudes towards redistribution. Individuals and societies who believe hard work and effort are more important for outcomes than luck or ancestry often choose systems characterized by higher inequality and lower redistribution.[8] Past mobility experiences and future expectations of social position are also significant for attitudes towards income equalization.[9] If the forces driving higher inequality also alter these underlying beliefs, then social norms for equality may weaken. The analysis presented below controls for changes in these social-mobility beliefs to isolate the effect of inequality, concluding that the increase in inequality alone is insufficient for the formation of a vicious cycle. Additional research needs to evaluate whether other norms (and non-norms) multiplier mechanisms exist.

Finally, while inequality has risen throughout the income distribution, the exceptional increase in the very upper echelons (i.e., the top 1% and higher) is one of its more notable traits

[8] Alesina and Angeletos (2004) demonstrate how differences in these beliefs can create multiple equilibria among otherwise similar economies, as rational agents select taxation and redistribution policies (and their associated distortions) that fulfill their original expectations. Benabou and Tirole (2002) develop a related general-equilibrium model where different beliefs regarding how just the world is create two distinct redistribution states.

[9] Piketty (1995); Benabou and Ok (2001); and Alesina and La Ferrara (2001).

(e.g., Piketty and Saez 2003). Many suggest this concentration of wealth has lead to a substantial shift in norms regarding executive compensation and a disproportionate political influence for elites. Unfortunately, the data employed here do not afford an analysis of these super-wealthy individuals, and examinations of social norm changes by income classes are restricted to quintile groupings. Further research is required to assess whether this study's finding – that responses to inequality changes do not vary significantly by income level – can be applied to these extremely rich families, executive compensation committees, and the like.[10]

The next section visually presents the international findings before turning to a regression framework for detailed results. Section 4.3 then explores regional variation in the United States. The overall inequality metrics (i.e., gini, 80-20 income percentile differential) used for the United States study are also disaggregated into measures for the upper and lower halves of the income distribution (i.e., 80-50 and 50-20 differentials), tentatively finding changes in the lower half to be more significant for explaining shifts in social norms. Section 4.4 refines the United States findings through an instrumental-variable specification combining exogenous changes in the federal minimum wage with predetermined regional characteristics. Finally, Section 4.5 explores whether responses differ by income level or neighborhood racial heterogeneity. Section 4.6 concludes the paper with a further discussion of this developing literature strand and directions for future research.

4.2 International Evidence

The international portion of this study focuses on how social attitudes towards redistribution respond to changes in national income inequality. Evidence is drawn from the International Social Survey Programme (ISSP) and the World Value Survey (WVS) using fixed-effects estimations that combine repeated opinion surveys with aggregate inequality metrics. The questions taken from both surveys are described, followed by the important construction of the inequality series.

[10] The limited philanthropy of the super-wealthy is frequently criticized (e.g, Economist 1998). Norms for redistribution through non-government channels (e.g., churches, charities) are not considered in this study.

4.2.1 Data Structure

The ISSP conducts annual surveys in member countries (38 nations in 1999) on rotating topics ranging from religion to environmental protection. This study primarily considers questions that were included in the 1987, 1992, and 1999 Social Inequality module. Responses to three complementary questions proxy social norms for government-led income redistribution, the first focusing on the acceptability of current income differences (Inequality Acceptance), the second considering the role of the government in the transfer of income (Government Responsibility), and the last focusing on the progressive nature of taxation (Progressive Taxation). Higher responses on a five-point scale indicate more discontent with current inequality and greater support for government intervention or progressive taxation.

Respondents are also asked their opinions on the appropriate salaries for a variety of occupations. Instructions request preferences be pre-tax and regardless of perceptions of current pay scales. From these responses, a Proposed Unskilled/Doctor Wage Ratio is developed as the log ratio of the wages ascribed for an "unskilled worker in a factory" and a "doctor in general practice." A higher ratio indicates a more-compressed wage distribution (i.e., a log ratio of zero would indicate unskilled workers and doctors should earn the same amount), while a lower ratio indicates support for greater compensation differentials.

Finally, two questions regarding the presence of conflicts between social groups are considered. The first, focusing on conflicts between the poor and the rich (Poor-Rich Conflict), is used to validate respondents' awareness of the inequality in their countries, while a second question regarding conflict between young and old people is considered as a falsification exercise (Young-Old Conflict). A higher score on a four-point scale indicates a greater perception of conflict.

As a complement to the ISSP, responses to a question included in the 1990 and 1995 rounds of the WVS are studied. For this question (WVS Income Equalization) respondents are asked to rate their views regarding income equalization, with a higher score on a ten-point scale expressing greater concern. Table 4.1 details for both surveys the countries included, sample sizes, and average responses to these questions. The Data Appendix describes in detail the wording of each question.

As a final ingredient, this study estimates changes in national income inequality using log gini series constructed from the United Nations Development Programme's World Income Inequality Database (WIID), the Luxembourg Income Study (LIS), Gottschalk and Smeeding (2000), and various national statistics agencies. With a few exceptions, these gini estimates are estimated with national samples of disposable (after-transfers) household income and lagged one year. The Data Appendix details the international series constructed and the techniques employed.[11]

4.2.2 Graphical Analysis

Before considering detailed empirical estimations, it is helpful to discuss visually the main findings of this study. Figure 4.1 plots the mean country responses for four ISSP outcomes against the inequality levels at the time of the surveys. Trend lines indicate higher inequality levels are associated with lower average responses. That is, respondents in more-unequal countries are less likely to feel income differences are too high (Inequality Acceptance) or to assign transfer responsibilities to the government (Government Responsibility); they also propose a wider wage distribution evidenced in the smaller log Proposed Unskilled/Doctor Wage Ratio.

Three notes should be made. First, the negative correlations are not due to respondents being unable to gauge the inequality in their countries. The fourth graph of Poor-Rich Conflict indicates more-unequal societies are more likely to recognize social conflicts exist along income dimensions. Second, the negative correlations are not a product of pooling surveys – the majority of the individual cross-sections also associate higher-inequality areas with reduced concern. Finally, Figure 4.1 highlights that the extreme responses of transition or developing economies may overly influence the findings. To address this concern, Figure 4.2 restricts the sample to long-term OECD members and finds similar results.

The levels patterns evident in the cross-sections, however, do not necessarily dictate the movement of countries over time. Figures 4.3 and 4.4 thus take the next step of plotting how changes in inequality correlate with changes in social norms. The x92 (x99) observations are the

[11] In the visual and empirical work below, gini estimates are preferably lagged one year, although contemporaneous and two-year or three-year lags are accepted when necessary. Data restrictions prevent the use of gross (pre-transfers) family-income inequality estimates, a more theoretically appropriate metric that is less influenced by current and past norms for redistribution. Lagging disposable-income inequality one year allows it to be predetermined in the year of the survey. The United States study later finds that gross-inequality and disposable-inequality estimates yield similar results.

mean changes for country x between 1987-1992 (1992-1999). For both the whole sample and the OECD sub-sample, societies experiencing increases in inequality become more concerned about income differences and assign an increasing responsibility to the government for transferring income. Note, however, that these societies do support an increase in wage dispersion; the empirical estimations below more closely examine the magnitude of this increase. Finally, changes in Poor-Rich Conflict ratings indicate that inequality changes are being perceived.

4.2.3 Empirical Estimations

While important for framing the analysis, the visual correlations fail to control adequately for factors influencing both inequality and social attitudes for redistribution. First, common shifts in attitudes over time (e.g., a greater worldwide concern for inequality not necessarily linked to changes in the inequalities of individual countries) can affect the results. A robust analysis should also control for changes between surveys in national income and demography (e.g., an aging population). Finally, and most importantly, social-mobility experiences and beliefs regarding the sources of success are primary determinants of attitudes toward redistribution. It is important to account for changes in these experiences and perceptions to isolate the role of increasing inequality.

To characterize more rigorously the visual correlations, a series of regressions are estimated with individual responses to the surveys as dependent variables. For simplicity, only least-squares specifications are discussed; ordered-logit specifications that allow for non-linearities in responses yield similar results. The primary estimation equation takes the following form (person i, country c, year t):

$$RESP_{i,c,t} = \alpha + \beta_1 \ln(GINI_{c,t-1}) + \beta_2 \ln(GDP/CAP_{c,t}) + \beta_3 X_{i,c,t} + \phi_c + \eta_t + \epsilon_{i,c,t}, \quad (4.1)$$

The β_1 coefficient is the focus of this study. Survey responses are ordered so that a positive β_1 coefficient reflects a more-concerned position: greater concern for inequality, more support for government intervention, a more-compressed wage structure, etc. The log GDP per capita controls for national wealth at the time of the survey. The $X_{i,c,t}$ vector of covariates includes personal demographics and responses to social-mobility questions as controls. A vector of country effects ϕ_c control for systematic level differences among countries, while a vector of year effects η_t absorb systematic differences in responses among surveys.

Table 4.2 presents the international results for the β_1 coefficient, with each row representing a separate set of regressions for the ISSP or WVS dependent variable indicated. To conserve space, only the observations for the Government Responsibility regressions are listed, but these are representative for the other ISSP estimations. Variables are transformed to have a mean of zero and a standard deviation of one to aid in interpretation. Thus, the 0.179 coefficient on the gini estimate in the first regression for Government Responsibility indicates a one standard-deviation change in the inequality level is estimated to be partially correlated with a change of about 18% of one standard deviation in survey responses.

The first column of results is for regressions that include only country and year fixed effects.[12] It is clear that the correlations noted earlier are statistically significant, but of a modest magnitude. An increase inequality is met with greater concern for income differences, a heightened role for government intervention, and a desire for a more-progressive tax structure.[13] Statistically significant increases in awareness of social conflict between poor and rich again highlight that changes in inequality are being perceived. In a falsification exercise, Table 4.2 also finds inequality changes are not correlated with changes in awareness of social conflict between young and old people.

Respondents are more likely, however, to propose a wider wage distribution. This is not very surprising as the productive force causing the inequality (e.g., skill-biased technical change) will certainly influence compensation norms. It is important to note, however, that the magnitude of this change is not very negative – that is, norms adjust to support additional inequality, but they fall short of endorsing the full expansion of inequality that is occurring.[14] An unreported

[12]Statistical agencies conducting surveys within each country provide weights for forming nationally representative samples; these weights are further adjusted so that all country-year observations carry the same significance. The results are not sensitive to employing weights or different weighting strategies. Standard errors are clustered on country-year observations.

[13]Levels regressions without country fixed effects also confirm the pooled cross-section correlations of Figures 4.1 and 4.2. Nations with greater inequality have a statistically significant reduced concern for income differences, weaker support for government intervention, and lower desire for a progressive tax structure. While critical for the results, only one other study considering fixed-effect specifications with inequality levels has been identified. Alesina, Di Tella, and MacCulloch (2001) employ a similar strategy in their study of differences in happiness between the United States and Europe. Suhrcke (2001) combines measures of inequality and an indicator dummy for post-socialist countries in a single cross-section from the 1999 ISSP Social Inequality survey.

[14]The ISSP surveys also ask respondents, in addition to proposing wages for various occupations, what they

disaggregation of changes in the Proposed Unskilled/Doctor Wage Ratio finds the expansion to be primarily occurring between doctors and skilled workers rather than skilled workers and unskilled workers.

The second column adds each nation's log GDP per capita to capture movements in the overall wealth of the country, as well as Demographic Controls and Mobility Controls. Demographic Controls include sex, married, age, education, and income dummies. Mobility Controls incorporate respondents' answers to other ISSP and WVS questions that reveal beliefs and experiences regarding social mobility. ISSP regressions include two questions asking respondents to rate the importance of being from a wealthy family or of knowing the right people for getting ahead. Respondents believing these important significantly favor more redistribution. Past mobility experiences are also modeled by respondents' ratings of the status of their jobs compared to their fathers' jobs; respondents believing their jobs are better than their fathers' are significantly less likely to support redistribution. WVS regressions incorporate a question asking respondents to rate whether hard work or luck determines success or failure.

The magnitudes and significance of the coefficients on the gini estimates are robust to including these Demographic and Mobility Controls. Column 3 further shows the results to be robust to including Work Controls of dummies for self-employed, supervisor, unemployed, and a union member.[15] After including these covariates, a one standard-deviation change in inequality now accounts for 20%-25% of a standard-deviation change in responses for most ISSP variables. Note, however, that the coefficients in the WVS regressions suggest a substantially higher explanatory power of 40%-60%. The higher percentage of developing countries in the WVS sample likely plays a role in these larger partial correlations. Also, the larger estimates may be the product of offering respondents ten choices rather than five, making it easier to

think those occupations actually earn. Regressions of proposed wage ratios on perceived wage ratios yield significant coefficients but with magnitude less than one-for-one. That is, respondents with a higher perception of current wage inequality do propose a wider wage structure, but not as wide as the inequality they perceive.

[15] The coefficients on the Demographic and Work Controls follow the patterns found in previous cross-sectional studies and are not reported here (e.g., Suhckre 2001, Alesina and La Ferrara 2001). As the quality of income data varies substantially across surveys and countries, respondents are grouped into family-income quintiles for each survey year. Support for redistribution declines with income; support also tends to be lower among male, older, and more-educated respondents. Self-employed workers and supervisors tend to have less support for redistribution, while unemployed workers and union members are more supportive. While reasonable, the direction of these findings should be treated with caution as income variation not captured by the quintile groupings may be loading onto other demographic and work characteristics. Finally, race/minority status is not included in the demographics; later results indicate this is an important factor for the United States.

capture shifts in attitude.[16]

As discussed earlier, poorer and transitional countries appear visually to possess substantially higher support for redistribution than their OECD counterparts with similar levels of inequality.[17] Moreover, they demonstrate significant changes in attitudes and inequality levels that dwarf the more-stable advanced nations. To ensure the sample composition between OECD and non-OECD countries is not driving the results, Column 4 includes Year x OECD dummies. Likewise, the fifth column incorporates Year x Transition Economy dummies. With the exception of the ISSP Progressive Taxation variable, the significance levels of the gini estimates are robust to forcing the variation into the subgroups. Column 6 also shows the results are robust to substituting a time trend for the year dummies.

While the results in Table 4.2 are for regressions employing only responses to the ISSP Social Inequality module, the Government Responsibility and Progressive Taxation questions are also included in the 1985, 1990, and 1995 Role of the Government module. A longer panel can be constructed that combines surveys from these two modules. While the panel enjoys more countries and higher-frequency variation in macroeconomic conditions, it unfortunately does not afford the inclusion of the important Mobility Controls. The findings from this longer panel (presented in an earlier version of this paper) mirror those in Table 4.2, with the positive coefficient for the Progressive Taxation question more robust to forcing the variation into the OECD and Transition Economy subgroups. A second version of the Government Responsibility question is also included in the Role of the Government surveys and the 1991 and 1998 ISSP Religion module. Results from this third panel are also consistent with those presented in Table 4.2. The stability of the findings through shifting time intervals and countries surveyed speaks to the robustness of the redistribution response.

4.2.4 Discussion of Results and Identification

The findings of Table 4.2 suggest a one standard-deviation increase in inequality is partially correlated with about 20% of a standard-deviation change in social norms. In words, it suggests that increases in inequality are met with a greater concern for inequality and a stronger

[16]It is also possible that the relative wording of the WVS question is responsible for these higher coefficients (see the Data Appendix).

[17]Austen (1999) and Suhckre (2001).

desire for government-led redistribution, but this increased concern is modest and reflects small movements around the long-term levels of the countries. Taking the United States as a specific example, a 20% response would be sufficient to achieve the average responses of other Anglo-Saxon countries (e.g., Canada, Australia, and Great Britain), but would fall short of the levels of Continental Europe and especially transition economies.

A causal interpretation for these results is reasonable, although not assured. Two basic concerns are the endogenous relationship between inequality and norms (i.e., that norms also influence the inequality levels) and omitted-variable biases. In addition to the lagging of inequality one period, the direction of the results suggests that the reverse-causality concern is weak. It could not have been the case that changes in social norms to favor more income equalization produced increases in inequality, while it is very reasonable that increased inequality led to greater support for redistribution. Employing disposable-income inequalities rather than gross-income inequalities may affect the coefficient magnitudes slightly, but will not change the direction of the findings.

It may be possible, however, to argue an omitted factor prompted both the increases in inequality and the changes in social norms. For example, an increased openness to trade may have raised inequality and also increased desire for government income stabilization out of fear of globalization (and unrelated to the change in inequality itself). The consistent results of higher inequality being associated with higher concern over disparities and increased conflict between the poor and rich, however, suggest that the most-plausible interpretation is the increased inequality acted directly on social norms. A more-rigorous instrument strategy employed with the United States data will also support this interpretation.[18]

4.3 United States Evidence

To complement and verify the international findings, regional variation in inequality and support for income equalization from the United States is explored next. In addition to being a

[18] It is, however, quite possible that weakening concern over compensation differentials contributes to rising wage inequality (reverse causality), particularly the significant increase in executive salaries. Unfortunately, the United States survey employed in the next section does not contain wage differential questions like the ISSP. Thus, while the instrumentation will be able to assign causality for the general redistribution result, this study is limited to partial correlations regarding rising inequality and compensation differentials.

check for the earlier results, this study is of interest for three reasons. First, while national inequality would be the most-perceived dimension for smaller countries such as Bulgaria or Ireland, regional differences may be more important for large nations that display significant heterogeneity in economic activity. Moreover, a substantial fraction of policy and budget decisions in the United States are made at the state or city level, with officials accountable to their local constituents. Finally, but certainly not least from a research perspective, the quality and quantity of United States data afford extensions and instruments that are not possible in international studies.

4.3.1 Data Structure

United States social norms are estimated from the General Social Survey (GSS), which has been conducted on an annual or biennial basis since 1972 with sample sizes ranging from 1400 to 3000 adults. The analysis considers four questions. The first question asks on a three-point scale whether the United States should be spending more or less money on welfare (Welfare Spending); an identical question regarding spending for the space exploration program (Space Exploration Program Spending) is also considered as a falsification exercise similar to the conflict between the young and old question in the international study. A third question (GSS Income Equalization) documents respondent support on a seven-point scale for the federal government's reduction of income differences between the rich and the poor. A fourth question surveying political-party affiliation is described below. Responses are again ordered so that higher values correspond to higher support for the reduction of inequality.

An important criticism of studies employing opinion polls is that they may be capturing only cheap talk – that is, respondents are willing to say redistribution should be higher, but they do not expect the government to take serious action and do not change their own behavior. There are a number of ways to substantiate that norms regarding redistribution do matter. Luttmer (2001), for instance, demonstrates that over 30% of the variation in state welfare-benefit levels can be explained through an interaction of attitudes towards welfare with state demographic compositions. He also considers how norms for redistribution modeled with the GSS mirror voting patterns in a California proposition.

Keeping the analysis focused on the GSS survey, this project instead considers how shifts in reported political-party affiliation correlate with changing inequality levels. Respondents are asked to state their party preference and the strength of this association on a seven-point scale (Party Identification), with one being strongly Republican and seven being strongly Democrat. Of course, many other factors influence party affiliation, and the platforms of parties demonstrate temporal and regional variation. Nevertheless, it is reasonable to portray the Democratic Party over the last three decades as supporting higher levels of transfer from the United States' wealthy classes to its poorer classes than the Republican Party. Regressions with this question study whether higher inequality is associated with changes in political affiliation, in addition to changes in support for welfare programs. The Data Appendix details the wording of these four questions.

The final requirements for the United States analyses are the important inequality metrics. The richness of United States data offers additional flexibility, and two metrics of overall inequality are considered. Modeling inequality with regional log gini estimates affords comparisons to the earlier international work. The detailed data also allow consideration of inequality trends for different parts of the income distribution. Thus, overall inequality is additionally modeled as the differential between the log 80th and 20th percentiles. After considering overall inequality, the 80-20 differential is disaggregated into the changes in inequality in the upper and lower halves of the distribution (i.e, the 80-50 and 50-20 differentials). Inequality estimates in this section are calculated over disposable family income for the four primary Census regions (i.e., Northeast, Midwest, South, and West) from the March Current Population Surveys (CPS); in the next section these results will be shown representative of other income definitions (e.g., pre-tax family labor earnings, hourly wage) and lower levels of regional aggregation (e.g., nine Census regions, states).[19]

[19] Three levels of geographic aggregation and three forms of inequality are considered for the United States. On the geographic dimension, inequality estimates for Census regions (four or nine) are calculated from the March CPS files. These annual measures are preferred since decade-based measurements can miss important fluctuations (most noticeably the significant expansion in family-income inequality during the recessions of the early 1980s and 1990s). The sample sizes of the March CPS are insufficient, however, for state-level analyses (and states are not identified until 1977). State-level statistics are instead calculated from the Census for each decade, with standard errors clustered at the decade level.

Three income definitions are considered: post-tax disposable family income from all sources, pre-tax family labor earnings, and hourly wages. The first two family measures are calculated over family equivalents using Danziger and Gottschalk's (1995) procedure of dividing by an inflation-adjusted poverty-line estimate for a family of similar composition (i.e., the number and ages of adults and children in the family unit). Additional procedures for preparing the sample (e.g., the exclusion of military families, adjustment of top-codes) follow the common

4.3.2 Discussion of Identification

It is important to highlight some identification issues for the United States findings before discussing the empirical results. While a motivation for this exercise is to explore whether regional inequalities matter more for social norms than national trends in large countries, the data suggest they are in fact second-order for the United States. Figure 4.5 plots the mean response to the GSS Welfare Spending question and the 80-20 income differential for each region by year.[20]

Two features of this graph deserve inspection. First, some differences in regional inequality exist (the solid line). While the South begins with significantly higher inequality than the other regions in the early 1970s, the strong growth in inequality in the Northeast and West results in the three regions being approximately equal by the late 1990s. The Midwest, while also experiencing an increase in inequality, remains significantly lower than the South throughout the period. Unlike the international analysis, however, none of the regions experience a period of substantial decline in inequality. Thus the inference is from stable inequality or increases in inequality. Second, the dramatic swings in the mid-1970s and 1990s highlight that regional variation in welfare support is second-order to large national shifts, likely due to political swings.[21]

The national trends in inequality and social norms are absorbed by the year effects, while systematic levels differences between regions are controlled for by geographic fixed effects. Given the importance of these national elements, the regression coefficients for the regional variation should be smaller than those captured in the international estimations.[22]

---

practices outlined in Danzinger and Gottschalk (1995); Autor, Katz, and Krueger (1998); and Katz and Autor (1999).

In each analysis, the regional fixed effects, median income levels, and standard-error clustering are adjusted to the appropriate geographic aggregation; median income levels are additionally adjusted to reflect the income definition used in the inequality calculation. The Data Appendix reports the regional disposable-income 80-20 differential estimates employed in the primary regressions.

[20]While representative, the mean regional responses should be treated with caution. The sampling design of the GSS results in certain states or metropolitan areas with distinct differences in social norms from their surrounding region entering and leaving the survey (e.g., the more-religious Utah in the West). While the regression results control for these shifts, the regional mean responses do not.

[21]The significant decline in support in the mid-1970s is linked to the explosion in welfare caseloads in the prior decade (e.g., Moffitt, Ribar, and Wilhelm 1998), while the large dip in the mid-1990s surrounds the 1994 Republican Revolution during Clinton's first term. The close co-movement of regional inequality and Welfare Spending norms between these periods is quite striking.

[22]To conserve space, a graphical analysis similar to that presented for the international evidence is omitted.

4.3.3 Empirical Estimations

Table 4.3A considers a set of specifications similar to the international regressions studied in Table 4.2; Table 4.3B replaces the log gini inequality metrics with log 80-20 income differentials that are employed throughout the remainder of this section. Column 1 of both specifications finds changes in regional inequality partially correlate with a statistically significant increase in support for all three norms when only year and region fixed effects are included. As expected, the coefficients are smaller than those found in the international regressions, as the regional variation is second-order to national trends. As a falsification exercise for Welfare Spending, no significant correlation is registered for Space Exploration Program Spending.

As before, Columns 2 and 3 further show the magnitudes and statistical significance of the coefficients are robust to including the regional median income (akin to the national GDP per capita) and Demographic Controls, Mobility Controls, and Work Controls. Unfortunately, incorporating many GSS social-mobility variables severely limits the sample size; the regressions only include a question that asks whether the financial position of a respondent's family has improved, worsened, or stayed the same over the last few years.[23] The GSS does, however, collect race data. Non-white respondents are found in the fourth column to have significantly higher support for redistribution, even after including income levels and the other Demographic Controls. The coefficients for Welfare Spending and Party Identification remain of similar size and significance, but those for Income Equalization diminish.

The last two columns offer some robustness checks. Excluding the South in Column 5 affects the significance of several estimates, but the shifts are sporadic. The seventeen states defined as the South comprise about a third of all GSS respondents, and it is not too surprising that the smaller sample size influences several estimations. The final regression includes a time

While the same patterns generally exist, they are weaker (as the upcoming regressions quantify). In the pooled cross-section, regions with higher inequality exhibit reduced support for income equalization and more right-winged politics, although a levels correlation with greater support for welfare spending is evident. While most individual ISSP cross-sections also display the negative relationships evident in the pooled graphs, the substantial national shifts in GSS responses do produce some positive trends when examining smaller time intervals. As in the international presentation, increases in inequality visually correlate with greater support for income equalization and welfare spending; the mean party affiliation also shifts towards the Democratic Party. Employing disaggregated inequality metrics does not affect the direction of these responses.

[23] Demographic surveys often find respondents over-estimate their relative financial position (e.g., Brooks 2002). In addition to actual incomes, the GSS collects respondents' perceptions of their incomes compared to the national average. The results are robust to using these perceptions rather than actual income levels.

trend. The coefficients on all the variables shift substantially, but the large increase for Welfare Spending is particularly noticeable. As Figure 4.5 highlighted, the norms series, and to some extent inequality, exhibit significant, non-linear national shifts. Replacing the year effects with a linear trend allows more of this variation to load onto the regional inequalities.

A significant concern about the analysis thus far is that gini estimates only measure overall inequality. A detailed exploration should further identify the subsets of the income distribution that are most important for changes in social norms. While more-disaggregated international statistics are very rare and typically of poor quality, United States data are available. Table 4.4 decomposes the 80-20 inequality into the 80-50 and 50-20 differentials. The results suggest that trends in inequality in the lower half of the distribution (i.e., the poor being increasingly left behind) are most responsible for the aggregate results previously identified for the United States.[24,25]

4.4 Minimum-Wage Instrument

United States regional estimations agree with the earlier international results: increases in inequality partially correlate with increases in desire for government-led redistribution. In addition to finding this effect on two levels, it was earlier noted that the direction of the results, the lagging of inequality, and the significance of survey questions focused on inequality itself suggest a causal interpretation is reasonable, although still not assured. In this section, an instrument designed for the United States regional variation further undergirds this claim.

In recent empirical studies, labor economists note the role of the minimum wage in rising United States inequality, especially during the 1979-1989 period when the real (i.e., inflation-

[24]Using 90-50 and 50-10 trends, which demonstrate less co-movement than the 80-50 and 50-20 series, yields significant results for the 50-10 ratio in all regressions (including Income Equalization). It should be noted, however, that a rigorous characterization of the relative contributions of inequality in the two halves of the income distribution requires separate instruments be designed for each portion. This is left for future research. These results should not be applied to the earlier international findings, especially the compensation differential metrics between doctors and factory workers.

[25]Moffitt, Ribar, and Wilhelm (1998) find evidence that declining welfare-benefit levels can be linked to declining low-skill wages, as voters seek to maintain a target benefit-wage ratio (perhaps to preserve equity between working and non-working poor or to minimize employment disincentives). The disaggregated income inequality results (in particular, the positive and significant coefficient on the 50-20 ratio) are robust to including measures of the 15th or 25th percentile wages.

adjusted) value of the federal rate declined by 24%.[26] While these substantial swings in mandated federal rates can be taken as exogenous from the perspective of individual states or regions, they do not provide the necessary regional variation by themselves. A credible instrument can be designed, however, through the interaction of these national trends with predetermined regional characteristics that govern how important minimum-wage mandates are for the local economy. The year effects absorb the national dynamics of the changing federal rate, and the pre-existing regional traits are controlled for by the geographic fixed effects. The identifying assumption is that the residual region-year interactions can serve as an instrument for the region-year inequality trends (which are themselves also subject to the fixed effects).

This study employs regional coverage ratios, defined as the percent of the working population protected by the minimum-wage statutes, as its interaction term. Regions differ in the composition of their economic activity, and the federal minimum-wage mandates are not applied equally to industries (e.g., 1970-2000 coverage rates in agriculture averaged 41% versus manufacturing's 97%). The larger the fraction of a region's population covered by the federal statutes, the more impact federal rates have on the local economy. The simplest interaction term would be the 1970 coverage rate; in a slight design improvement, the interaction term is built instead as the expected coverage in year t for each region. This modification allows incorporation of trends in national coverage rates due to changing federal legislation (especially in the mid 1970s), thereby raising the quality of the first-stage estimations. The inequality instrument for region r and year t takes the form[27]

$$INEQIV_{r,t} = \ln(FED_{1970}/FED_t) \cdot E_{1970}COV_{r,t},$$

where

$$E_{1970}COV_{r,t} = 1 - \textstyle\sum_j IND\%_{j,r,1970} \cdot (COV_{j,1970}/COV_{j,t}),$$

with j indexing industries. The two parts of this interaction deserve careful explanation. The construction of the second element $E_{1970}COV_{r,t}$ is the more complicated. It is the expected coverage rate in region r for year t, estimated from the 1970 industrial composition of the

[26] DiNardo, Fortin, and Lemieux (1996); Lee (1999); and Golan, Perloff, and Wu (2001).

[27] Recall that inequality is lagged one year in the estimations; the instrument will be lagged as well.

working poor and changes in national coverage rates by industry. $IND\%_{j,r,1970}$ is the percent of a region's workforce from the 1970 Census who are both earning less than the minimum wage and working in industry j. By itself, $\sum_j IND\%_{j,r,1970}$ would produce the actual percentage of the region's working population earning less than the federal minimum wage in 1970. $COV_{j,1970}/COV_{j,t}$ is the ratio of the national coverage rate for industry j in 1970 to that in year t. From a starting value of one, the ratio moves above (below) one for industries where the coverage rates decrease (increase) compared to 1970 levels.[28]

The combination of these terms is the expected percentage of a region's workforce earning below the minimum wage in year t. The starting 1970 level of $\sum_j IND\%_{j,r,1970} \cdot (COV_{j,1970}/COV_{j,t})$ is still the actual workforce percentage earning below the 1970 federal rate in each region (as the coverage ratio for all industries is one). For subsequent years, it is expected that the percentage of the population earning below the minimum wage will decline in region r if its poor workers were primarily employed in industries where the coverage rate later increased. On the other hand, little change is expected in states or regions where very few workers were initially below the minimum wage or where the poor worked in industries for which the coverage rate did not change significantly. Finally, $1 - \sum_j IND\%_{j,r,1970} \cdot (COV_{j,1970}/COV_{j,t})$ estimates the percent of the population covered by the minimum-wage mandates and thus the potential importance of changes in the federal rate for the region's inequality level.

Turning to the first term, $\ln(FED_{1970}/FED_t)$, the log ratio of the real federal minimum-wage rate in 1970 to the rate in year t takes an initial value at zero for 1970. In years when the real federal rate is greater (less) than the real federal rate for 1970, this component of the instrument has a negative (positive) value. Note that some states have mandated minimum wages that exceed the federal rate. These are not considered as the local legislation could clearly be endogenous to the inequality levels. The Data Appendix provides descriptive statistics for these two components of the instrument.

The instrument is then the interaction of shifts in the real federal rate with the expected

[28] Coverage rates are at the one-digit SIC level and exclude government employees (e.g., Nordlund 1997, United States Department of Labor 1998). Coverage rates have not been identified for 1989 or after 1996. For the main estimations, a linear interpolation is employed for 1989 and observations post-1997 are assigned 1996 values; the results are robust to dropping these missing years. Unfortunately, the coverage data are not disaggregated to where each observation's own region could be excluded. As Data Appendix shows, the expected coverage rate calculations produce only a slight trend vis-à-vis fixed 1970 levels.

coverage level, or how much the federal legislation matters for a region. Note again that the instrument comes only from the interaction between these two elements. The individual trends of the real federal rate and industry coverage rates are absorbed by the year effects. Geographic fixed effects control for the region's predetermined industrial composition of poor workers. Note too that the instrument does not have a level *per se* – its value for all regions is zero when the real federal rate is equal to its 1970 level (i.e., 1970 itself, approximately so in 1975/1976 and 1981). It relies on the region fixed effects to control for the mean inequality positions of each area. Finally, the instrument is designed to have a positive first-stage coefficient. The $E_{1970}COV_{r,t}$ term is always positive and only governs the magnitude of the response; the $\ln(FED_{1970}/FED_t)$ component is positive when the current federal rate is below its 1970 level, which should correspond to rising inequality, and vice versa.[29]

Table 4.5 presents the detailed results of the instrumental-variable specifications for the log 80-20 differential.[30] The first-stage results are presented for the Welfare Spending sample; the positive coefficients and R^2 values are reflective of the samples of the other dependent variables. Figure 4.6 plots for each region the residual trends (i.e., after year and geographic fixed effects are removed) for the minimum-wage instrument (the solid line) and the inequality level (the line with circles). The expected first-stage relationship is apparent within each region.

The second-stage results confirm the least-square specifications discussed earlier; a one

[29]The robustness of the instrument design has been verified on several dimensions. First, the results are mostly robust to simply fixing the coverage rate at its 1970 level for each region; the only trouble spot is in regressions that contain only year and region fixed effects, as the simpler interaction captures some of the median-income level trend when it is excluded. Second, the total industrial composition of the region can be substituted for the industrial composition of the poor workers. Finally, as noted above, the instrument incorporates two aggregate trends – changes in the federal rate and changes in industry coverage rates. Close observation shows the instrument can work against itself. Focusing on movements in the minimum-wage level, the instrument correctly predicts regions with higher coverage levels will be more affected by federal changes. Yet, over a short horizon and holding the minimum wage fixed, the instrument incorrectly predicts an increase in the coverage rate will raise inequality if the real federal rate is below its 1970 level; its predicted direction is correct if the real federal rate is above its 1970 level. An alternative specification removes the competing effects by using two instruments, one interacting the dynamics of the federal rate with fixed 1970 coverage rates and the second interacting industry coverage rate trends with the 1970 industrial composition. The results are again very close to those presented in the main text.

[30]The battery of regressions is similar to Table 4.3, although the time-trend specification is dropped since the instrument design requires year fixed effects be included. The instrument specifications are robust to using other forms of aggregate inequality (e.g., gini, 90-10, entropy). Estimations employing only the 50-20 differential also yield similar results, but excluding inequality in the upper half of the distribution may create a bias since some workers in high-income families are affected by minimum-wage legislation (Card and Krueger 1995). While first-stage coefficients for the 80-20 or 50-20 specifications are almost always positive and highly significant, potential first-stage coefficients for the 80-50 inequality are of mixed sign and significance.

standard-deviation increase in inequality is now found to produce 10%-20% of a standard-deviation shift in support for government-led redistribution. Table 4.6 concludes by replicating the Column 2 regressions of Tables 4.3 and 4.5 (i.e., estimations including median income levels, Demographic Controls, and Mobility Controls) across three levels of geographic aggregation and three income definitions. The first three columns are for least-squares regressions, while the last three columns are for instrumental-variable specifications.[31] The least-squares permutations are well-behaved and generally indicate a moderate decline in coefficient size as specifications move away from disposable family income towards the hourly wage definition. The declining coefficient sizes with lower levels of geographic aggregation mirror the earlier coefficient reduction from the international regressions to the four Census regions variation. However, these two trends are weaker in the instrumental-variable permutations. While larger standard errors are evident in some state-level or hourly wage specifications, the instrumental-variable results in general are robust across these dimensions.

4.5 Income and Neighborhood-Heterogeneity Extensions

This final section extends the United States analysis to consider whether the average increase in support for redistribution with rising inequality masks differences among income classes.[32] While the demographic characteristics of respondents are statistically significant for explaining survey answers, Piketty (1996, 1999) notes the overall level of disagreement within a country about distributive equality is usually small vis-à-vis other social issues (e.g., death penalty). Section 4.2 found, however, that perception of conflict between the poor and the rich increases with rising inequality, and it is important to clarify if the average response belies increasing disagreement among classes about appropriate redistribution levels. The rich may become more protective of their wealth as the gap grows, perhaps out of concern over larger transfers or perhaps out of reduced fear that they too may one day be poor. Altruistic motives, however,

[31] The regional specifications are annual and derived from the March CPS while the state specifications are at the decade level and derived from the Census.

[32] An earlier version of this paper replicated this analysis for the ISSP and WVS panels, finding similar results for the income-quintile interactions. These results are available upon request. Neighborhood racial heterogeneity is not available for study at the international level.

may yield greater assistance from the wealthy as disparity widens.[33]

Exploring this issue, Table 4.7 presents three least-squares regressions for the United States norms studied. The first regression of each triplet simply includes Demographic Controls, Mobility Controls, and Racial Controls (i.e., a replication of Column 4 from Table 4.3). The second regression interacts the 80-20 differential with whether respondents are in the top-two income quintiles or the bottom-two income quintiles. The estimations do not find significant differences by class for the Welfare Spending or Income Equalization variables. Respondents in the bottom-two quintiles are more likely to align themselves with the Democratic Party as inequalities in their regions increase. This result is not robust, however, to interacting a time trend with being in the upper-two or lower-two income quintiles.[34]

The third regression of each trio interacts the 80-20 differential with whether the respondent lives near someone of the opposite race (and also adds a main effect for being in a heterogeneous neighborhood). Luttmer (2001) finds support for welfare spending increases as the share of local recipients from a respondent's racial group rises. Lind (2003) also finds aggregate evidence that inequality between racial groups versus inequality within racial groups can have opposite effects for redistribution outcomes. The interacted coefficient for the Welfare Spending regression (but not Income Equalization) agrees with these studies – the increase in redistribution support associated with rising inequality is diminished in racially heterogeneous neighborhoods.[35]

The results of this section suggest changes in support for government-led redistribution are fairly uniform across income groups. The data do not support hypotheses of rising class warfare as inequality increases. This finding is in agreement with Rawlsian models like Piketty (1995), where different classes have similar views on distributive equality holding fixed beliefs about incentive costs. A limitation to these findings, however, is important to note. Piketty and

[33]The Economist (1998, 2001, 2003) provides several discussions of the class-warfare issue and its relationship to redistribution and the American political landscape.

[34]McCarty, Poole, and Rosenthal (2003) note increases in United States inequality have moved in tandem with stronger ideological differences over redistribution and more-polarized party politics. While income has become a stronger predictor of party affiliation over the last twenty-five years, their work also suggests inequality bears limited responsibility for the polarization.

[35]The significant coefficient on the Party Identification interaction (i.e., that respondents are significantly less likely to lean towards the Democratic Party as inequality increases in their region if they live in a racially heterogeneous community) should again be treated with caution. It is a product of the increasing popularity of the Republican Party in the South during this period; dropping this region from the regression yields an insignificant interaction

Saez (2003) find a tremendous increase in the concentration of wealth among the very rich in the United States (i.e., the top 1% and even smaller fractions). Unfortunately, the data cannot be used for an analysis for these super-wealthy individuals and their disproportionate influence (e.g., Krugman 2002).

4.6 Conclusions

This study characterizes how changes in inequality affect social attitudes for government-led redistribution and compensation differentials. Market-based factors have substantially increased inequality in the United States over the last three decades. If the inequality caused by these mechanisms reduces social norms regarding distributive equality, the inequality can become amplified and entrenched. While international and United States regional cross-sections often display a strong, negative correlation between inequality and support for redistribution, this study finds countries and states do not evolve along this pattern in the short-run.

Controlling for initial positions and respondent views of social mobility, local changes in inequality are positively and significantly correlated with changes in support for government-led redistribution. Acceptance of wage disparity does increase with higher inequality levels, but the response is less than one-for-one. While greater class conflict is perceived along income dimensions, the increases in support for redistribution among wealthy individuals are as strong as those of poorer individuals. To the extent the forces driving inequality also alter the underlying beliefs (e.g., determinants of success, mobility experiences, incentive costs) most important for determining the long-term tradeoff between inequality and redistribution norms, then these forces may contribute to reduced concern over the disparity. The conclusion of this study, however, is that the increase in inequality itself is insufficient for weakened social norms for equality.

Several important areas for future research exist. Political economists have long studied reasons for the negative cross-sectional relationship between inequality and support for redistribution; this study explored localized movements around these long-run positions. Recent theoretical research considers endogenous shifts in long-term positions[36]; as more data become

[36] Benabou (2002) and Hassler *et. al.* (2003).

available, future research should empirically test these longer-term dynamics. Such shifts will further clarify the primal factors behind cross-sectional differences, highlight whether the concerned responses noted here are governed by important thresholds or critical-mass points, and identify mechanisms beyond ideology that can contribute to the formation of vicious cycles.

It is also important to characterize the channels through which inequality and norms interact. For instance, increasing social stratification[37] may amplify or diminish the direct effect of increasing inequality on social norms. Alesina and La Ferrara (2000) report greater inequality is particularly correlated with reduced membership in church and service groups, activities often associated with assisting the less fortunate. This deterioration of civic bonds may weaken support for redistribution. On the other hand, Luttmer (2001) argues free-rider concerns likely reduce support for welfare policies, and perhaps these concerns are weakened in more-segmented communities. It is also unclear how the non-pecuniary status desires that can limit support for redistribution change in a more-stratified society.[38] A better understanding of how stratification and other channels facilitate the interaction of inequality and norms will afford more-causal assessments and aid in policy recommendations.

Finally, and most importantly, future research should trace how different political systems (including such diverse issues as government structure, campaign financing laws, voter participation, etc.) govern the translation of changes in social norms into policy outcomes. Recent research notes in particular the importance of franchising groups favoring higher redistribution and the disproportionate influence of elites.[39] The adoption of more-conservative redistribution policies in several Anglo-Saxon countries during periods of rising inequality suggests this issue is a primary concern.[40] How political systems are structured will govern whether rising latent concerns for redistribution produce higher effective support to which politicians are held accountable.

[37] Putman (2000); Benabou (1993, 1996); and Bertrand, Luttmer, and Mullainathan (2000).

[38] Corneo and Gruner (2000, 2002).

[39] Husted and Kenny (1997) and Lott and Kenny (1999).

[40] Caminada and Goudswaard (2001) and Hassler et. al. (2003). Interestingly, little correlation exists at the state level between inequality and Democrat vote percentages in United States Presidential elections (Rodriguez 1999). Whether uneven declines in voter participation can reconcile the findings of this study with the aggregate outcomes is being explored in current research.

4.7 Data Appendix

4.7.1 International Opinion Polls (ISSP and WVS)

The international exercises employ the International Social Survey Programme (ISSP) and the World Value Survey (WVS). To maintain a consistent presentation across international and United States surveys, responses are ordered such that more-concerned views are associated with higher numbers.

The ISSP analysis focuses on the 1987, 1992, and 1999 Social Inequality module; the Government Responsibility and Progressive Taxation questions are also included in the 1985, 1990, and 1996 Role of the Government module. Responses to three complementary questions proxy social norms for government-led income redistribution: the first focusing on the acceptability of current income differences, the second considering the role of the government in the transfer of income, and the last focusing on progressive taxation:

Q. (Inequality Acceptance) "Are differences in income in <Respondent's country> too large?"

1. Disagree strongly
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Agree strongly

Q. (Government Redistribution) "It is the responsibility of the government to reduce the differences in income between people with high incomes and those with low incomes."

1. Disagree strongly
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Agree strongly

Q. (Progressive Taxation) "Do you think that people with high incomes should pay a larger share of their income in taxes than those with low incomes, the same share, or a smaller share?"

1. Much smaller share
2. Smaller
3. The same share
4. Larger
5. Much larger share

Three important characteristics of these questions should be noted. They shy away from sensitive wording (e.g., words like "welfare" carry negative connotations) and they offer respondents a range of options that include a neutral stance. The Government Redistribution and Progressive Taxation questions also do not reference a country's current policy position (e.g., "do you think the government should be doing more to reduce the differences..."). Such relative questions are more difficult to evaluate in panel exercises.

Respondents are also asked their opinions on the appropriate salaries for a variety of occupations. Instructions request preferences be pre-tax and regardless of perceptions of current pay scales. From these responses, a Proposed Unskilled/Doctor Wage Ratio is developed as the log ratio of the wages ascribed for an "unskilled worker in a factory" and a "doctor in general practice." A higher ratio indicates a more-compressed wage distribution (i.e., a ratio of one would indicate unskilled workers and doctors should earn the same amount), while a lower ratio indicates support for greater compensation differentials.[41]

Finally, two questions regarding the presence of conflicts between social groups are employed. The first focuses on conflicts between the poor and the rich to validate respondents' awareness of the inequality in their countries, while a second question regarding conflict between young and old people is considered as a falsification exercise.

Q. (Poor-Rich Conflict) "In all countries there are differences or even conflicts

[41] Other occupations present in all three Social Inequality surveys include a skilled factory worker, a government minister, and a chairman of a large national company. When discussing compensation differentials, the text also describes the evolution of the wage premiums between skilled workers and unskilled workers or doctors. An unfortunate top-coding change in the 1999 survey restricts analysis of the proposed chairman salary, although rough estimations not correcting for this top-coding yield results comparable to those presented. This top-coding also has the potential to affect the doctor wage rate; regressions excluding the 1999 survey demonstrated similar outcomes to primary panel.

between different social groups. In your opinion, in <R's country> how much conflict is there between poor people and rich people?"

1. No conflicts
2. Not very strong conflicts
3. Strong conflicts
4. Very strong conflicts

Q. (Young-Old Conflict) "... between young people and older people?"

1. No conflicts
2. Not very strong conflicts
3. Strong conflicts
4. Very strong conflicts

As a complement to the ISSP, this study also considers responses to a question included in the 1990 and 1995 rounds of the WVS. This question (WVS Income Equalization) asks respondents to rate their views regarding income equalization on a ten-point scale. Ten is labeled, "Incomes should be made more equal." One is labeled, "We need larger income differences as incentives for individual effort." While the WVS panel enjoys a more-diverse group of developing economies, interpretation of this question is limited by its reference to the country's current position (i.e., *more* equal, *larger* differences) and asymmetric labeling of the two extreme values. Only being able to consider one period of change is also a handicap. Nevertheless, finding quantitatively and qualitatively similar results in a different sample is an important robustness check.

4.7.2 International Inequality Series

This subsection details the construction of the international gini estimates employed in the main text. Nations participating in multiple International Social Survey Programme (ISSP) or World Values Survey (WVS) rounds are included, although the former is this study's primary interest. Table 4.A1 outlines the sources and their characteristics (e.g., income definition); data collection relies heavily on the United Nations Development Programme's World Income Inequality Database (WIID), the Luxembourg Income Study (LIS), Gottschalk and Smeeding

(1997, 2000), and the individual publications of national statistics agencies. The WIID includes the earlier work of Deininger and Squire (1996). Table 4.A2 documents the constructed series, with shaded boxes highlighting the years in which countries participated in the ISSP Role of the Government or Social Inequality modules employed in the primary regressions.[42]

The target gini concept is disposable household income based upon a nationally representative sample. Although many sources, including LIS, divide by the square root of the household size, equivalency scales are not consistent across countries. Data limitations prevent consideration of gross household-income inequality, a more theoretically sound measure (although one can argue disposable-income differences are what respondents are recalling when questioned). In the United States portion of this study, the form of inequality (e.g., gross versus disposable household income, household labor earnings, hourly wage) is not critical for the results. A one-year lag in inequality is targeted for each survey round, but contemporaneous and two-year or three-year lagged measures are also accepted when necessary.

Selected series include multiple observations derived with a consistent technique and dataset. Other sources not listed in Table 4.A1 are also used to substantiate both levels and trends of the chosen series, as well as to provide comparisons for how other income concepts are behaving during the same period. In a number of cases, two or three series are pieced together to span the time frame of this study (or as much of it as possible). In such cases, observations must share a common or adjoining year as a levels check; moreover, overlapping intervals are examined when available to ensure the series are following similar trends. Auxiliary series are also employed in these exercises for verification purposes. Finally, the gini estimates are rescaled to match the levels of LIS estimates around 1990 if the LIS is not employed directly in the construction of the series (participating countries only).[43]

Atkinson and Brandolini (2001) outline a number of pitfalls that can occur when piecing together series from secondary datasets. The dataset developed for this study attempts to address these concerns while still assembling a meaningful panel of countries. However, it

[42] The task here is to develop gini series covering the years included in the two survey programs. In doing so, a longer horizon is often considered than what the surveys require for a particular country to establish more confidence in the trends developed. These series, however, do not exhaust the inequality data available; gaps in the sequences do not necessarily mean appropriate gini estimates are not available.

[43] These adjustments produce minor differences between the reported series and source data. The notes column of Table 4.A1 highlights how the LIS is employed with each country if it does not serve directly as a source for the series.

certainly falls short of achieving "double harmonization" across countries and time, and Table 4.A1 identifies questionable series due to poor quality data, alternative income concepts, splicing concerns, and so on. The consistency of the results across the ISSP and WVS samples, dropping low-quality series, and looking at harmonized United States inequalities should nevertheless instill confidence that the findings of this study are not the product of irregularities in the constructed series.

4.7.3 United States Opinion Poll (GSS)

Social norms for the United States are estimated from the General Social Survey (GSS), which has been conducted on an annual or biennial basis since 1972 with sample sizes ranging from 1400 to 3000 adults. This study focuses on two complementary questions that are included for the full term of the survey. The first gauges respondent attitudes towards spending more or less money on welfare, while the second asks a similar question regarding the space exploration program (another falsification exercise):

> Q. (Welfare Spending) "Are we spending too much money, too little money, or about the right amount on welfare?"
>
> 1. Too much
> 2. About right
> 3. Too little

A second question, included in most surveys since 1978, asks respondents to rate on a seven-point scale how much the federal government should concern itself with the income differences between the rich and poor (GSS Income Equalization). Seven is labeled, "The government ought to reduce income differences between the rich and poor." One is labeled, "The government should not concern itself with reducing income differences."

For both the Welfare Spending and GSS Income Equalization questions, alternative versions are included in some years (e.g., substituting "assistance to the poor" for "welfare"). As the mean responses shift significantly with these alternative word choices, these questions are not incorporated; a visual check indicates trends for these alternative questions mirror those of the main questions. It should also be noted that the Welfare Spending question references current policies. Luttmer (2001) considers several corrections for this relative inquiry, and finds his results using the base question alone are robust. This study does not attempt any such corrections.

Finally, respondents since 1972 are asked their political-party preference and the strength of this association on a seven-point scale.

Q. (Party Identification) "Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?"

1. Strong Republican
2. Not very strong Republican
3. Independent, close to Republican
4. Independent (Neither, No Response)
5. Independent, close to Democrat
6. Not very strong Democrat
7. Strong Democrat

4.7.4 United States Inequality Series

Table 4A.3 provides the federal minimum wage ratios and expected regional coverage ratios used to construct the minimum-wage instrument employed in the United States analysis. The last four columns also document the log 80-20 income ratios employed in the primary estimations.

Bibliography

[1] Acemoglu, Daron, Phillipe Aghion, and Giovanni Violante, "Deunionization, Technical Change and Inequality", *Carnegie-Rochester Conference Series on Public Policy* (2001).

[2] Alesina, Alberto, and George-Marios Angeletos, "Fairness and Redistribution", *The American Economic Review* (forthcoming 2004).

[3] Alesina, Alberto, Rafael Di Tella, and Robert MacCulloch, "Inequality and Happiness: Are Europeans and Americans Different?", NBER Working Paper 8198 (2001).

[4] Alesina, Alberto, Edward Glaeser, and Bruce Sacerdote, "Why Doesn't the United States Have a European-Style Welfare State?", *Brookings Papers on Economic Activity*, 2001:2 (2001), 187-277.

[5] Alesina, Alberto, and Eliana La Ferrara, "Participation in Heterogeneous Communities", *The Quarterly Journal of Economics*, 115:3 (2000), 847-904.

[6] Alesina, Alberto, and Eliana La Ferrara, "Preferences for Redistribution in the Land of Opportunities", NBER Working Paper 8267 (2001).

[7] Atkinson, Anthony, "The Changing Distribution of Income: Evidence and Explanations", *German Economic Review*, 1:1 (2000), 3-18.

[8] Atkinson, Anthony, and Andrea Brandolini, "Promise and Pitfalls in the Use of Secondary Data-Sets: Income Inequality in OECD Countries as a Case Study", *Journal of Economic Literature*, 34:3 (2001), 771-799.

[9] Austen, Sioghan, "Norms of Inequality", *Journal of Economic Issues*, 33:2 (1999), 435-442.

[10] Autor, David, Lawrence Katz, and Alan Krueger, "Computing Inequality: Have Computers Changed the Labor Market?", *The Quarterly Journal of Economics*, 113:4 (1998), 1169-1213.

[11] Benabou, Roland, "Workings of a City: Location, Education, and Production", *The Quarterly Journal of Economics*, 108:3 (1993), 619-652.

[12] Benabou, Roland, "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance", *The American Economic Review*, 86:3 (1996), 584-609.

[13] Benabou, Roland, "Inequality and Growth", in Bernanke, Benjamin, and Julio Rotemberg, ed., *NBER Macroeconomic Annual*, 1996, 11-74.

[14] Benabou, Roland, "Unequal Societies: Income Distribution and the Social Contract", *The American Economic Review*, 90:1 (2000), 96-129.

[15] Benabou, Roland, "Human Capital, Technical Change, and the Welfare State", Working Paper (2002).

[16] Benabou, Roland, and Efe Ok, "Social Mobility and the Demand for Redistribution: The POUM Hypothesis", *The Quarterly Journal of Economics*, 116:2 (2001).

[17] Benabou, Roland, and Jean Tirole, "Belief in a Just World and Redistributive Politics", Working Paper (2002).

[18] Berman, Eli, John Bound, and Zvi Griliches, "Changes in the Demand for Skilled Labor Within U.S. Manufacturing: Evidence from the Annual Survey of Manufacturers", *The Quarterly Journal of Economics*, 109 (1994), 367-397.

[19] Bertrand, Marianne, Erzo Luttmer, and Sendhil Mullainathan, "Network Effects and Welfare Culture", *The Quarterly Journal of Economics*, 115:4 (2000), 1019-1055.

[20] Bok, Derek, *The Cost of Talent: How Executives and Professionals Are Paid and How It Affects America* (New York, N.Y.: Free Press, 1993).

[21] Brandolini, Andrea, "The Distribution of Personal Income in Post-War Italy: Source Description, Data Quality, and the Time Pattern of Income Inequality", Banca D'Italia Working Paper 350 (1999).

[22] Brooks, David, "Superiority Complex", *The Atlantic Monthly*, November 2002.

[23] Caminada, Koen, and Kees Goudswaard, "International Trends in Income Inequality and Social Policy", *International Tax and Public Finance*, 8 (2001), 395-415.

[24] Card, David, and Alan Krueger, *Myth and Measurement* (Princeton, N.J.: Princeton University Press, 1995).

[25] Card, David, "The Effect of Unions on Wage Inequality in the U.S. Labor Market", *Industrial and Labor Relations Review*, 54:2 (2001), 296-315.

[26] Card, David, and Thomas Lemieux, "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis", *The Quarterly Journal of Economics*, 116:2 (2001), 705-746.

[27] Corneo, Giacomo, and Hans Peter Gruner, "Individual Preferences for Political Redistribution", *The Journal of Public Economics*, 83 (2002), 81-107.

[28] Corneo, Giacomo, and Hans Peter Gruner, "Social Limits to Redistribution", *The American Economic Review*, 90:5 (2000), 1491-1507.

[29] Danziger, Sheldon, and Peter Gottschalk, *America Unequal* (Cambridge, M.A.: Harvard University Press, 1995).

[30] Deininger, Klaus, and Lyn Squire, "A New Data Set Measuring Income Inequality", *The World Bank Economic Review*, 10:3 (1996), 565-592.

[31] DiNardo, John, Nicole Fortin, and Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach", *Econometrica*, 64:5 (1996), 1001-1044.

[32] Economist, "The Gospel of Wealth", 28 May 1998.

[33] Economist, "Room at the Bottom", 1 January 1998.

[34] Economist, "Survey: Pay", 6 May 1999.

[35] Economist, "Does Inequality Matter?", 14 June 2001.

[36] Economist, "Would You Like Your Class War Shaken or Stirred, Sir?", 6 September 2003.

[37] Fanjul, Gonzalo, and Victor Renes, "Spain Report" (2002).

[38] Frank, Robert, and Philip Cook, *The Winner-Take-All Society* (New York, N.Y.: Free Press, 1995).

[39] Frick, Jaochim, and Markus Grabka, "The Personal Distribution of Income and Imputed Rent – A Cross-National Comparison for the UK, West Germany, and the USA", DIW Berlin Working Paper 271 (2002).

[40] Golan, Amos, Jeffrey Perloff, and Ximing Wu, "Welfare Effects of Minimum Wages and Other Government Policies", Working Paper (2001).

[41] Golan, Amos, Jeffrey Perloff, and Ximing Wu, "Effects of Government Policies on Income Distribution and Welfare", Working Paper (2001).

[42] Goodman, Alissa, "Inequality and Living Standards in Great Britain: Some Facts", Institute for Fiscal Studies Briefing Note 19 (2001).

[43] Gottschalk, Peter, and Timothy Smeeding, "Cross-National Comparisons of Earnings and Income Inequality", *Journal of Economic Literature*, 35:2 (1997), 633-687.

[44] Gottschalk, Peter, and Timothy Smeeding, "Empirical Evidence on Income Inequality in Industrialized Countries", LIS Working Paper 154 (1999).

[45] Gottschalk, Peter, and Timothy Smeeding, "Empirical Evidence on Income Inequality in Industrialized Countries", in Atkinson, Anthony, and Francois Bourguignon, ed., *Handbook of Income Distribution, Volume 1* (Elsevier North-Holland Pergamon, 2000), 261-307.

[46] Hassler, John, Jose Rodriguez, Kjetil Storesletten, and Fabrizio Zilibotti, "The Survival of the Welfare State", *The American Economic Review*, 93:1 (2003).

[47] Hauser, Richard, and Irene Becker, "Changes in the Distribution of Pre-Government and Post-Government Income in Germany 1973-1993", presented to The Personal Distribution of Income in an International Perspective at the Hanse Institute for Advanced Study (July 1999).

[48] Husted, Thomas, and Lawrence Kenny, "The Effect of the Expansion of the Voting Franchise on the Size of Government", *The Journal of Political Economy*, 105:1 (1997), 54-82.

[49] Katz, Lawrence, and David Autor, "Changes in the Wage Structure and Earnings Inequality", in Ashenfelter, Orley, and David Card, ed., *Handbook of Labor Economics, Volume 3* (Elsevier North-Holland Pergamon, 1999), 1463-1555.

[50] Katz, Lawrence, and Kevin Murphy, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors", *The Quarterly Journal of Economics*, 107 (1995), 35-78.

[51] Kremer, Michael, and Eric Maskin, "Wage Inequality and Segregation By Skill", NBER Working Paper 5718 (1996).

[52] Krugman, Paul, "For Richer", *New York Times*, 20 October 2002.

[53] Krusell, Per, Lee Ohanian, Jose-Victor Rios-Rull, and Giovanni Violante, "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis", *Econometrica*, 68:5 (2000), 1029-1053.

[54] Lee, David, "Wage Inequality in the United States During the 1980s: Rising Dispersion or Falling Minimum Wage?", *The Quarterly Journal of Economics*, 114:3, 977-1023.

[55] Levin-Waldman, Oren, *The Case of the Minimum Wage: Competing Policy Models* (Albany, N.Y.: State University of New York Press, 2001).

[56] Lind, Jo Thori, "Fractionalization and the Size of Government", Working Paper (2003).

[57] Lott, John, and Lawrence Kenny, "Did Women's Suffrage Change the Size and Scope of Government?", *The Journal of Political Economy*, 107:6 (1999), 1163-1198.

[58] Luttmer, Erzo, "Group Loyalty and the Taste for Redistribution", *The Journal of Political Economy*, 109:3 (2001), 500-528.

[59] Luxembourg Income Study (LIS) Key Figures, accessed at http://www.lisproject.org/keyfigures.htm on 4 December 2003.

[60] McCarty, Nolan, Keith Poole, and Howard Rosenthal, "Political Polarization and Income Inequality", Working Paper (2003).

[61] Meltzer, Allan, and Scott Richards, "A Rational Theory of the Size of Government", *The Journal of Political Economy*, 89:5 (1981), 914-927.

[62] Moffitt, Robert, David Ribar, and Mark Wilhelm, "The Decline of Welfare Benefits in the U.S.: The Role of Wage Inequality", *The Journal of Public Economics*, 68 (1998), 421-452.

[63] Nordlund, William, *The Quest for a Living Wage: The History of the Federal Minimum Wage Program* (Westport, C.T.: Greenwood Press, 1997).

[64] Ovtcharova, Lilia, "What Kind of Poverty Alleviation Policy Does Russia Need?", Russian-European Centre for Economic Policy Research Paper (2001).

[65] Piketty, Thomas, "Social Mobility and Redistributive Politics", *The Quarterly Journal of Economics*, 110:3 (1995), 551-584.

[66] Piketty, Thomas, "The Politics of Redistribution: Recent Developments and Research Perspectives", presented to the McArthur Foundation Costs of Inequality Project meeting (May 3-5, 1996).

[67] Piketty, Thomas, "Mobilite Economique et Attitudes Politiques Face a la Redistribution", CEPRE Working Paper 9603 (1996).

[68] Piketty, Thomas, "Attitudes Toward Income Inequality in France: Do People Really Disagree?", CEPRE Working Paper 9918 (1999).

[69] Piketty, Thomas, and Emmanuel Saez, "Income Inequality in the United States, 1913-1998", *The Quarterly Journal of Economics*, 118:1 (2003), 1-39.

[70] Pineda, Jose, and Francisco Rodriguez, "The Political Economy of Human Capital Accumulation", Working Paper (2000).

[71] Putnam, Robert, *Bowling Alone: The Collapse and Revival of American Community* (New York, N.Y.: Touchstone 2000).

[72] Rodriguez, Francisco, "Does Distributional Skewness lead to Redistribution?: Evidence from the United States", *Economics and Politics*, 11:2 (1999).

[73] Rosen, Sherwin, "The Economics of Superstars", *The American Economic Review*, 71:5 (1981), 845-858.

[74] Rupnik, Carlo, Margaret Thompson-James, and Ray Bollman, "Measuring Economic Well-Being of Rural Canadians Using Income Indicators", Statistics Canada Agriculture and Rural Working Paper Series 45 (2001).

[75] Saint-Paul, Gilles, "The Dynamics of Exclusion and Fiscal Conservatism", *Review of Economic Dynamics*, 4 (2001), 275-302.

[76] Statistics Austrialia: Australia Bureau of Statistics, "Australian Social Trends 2002" (2002).

[77] Statistics Japan: Statistics Bureau and Statistics Center, "1999 National Survey of Family Income and Expenditure" (2002).

[78] Statistics Korea: Korea National Statistics Office, "Household Income and Expenditure Survey" (2002). Data collected from multiple surveys.

[79] Statistics New Zealand: Statistics New Zealand, "New Zealand Now: Incomes" (1999).

[80] Statistics Norway: Statistics Norway, "Recent Trends in Income Inequality in Norway" (2002).

[81] Statistics Philippines: National Statistics Office, "Family Income and Expenditure Survey" (2002). Data collected from multiple surveys.

[82] Statistics United States: United States Census Bureau, "The Changing Shape of the Nation's Income Distribution: Consumer Income" (2000).

[83] Suhrcke, Marc, "Preferences for Inequality: East vs. West", HWWA Discussion Paper 150 (2001).

[84] United States Department of Labor, Employment Standards Administration, *Minimum Wage and Overtime Hours Under the Fair Labor Standards Act* (Washington, D.C.: Government Printing Office, 1998).

[85] UNU/WIDER-UNDP World Income Inequality Database, Version 1.0, 12 September 2000.

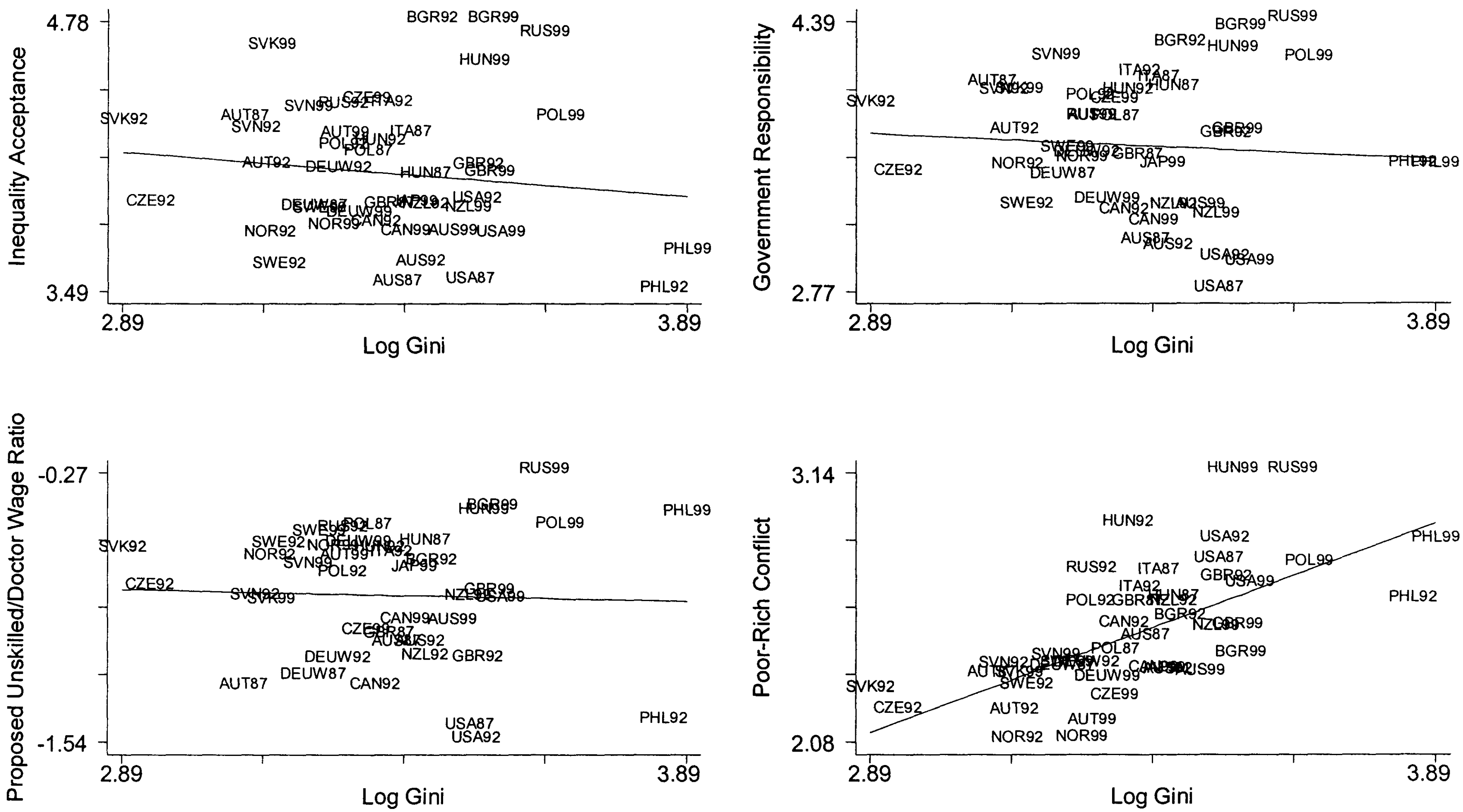


Figure 4.1: ISSP Inequality Norm Levels

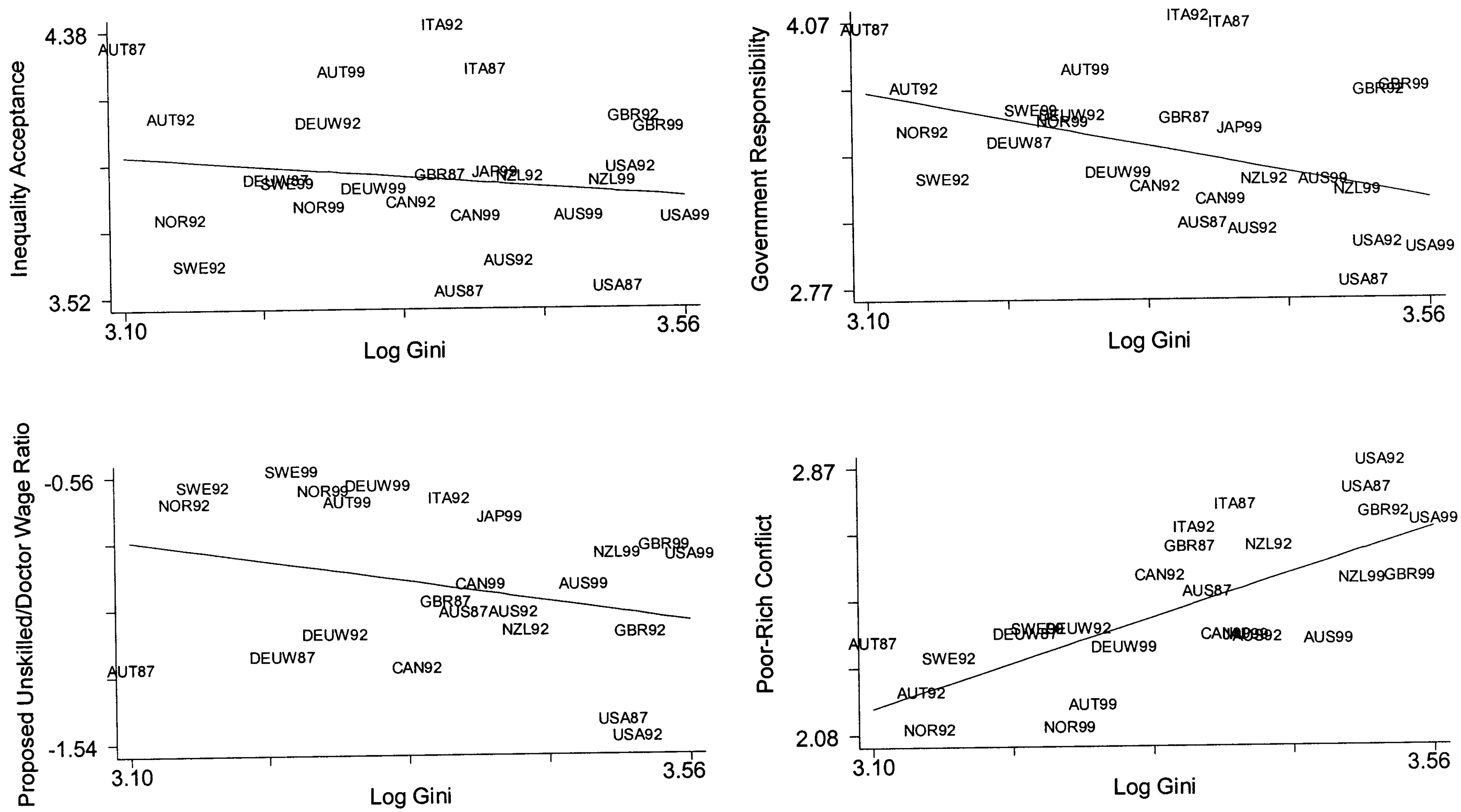


Figure 4.2: ISSP Inequality Norms Levels - OECD

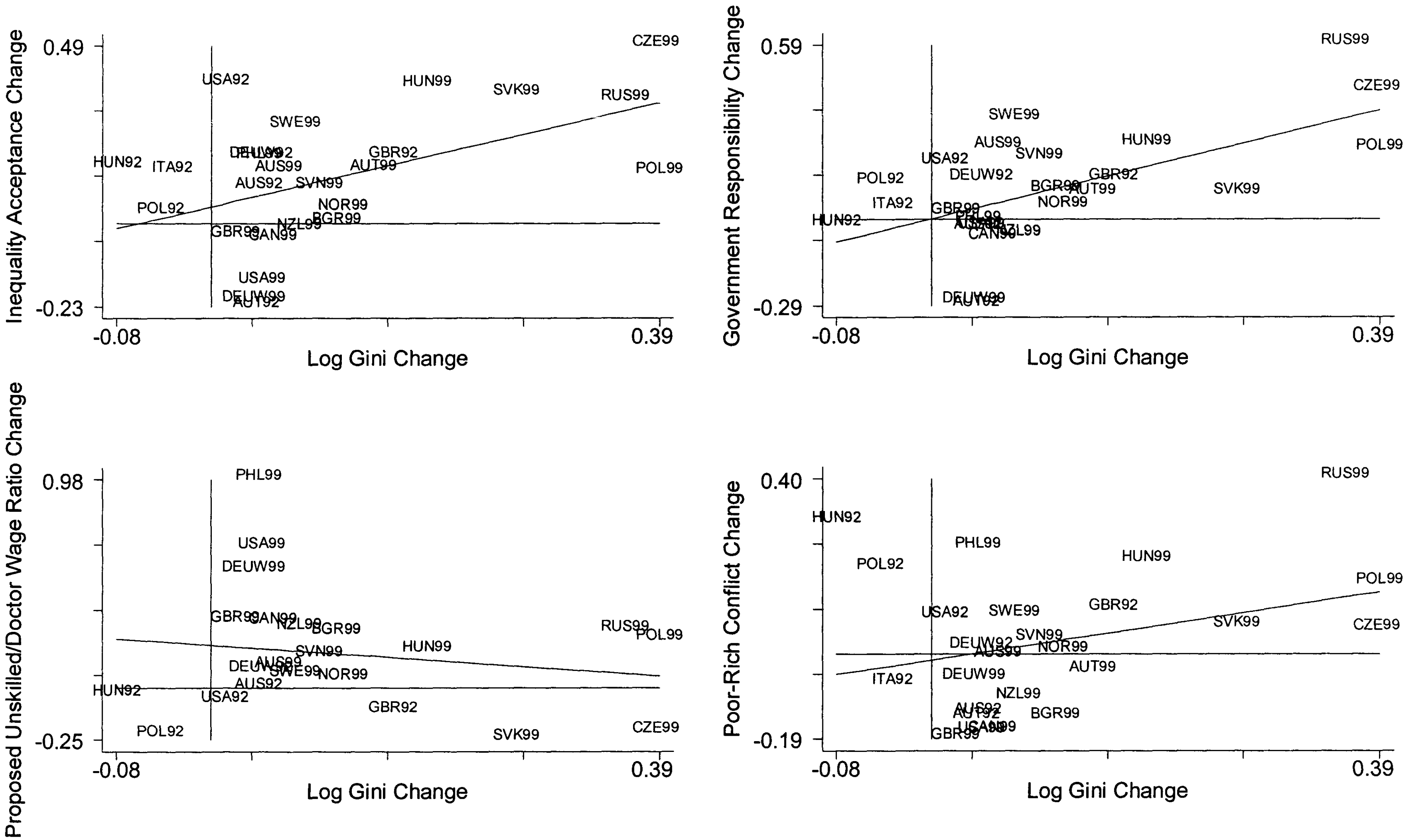


Figure 4.3: ISSP Inequality Norms Changes

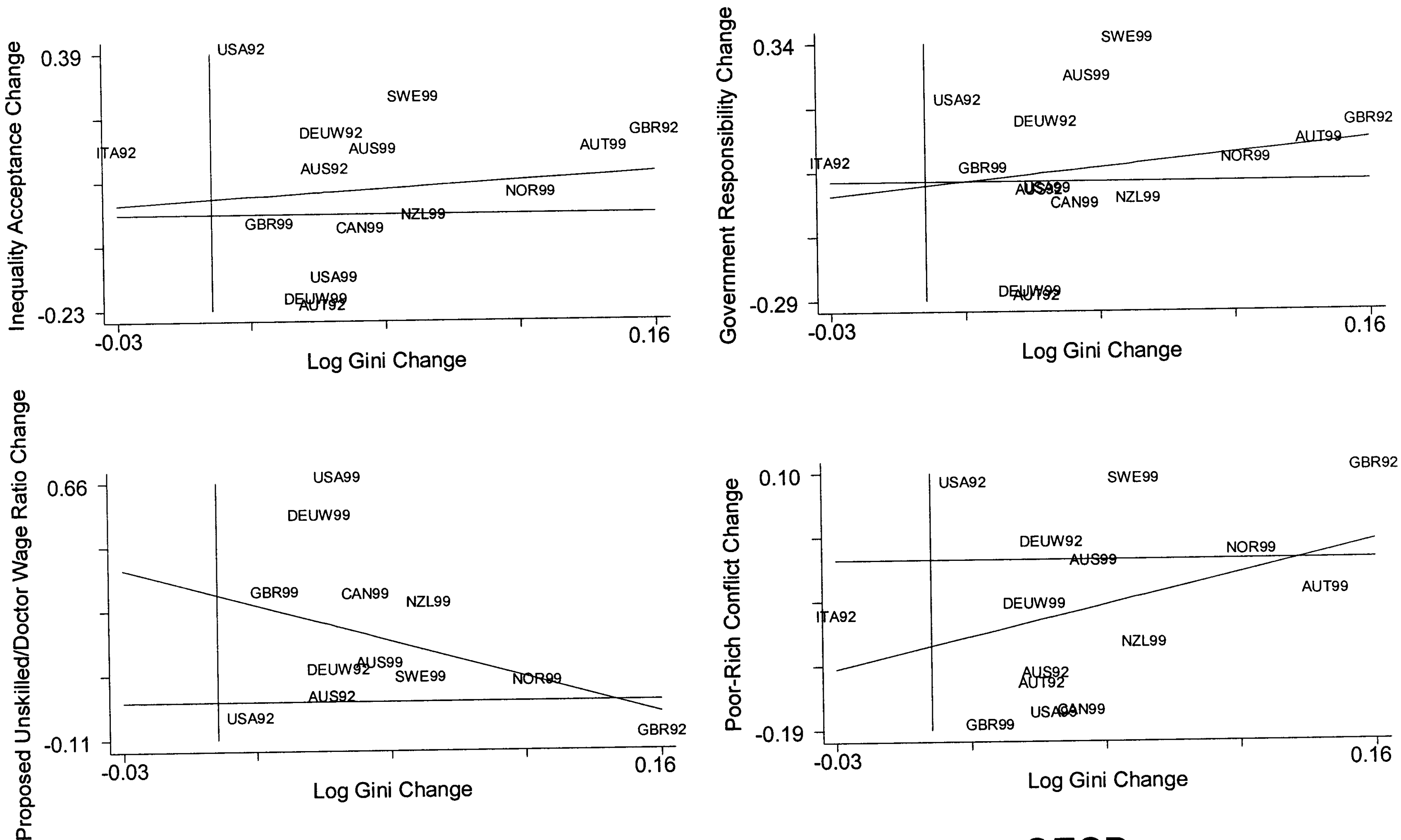


Figure 4.4: ISSP Inequality Norms Changes - OECD

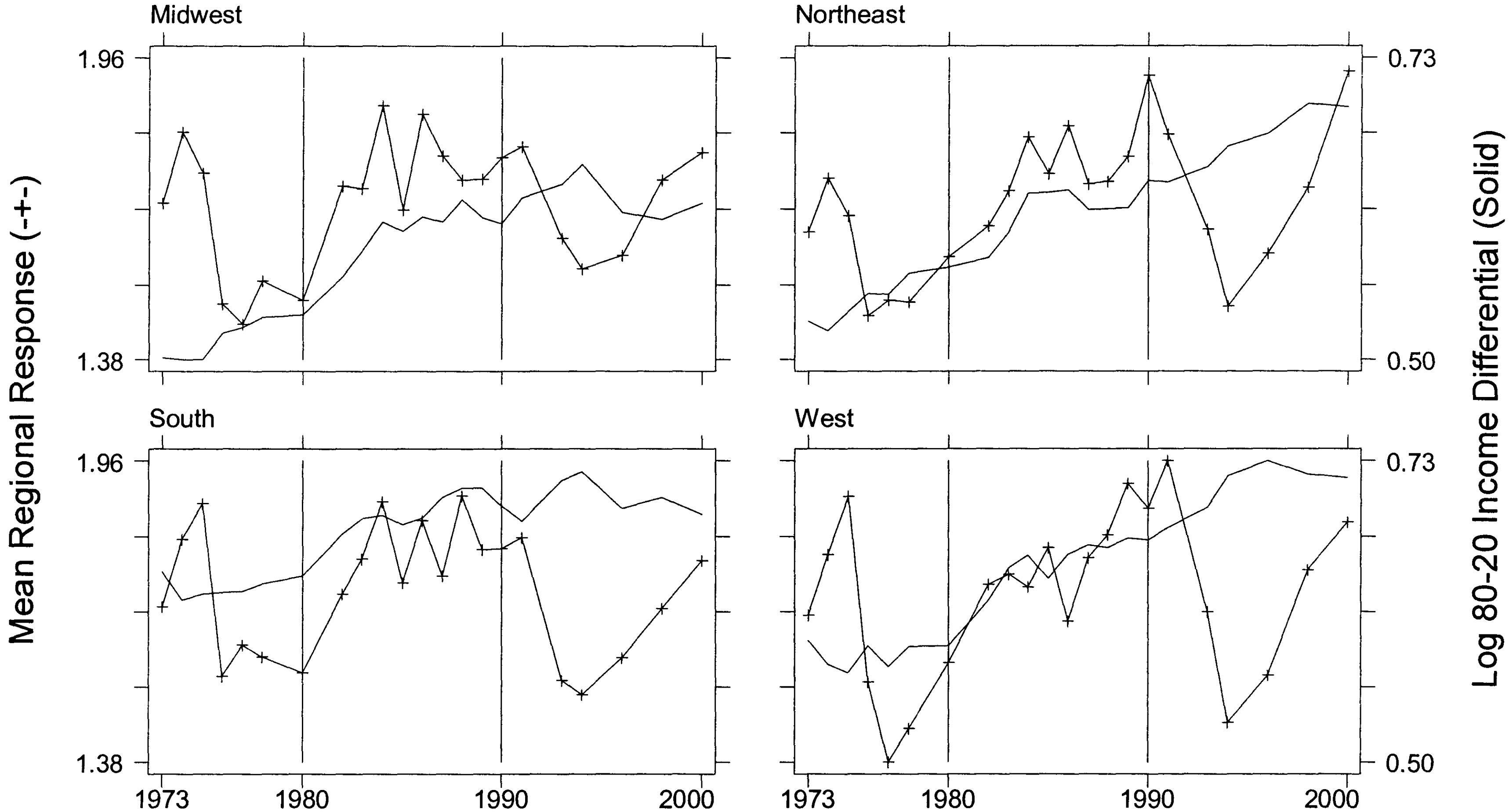


Figure 4.5: United States Welfare Spending Norms Trends

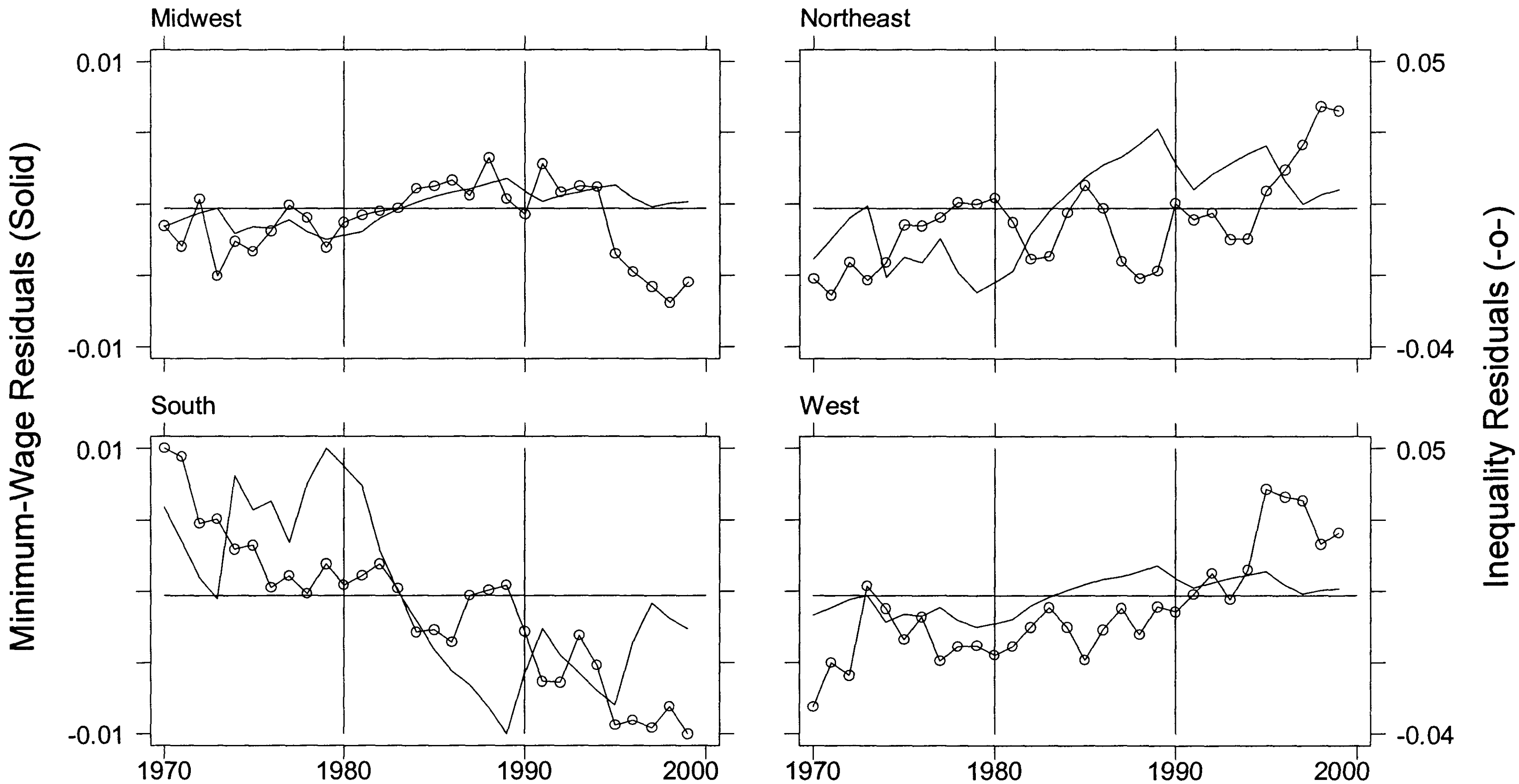


Figure 4.6: First-Stage Residuals

Table 4.1: ISSP and WVS Descriptive Statistics

	Total Sample	Long-Term OECD	Non Long-Term OECD
	A. ISSP Social Inequality Panel		
Countries	18	10	8
Respondents	64,424	34,375	30,049
Inequality Acceptance	4.06	3.87	4.30
(1-5 Scale)	(0.99)	(0.99)	(0.93)
Government Responsibility	3.66	3.42	3.97
(1-5 Scale)	(1.18)	(1.19)	(1.08)
Progressive Taxation	4.02	3.97	4.10
(1-5 Scale)	(0.77)	(0.73)	(0.83)
Unskilled/Doctor Wage Ratio	0.50	0.44	0.58
	(0.45)	(0.40)	(0.50)
Poor-Rich Conflict	2.55	2.48	2.64
(1-4 Scale)	(0.85)	(0.78)	(0.92)
Young-Old Conflict	2.22	2.23	2.20
(1-4 Scale)	(0.81)	(0.75)	(0.87)
Log Gini Coefficient	3.39	3.35	3.44
	(0.20)	(0.13)	(0.25)
	B. WVS Social Inequality Panel		
Countries	22	7	15
Respondents	79,127	32,989	46,138
WVS Income Equalization	5.04	5.18	4.95
(1-10 Scale)	(2.95)	(2.69)	(3.11)
Log Gini Coefficient	3.48	3.33	3.63
	(0.29)	(0.15)	(0.32)

Notes: Standard deviations are indicated in parentheses. To be included, a country must have participated in at least two surveys and have appropriate inequality data for those survey periods. Sample sizes in regressions are smaller than total respondents as some respondents skipped questions; surveys also varied on the demographic and mobility information collected. ISSP Long-Term OECD Members include AUS, AUT, CAN, DEU, GBR, ITA, NOR, NZL, SWE, and USA. ISSP Non-Long-Term OECD Members include BGR, CZE, HUN, PHL, POL, RUS, SVK, and SVN. WVS Long-Term OECD Members include ESP, FIN, GBR, JAP, NOR, SWE, and USA. WVS Non-Long-Term OECD Members include ARG, BGR, BLR, BRA, CHL, CHN, IND, KOR, LTU, LVA, MEX, POL, RUS, SVN, and ZAF.

Table 4.2: ISSP and WVS Regressions with Aggregate Inequality

	Base Regression	Including Demographic and Mobility Controls				
		Base Regression	Including Worker Controls	Including OECD-Yr Effects	Including Trans.-Yr Effects	Including Time Trend
	(1)	(2)	(3)	(4)	(5)	(6)
A. ISSP Social Inequality Panel						
Government Responsibility Responses						
Log National Gini Coefficient	0.179 (0.038)	0.186 (0.035)	0.261 (0.037)	0.158 (0.038)	0.149 (0.046)	0.185 (0.031)
Observations	54,054	38,066	19,970	38,066	38,066	38,066
Progressive Taxation Responses						
Log National Gini Coefficient	0.187 (0.102)	0.197 (0.097)	0.200 (0.100)	0.129 (0.113)	0.187 (0.121)	0.135 (0.087)
Inequality Acceptance Responses						
Log National Gini Coefficient	0.165 (0.045)	0.171 (0.048)	0.243 (0.056)	0.124 (0.056)	0.141 (0.064)	0.130 (0.048)
Log Unskilled/Doctor Wage Ratio Responses						
Log National Gini Coefficient	-0.452 (0.121)	-0.396 (0.120)	-0.365 (0.112)	-0.535 (0.186)	-0.331 (0.150)	-0.282 (0.120)
Poor-Rich Conflict Responses						
Log National Gini Coefficient	0.129 (0.047)	0.132 (0.035)	0.138 (0.039)	0.103 (0.053)	0.157 (0.051)	0.107 (0.036)
Young-Old Conflict Responses						
Log National Gini Coefficient	-0.014 (0.026)	-0.006 (0.023)	-0.029 (0.020)	-0.048 (0.050)	0.023 (0.024)	0.016 (0.017)
B. WVS Panel						
WVS Income Equalization Responses						
Log National Gini Coefficient	0.224 (0.098)	0.443 (0.199)	0.578 (0.224)	0.473 (0.216)	0.542 (0.340)	0.443 (0.199)
Observations	79,127	62,378	49,574	62,378	62,378	62,378

Notes: Regressions include the log GDP per capita for each country, country fixed effects, and year fixed effects. Clustered standard errors are in parentheses. Observations for Government Responsibility are representative for other ISSP variables. Variables are transformed to have a mean of zero and a standard deviation of one for presentation. Demographic Controls include sex, married, age, education, and income dummies. Mobility Controls include respondents' views on the determinants of success and comparisons of their jobs to their fathers' jobs (ISSP). Work Controls include self-employed, unemployed, supervisor, and union-member dummies. Regressions are weighted for nationally representative samples and equal cross-national weight.

Table 4.3A: GSS Regressions with Aggregate Inequality (Gini)

	Base Regression	Including Demographic and Mobility Controls: Base Regression	Including Worker Controls	Including Racial Controls	Excluding South Region	Including Time Trend
	(1)	(2)	(3)	(4)	(5)	(6)
Welfare Spending Responses						
Log Regional Gini Coefficient	0.130	0.135	0.114	0.132	0.076	0.239
	(0.030)	(0.032)	(0.033)	(0.031)	(0.053)	(0.051)
Observations	24,247	21,965	14,704	21,965	14,658	21,965
Income Equalization Responses						
Log Regional Gini Coefficient	0.086	0.040	0.059	0.023	0.072	0.016
	(0.033)	(0.032)	(0.033)	(0.032)	(0.049)	(0.044)
Observations	20,414	18,344	17,293	18,344	12,129	18,344
Party Identification Responses						
Log Regional Gini Coefficient	0.198	0.206	0.217	0.196	0.183	0.129
	(0.030)	(0.032)	(0.037)	(0.029)	(0.048)	(0.030)
Observations	37,763	33,971	23,026	33,791	22,469	33,791
Space Exploration Program Spending Responses						
Log Regional Gini Coefficient	-0.044	-0.047	-0.067	-0.047	-0.117	-0.149
	(0.032)	(0.033)	(0.045)	(0.033)	(0.049)	(0.040)
Observations	23,942	21,757	14,574	21,757	14,592	21,757

Notes: Regressions include the log median income for each region, region fixed effects, and year fixed effects. Clustered standard errors are in parentheses. Variables are transformed to have a mean of zero and a standard deviation of one for presentation. Demographic Controls include sex, married, age, education, and income dummies. Mobility Controls include recent changes in family financial position. Work Controls include self-employed, unemployed, and union-member dummies. Racial Controls include non-white respondent dummy. Regressions are weighted for nationally representative samples.

Table 4.3B: GSS Regressions with Aggregate Inequality (80-20 Differential)

	Base Regression	Including Demographic and Mobility Controls: Base Regression	Including Worker Controls	Including Racial Controls	Excluding South Region	Including Time Trend
	(1)	(2)	(3)	(4)	(5)	(6)
			Welfare Spending Responses			
Log Regional 80/20 Differential	0.098	0.114	0.127	0.112	0.081	0.217
	(0.027)	(0.032)	(0.034)	(0.032)	(0.052)	(0.049)
Observations	24,247	21,965	14,704	21,965	14,658	21,965
			Income Equalization Responses			
Log Regional 80/20 Differential	0.099	0.040	0.051	0.026	0.103	0.132
	(0.025)	(0.032)	(0.033)	(0.032)	(0.051)	(0.033)
Observations	20,414	18,344	17,293	18,344	12,129	18,344
			Party Identification Responses			
Log Regional 80/20 Differential	0.135	0.164	0.173	0.158	0.108	0.036
	(0.024)	(0.027)	(0.035)	(0.026)	(0.045)	(0.032)
Observations	37,763	33,971	23,026	33,791	22,469	33,791
			Space Exploration Program Spending Responses			
Log Regional 80/20 Differential	0.002	-0.015	-0.021	-0.016	-0.049	0.081
	(0.025)	(0.029)	(0.046)	(0.029)	(0.051)	(0.046)
Observations	23,942	21,757	14,574	21,757	14,592	21,757

Notes: See Table 4-3A.

Table 4.4: GSS Regressions with Disaggregated Inequality

	Base Regression	Base Regression	Including Worker Controls	Including Racial Controls	Excluding South Region	Including Time Trend
		Including Demographic and Mobility Controls				
	(1)	(2)	(3)	(4)	(5)	(6)
			Welfare Spending Responses			
Log Regional 80/50 Differential	0.013 (0.035)	0.013 (0.036)	0.002 (0.033)	0.001 (0.034)	-0.019 (0.032)	0.163 (0.073)
Log Regional 50/20 Differential	0.072 (0.018)	0.084 (0.021)	0.098 (0.023)	0.086 (0.021)	0.096 (0.045)	0.104 (0.036)
Observations	24,247	21,965	14,704	21,965	14,658	21,965
			Income Equalization Responses			
Log Regional 80/50 Differential	0.067 (0.030)	0.042 (0.033)	0.042 (0.035)	0.028 (0.032)	0.047 (0.036)	0.061 (0.046)
Log Regional 50/20 Differential	0.046 (0.020)	0.011 (0.019)	0.020 (0.020)	0.007 (0.019)	0.058 (0.046)	0.077 (0.030)
Observations	20,414	18,344	17,293	18,344	12,129	18,344
			Party Identification Responses			
Log Regional 80/50 Differential	0.036 (0.034)	0.035 (0.038)	0.002 (0.042)	0.015 (0.036)	-0.036 (0.038)	0.047 (0.042)
Log Regional 50/20 Differential	0.093 (0.020)	0.114 (0.020)	0.137 (0.025)	0.118 (0.018)	0.135 (0.038)	0.009 (0.020)
Observations	37,763	33,971	23,026	33,791	22,469	33,791
			Space Exploration Program Spending Responses			
Log Regional 80/50 Differential	0.017 (0.032)	0.002 (0.035)	-0.008 (0.046)	0.010 (0.034)	0.023 (0.039)	-0.093 (0.055)
Log Regional 50/20 Differential	-0.006 (0.019)	-0.012 (0.018)	-0.013 (0.026)	-0.016 (0.019)	-0.067 (0.039)	0.100 (0.030)
Observations	23,942	21,757	14,574	21,757	14,592	21,757

Notes: See Table 4-3A.

Table 4.5: GSS Regressions with Minimum-Wage Instrument

	Base Regression	Including Demographic and Mobility Controls: Base Regression	Including Worker Controls	Including Racial Controls	Excluding South Region
	(1)	(2)	(3)	(4)	(5)
		A. Second-Stage Coefficients			
		Welfare Spending Responses			
Log Regional 80/20 Differential	0.204 (0.075)	0.206 (0.081)	0.207 (0.067)	0.200 (0.083)	0.111 (0.082)
Observations	24,247	21,965	14,704	21,965	14,658
		Income Equalization Responses			
Log Regional 80/20 Differential	0.128 (0.077)	0.070 (0.064)	0.083 (0.063)	0.055 (0.063)	0.369 (0.181)
Observations	20,414	18,344	17,293	18,344	12,129
		Party Identification Responses			
Log Regional 80/20 Differential	0.209 (0.057)	0.232 (0.052)	0.196 (0.057)	0.224 (0.050)	0.330 (0.159)
Observations	37,763	33,971	23,026	33,971	22,469
		Space Exploration Program Spending Responses			
Log Regional 80/20 Differential	-0.054 (0.058)	-0.035 (0.056)	-0.034 (0.063)	-0.038 (0.056)	-0.353 (0.211)
Observations	23,942	21,757	14,574	21,757	14,592
		B. First-Stage Coefficients (Welfare Spending)			
		Log Regional 80/20 Differential			
Minimum Wage Instrument	2.216 (0.090)	1.958 (0.068)	1.958 (0.068)	1.958 (0.068)	6.004 (0.360)
R-Squared FS	0.18	0.23	0.23	0.23	0.12

Notes: See Table 4-3A.

Table 4.6: GSS Regressions with Extended Income Definitions and Regions

Source of Log 80/20 Inequality Metric	OLS			IV		
	Four Regions	Nine Regions	State Level	Four Regions	Nine Regions	State Level
	(1)	(2)	(3)	(4)	(5)	(6)
			Welfare Spending Responses			
Post-Tax Family Disposable Income	0.114 (0.032)	0.061 (0.028)	0.081 (0.023)	0.206 (0.081)	0.194 (0.062)	0.177 (0.066)
Pre-Tax Family Labor Earnings	0.105 (0.035)	0.068 (0.030)	0.041 (0.021)	0.209 (0.075)	0.215 (0.069)	0.295 (0.169)
Total Population Hourly Wage	0.030 (0.027)	0.056 (0.019)	0.067 (0.018)	0.593 (0.474)	0.227 (0.090)	0.217 (0.095)
			Income Equalization Responses			
Post-Tax Family Disposable Income	0.040 (0.032)	0.027 (0.026)	0.068 (0.029)	0.070 (0.064)	0.042 (0.075)	-0.077 (0.105)
Pre-Tax Family Labor Earnings	0.032 (0.030)	0.023 (0.025)	0.020 (0.022)	0.098 (0.097)	0.049 (0.092)	-0.204 (0.350)
Total Population Hourly Wage	0.054 (0.014)	0.018 (0.018)	0.053 (0.016)	0.305 (0.879)	0.047 (0.173)	0.321 (0.568)
			Party Identification Responses			
Post-Tax Family Disposable Income	0.164 (0.027)	0.099 (0.023)	0.050 (0.027)	0.232 (0.052)	0.202 (0.060)	0.224 (0.080)
Pre-Tax Family Labor Earnings	0.143 (0.029)	0.100 (0.024)	0.018 (0.025)	0.250 (0.060)	0.226 (0.063)	0.383 (0.225)
Total Population Hourly Wage	0.066 (0.019)	0.038 (0.016)	0.056 (0.020)	0.636 (0.411)	0.235 (0.090)	0.317 (0.148)
			Space Exploration Program Spending Responses			
Post-Tax Family Disposable Income	-0.015 (0.029)	-0.006 (0.024)	0.012 (0.021)	-0.035 (0.056)	-0.022 (0.050)	0.059 (0.054)
Pre-Tax Family Labor Earnings	-0.034 (0.028)	-0.055 (0.023)	-0.007 (0.018)	-0.033 (0.055)	-0.023 (0.054)	0.095 (0.095)
Total Population Hourly Wage	-0.022 (0.022)	-0.006 (0.013)	-0.012 (0.015)	-0.109 (0.183)	-0.032 (0.066)	0.076 (0.070)

Notes: Cells represent separate regressions. Regressions include Demographic and Mobility Controls, the log median income for each region, region fixed effects, and year fixed effects. Clustered standard errors are in parentheses. Variables are transformed to have a mean of zero and a standard deviation of one for presentation. Regressions are weighted for nationally representative samples.

Table 4.7: GSS Regressions with Income and Neighborhood Interactions

	Base Regression	Income Quintile Interaction	Racial Heterogeneity Interaction
	(1)	(2)	(3)
		Welfare Spending Responses	
Log Regional 80/20 Differential	0.114 (0.032)	0.105 (0.038)	0.141 (0.032)
Log 80/20 x Bottom Two Quintiles		-0.001 (0.024)	
Log 80/20 x Top Two Quintiles		0.023 (0.024)	
Log 80/20 x Racial Heterogeneity			-0.024 (0.013)
Observations	21,965	21,965	20,359
		Income Equalization Responses	
Log Regional 80/20 Differential	0.040 (0.032)	0.033 (0.036)	0.035 (0.033)
Log 80/20 x Bottom Two Quintiles		-0.011 (0.024)	
Log 80/20 x Top Two Quintiles		-0.005 (0.022)	
Log 80/20 x Racial Heterogeneity			-0.006 (0.015)
Observations	18,344	18,344	17,493
		Party Identification Responses	
Log Regional 80/20 Differential	0.164 (0.027)	0.135 (0.029)	0.192 (0.026)
Log 80/20 x Bottom Two Quintiles		0.036 (0.015)	
Log 80/20 x Top Two Quintiles		0.012 (0.014)	
Log 80/20 x Racial Heterogeneity			-0.039 (0.010)
Observations	33,791	33,791	31,730

Notes: Regressions include Demographic, Mobility, and Racial Controls, the log median income for each region, region fixed effects, and year fixed effects. Column 3 includes a main effect for Heterogeneous Neighborhood. Clustered standard errors are in parentheses. Variables are transformed to have a mean of zero and a standard deviation of one for presentation. Regressions are weighted for nationally representative samples.

Table 4.A1: Source Data for Gini Coefficients

Country	Years	Unit	Income Definition	Source	Notes
ARG (Argentina)	80, 89	Individual	Income	WIID (5 NOOK)	First series urban population.
	90-97	Household	Income	WIID (5 NOOK)	Questionable quality.
AUS (Australia)	81, 85, 89, 94	Household	Disposable Income	LIS, Gottschalk and Smeeding	
	95-98	Household	Disposable Income	Statistics Australia (2002)	
AUT (Austria)	87, 94, 95, 97	Household	Disposable Income	LIS	Interpolated series. Questionable quality.
	80-91	Individual	Gross Earnings	WIID (4 OKWN)	
BEL (Belgium)	85, 88, 92, 97	Household	Disposable Income	LIS	
BLR (Belarus)	88, 94	Unknown	Gross Income	WIID (5 NOOK)	Series consistent with available net household income statistics. Questionable quality.
BGR (Bulgaria)	89, 91	Household	Disposable Income	World Bank	
	92-97	Household	Disposable Income	WIID (1 OKIN)	
BRA (Brazil)	81, 83-89	Household	Gross Income	WIID (1 OKIN)	Interpolated series. Questionable quality.
	87, 96	Family	Gross Income	WIID (1 OKIN)	
CAN1 (Canada)	80-95	Household	Disposable Income	Gottschalk and Smeeding	Adjusted to 1991 LIS level; second series urban population; constructed series consistent with LIS data.
	90-98	Family	Income	Rupnik, Thompson-James, and Bollman (2001)	
CAN2 (Canada)	81, 87, 91, 94, 97, 98	Household	Disposable Income	LIS	
CHE (Switzerland)	82, 92	Household	Disposable Income	LIS	
CHL (Chile)	90, 92, 94, 96	Household	Gross Income	WIID (1 OKIN)	
CHN (China)	88-95	Household	Disposable Income	WIID (1 OKIN, 1 OKIR)	Interpolated series.
CZE1 (Czech Rep.)	91-97	Household	Disposable Income	WIID (1 OKIN)	Series has a sharper trend than in LIS.
CZE2 (Czech Rep.)	92, 96	Household	Disposable Income	LIS	
DEU1 (W. Germany)	83, 85, 87, 90-93	Household	Disposable Income	WIID (1 OKIN)	Interpolated series. Constructed series has a sharper trend than in LIS data.
	80, 83, 93, 95	Household	Disposable Income	Gottschalk and Smeeding	
	93, 98	Household	Disposable Income	Frick and Grabka (2002)	
DEU2 (W. Germany)	81, 83, 84, 89, 94	Household	Disposable Income	LIS	
DNK1 (Denmark)	81-90	Household	Disposable Income	Gottschalk and Smeeding	Adjusted to 1987 LIS level.
DNK2 (Denmark)	87, 92, 95, 97	Household	Disposable Income	LIS	
ESP (Spain)	80, 90	Household	Disposable Income	LIS	LIS currently negotiating for a 1995 dataset. Questionable series.
	90, 95	Unknown	Income	Fanjul and Renes (2002)	
EST (Estonia)	92-97	Household	Disposable Income	WIID (1 OKIN)	Consistent with 2000 LIS level.

Table 4.A1: Source Data for Gini Coefficients (continued)

Country	Years	Unit	Income Definition	Source	Notes
FIN (Finland)	86-97	Household	Disposable Income	Gottschalk and Smeeding , WIID (1 OKIN)	Adjusted to 1991 LIS level; series consistent with LIS data.
FRA (France)	81, 84, 89, 94	Household	Disposable Income	LIS, Gottschalk and Smeeding	
GBR (Great Britain)	80-99	Household	Disposable Income	Goodman (2001), Gottschalk and Smeeding	Consistent with LIS data.
HUN1 (Hungary)	82, 87, 89	Individual	Disposable Income	WIID (1 OKIN)	Adjusted to 1991 LIS level; constructed series consistent with LIS data except
	89, 91, 93-97	Individual	Disposable Income	WIID (1 OKIN)	
HUN2 (Hungary)	91, 94, 99	Household	Disposable Income	LIS	
IND (India)	83, 87-88, 90-92, 94-98	Household	Net Expenditure	WIID (3 OKIU)	Series urban population. Questionable quality.
IRL (Ireland)	80, 87	Household	Disposable Income	WIID (1 OKIN)	
	87, 94-96	Household	Disposable Income	LIS, Gottschalk and Smeeding	
ISR (Israel)	82, 86, 92	Household	Disposable Income	Gottschalk and Smeeding	
	86, 92, 97	Household	Disposable Income	LIS	
ITA (Italy)	80-87, 89, 91, 93, 95	Household	Disposable Income	Gottschalk and Smeeding, Brandolini (1999)	Adjusted to 1991 LIS level; constructed series consistent with LIS data.
JAP (Japan)	84, 89, 94, 99	Household	Income	Statistics Japan (2002), Gottschalk and Smeeding	
KOR (South Korea)	80, 85, 88	Household	Gross Income	WIID (1 OKIN)	Second series urban population.
	89-01	Household	Gross Income	Statistics Korea (2002)	
LVA (Lativa)	91-96	Individual	Gross Earnings	WIID (4 OKWN)	Levels consistent with available net household income statistics. Questionable quality.
LTU (Lithuania)	89, 92, 94-96	Individual	Gross Earnings	WIID (4 OKWN)	Levels consistent with available net household income statistics. Questionable quality.
MEX (Mexico)	84, 89, 92, 94, 96, 98	Household	Disposable Income	LIS	
NGA (Nigeria)	91, 93, 97	Household	Expenditure	WIID (1 OKIN)	Questionable quality.
NLD1 (Netherlands)	81, 83, 85, 88-95	Household	Disposable Income	Gottschalk and Smeeding	Adjusted to 1991 LIS level; series has a sharper trend than in LIS data.
NLD2 (Netherlands)	83, 87, 91, 94	Household	Disposable Income	LIS	

Table 4.A1: Source Data for Gini Coefficients (continued)

Country	Years	Unit	Income Definition	Source	Notes
NOR (Norway)	79, 91	Household	Disposable Income	Gottschalk and Smeeding , WIID (1 OKIN)	Constructed series consistent with LIS data. 1979 value listed as 1980.
	86-97	Household	Disposable Income	Statistics Norway (2002), Brandolini	
NZL (New Zealand)	82, 84, 86, 88-97	Household	Disposable Income	Statistics New Zealand (1999)	
PHL (Philippines)	85, 88, 91, 94, 97, 00	Family	Income	Statistics Philippines (2002)	2000 value listed as 1999.
POL1 (Poland)	86-92	Individual	Gross Income	WIID (1 OKIN)	Adjusted to 1992 LIS level; constructed series has a sharper trend than in LIS data.
	89-97	Household	Disposable Income	WIID (1 OKIN)	
POL2 (Poland)	86, 92, 95, 99	Household	Disposable Income	LIS	
RUS1 (Russia)	89-99	Household	Income	Ovtcharova (2001)	Series trend set off one year from LIS data. Questionable series.
RUS2 (Russia)	92, 95	Household	Disposable Income	LIS	
SVK (Slovakia)	89-97	Household	Disposable Income	WIID (1 OKIN)	Consistent with LIS data.
SVN (Slovenia)	91-96	Household	Disposable Income	WIID (1 OKIN)	Interpolated series.
	89-97	Individual	Gross Earnings	WIID (4 OKWN)	
	97, 99	Household	Disposable Income	LIS	
SWE (Sweden)	81, 87, 92, 95	Household	Disposable Income	LIS	Interpolated series.
	89-99	Family	Disposable Income	Gottschalk and Smeeding, WIID (1 OKIN)	
USA (United States)	80-98	Household	Disposable Income	United States Census Bureau (2000), Gottschalk and Smeeding	Adjusted to 1991 LIS level. Break in series between 1992 and 1993.
ZAF (South Africa)	90,95	Household	Gross Income	WIID (1 OKIN)	

Notes: The first inequality series is employed where two are indicated (e.g., RUS1 and RUS2), but the results are robust to substituting the second. Appropriate gini estimates have not been identified for CYP, ICL, PRT, and TUR.

Table 4.A2: Gini Coefficients

Country	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
ARG	37.0										43.1	44.1	43.6	43.8	46.3	45.8	44.8	46.7		
AUS		28.1				29.2				30.4					31.1	32.0	31.5	30.9	32.2	
AUT	21.8	21.8	21.8	22.3	22.3	22.3	22.3	22.7	22.7	22.7	22.7	23.1			28.0	27.7	26.6			
BEL						22.7			23.2				22.4					25.0		
BLR									23.0						28.0					
BGR										24.3			31.1	31.9	35.6	37.2	34.8	34.6		
BRA		55.3		56.2	55.5	56.7	55.9	57.0	58.6	59.5	58.3						58.1			
CAN1	28.2	27.6	27.8	28.6	28.3	28.2	28.3	28.1	27.9	27.7	27.7	28.1	27.8	28.1	27.7	28.0	29.0	29.5	29.6	
CAN2		28.4						28.3				28.1			28.4			29.1	30.5	
CHE			30.9										30.7							
CHL											54.7		52.2		55.6		56.4			
CHN									38.2	38.0	39.3	38.9	39.7	40.6	41.8	43.1				
CZE1												18.9	20.3	21.6	22.1	21.5	28.1	27.6		
CZE2													20.7				25.9			
DEU1	25.4			25.0		26.0		25.2			26.0	26.3	26.4	27.4		27.5			27.3	
DEU2		24.4		26.0	24.9					24.7		26.1								
DNK1		24.9	24.6	24.4	24.8	25.1	25.1	25.4	26.3	26.5	27.8									
DNK2								25.4					23.6			26.3		25.7		
ESP	31.8										30.3					31.2				
EST													41.2	38.8	39.6	39.0	37.4	34.1		
FIN							21.3	20.7	21.2	21.3	21.2	21.0	20.7	21.8	21.6	21.8	22.6	23.6		
FRA		28.8			29.2					28.7					28.8					
GBR	25.3	25.9	25.8	26.4	26.6	27.9	28.8	30.2	32.0	32.4	33.7	33.7	34.0	33.7	33.0	33.0	33.3	33.8	34.4	34.2
HUN1		26.7	26.7					30.7		29.7		28.3		31.4	32.1	33.7	33.9	34.1		
HUN2												28.3			32.3					29.5
IND				33.4				35.6	35.6		35.6	34.0	35.5		34.5	33.4	35.4	36.1		
IRL	33.5							32.8							33.3	33.6	32.5			
ISR			30.0				30.8						30.5					33.6		

Table 4.A2: Gini Coefficients (continued)

Country	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
ITA	30.9	30.3	28.4	28.9	29.8	29.9	29.9	31.6		29.3		28.9		32.2		32.0				
JAP					28.0					29.3					29.7					30.1
KOR	34.9					31.2				30.4	29.5	28.7	28.4	28.1	28.5	28.4	29.1	28.3	31.6	32.0
LVA											22.5	22.5	29.6	24.5	30.7	31.7	32.2			
LTU										26.0			37.2		34.9	34.1	35.0			
MEX					44.8					46.7			48.5		49.6		47.7		49.4	
NGA												45.0		45.0				50.6		
NLD1		23.9		23.5		24.7			25.6	25.5	25.8	26.6	27.8	27.7	27.8	28.3				
NLD2				26.0				25.6				26.6			25.3					
NOR	22.5						22.2	22.3	21.9	23.4	22.8	23.3	23.7	24.3	25.4	24.8	25.7	26.1		
NZL			25.9		26.0		25.3		25.8	28.0	29.9	30.7	29.9	31.8	31.0	31.8	32.2	33.1		
PHL						44.6			44.5			46.8			45.1			48.7		48.2
POL1							27.7	28.1	27.6	28.6		26.5	27.4	36.2	37.3	36.9	37.8	39.0		
POL2							27.1						27.4			31.8				29.3
RUS1										26.5	28.5	26.5	28.9	39.8	40.9	38.1	37.5	37.5	37.9	39.4
RUS2													39.3			44.7				
SVK										18.1	17.8	18.0	18.6	19.7	20.8	20.0	24.8	23.4		
SVN										19.0	20.1	22.7	22.6	25.0	22.0	23.4	24.0	25.0		24.9
SWE		19.7						21.8		22.1	22.3	23.7	22.9	23.4	26.2	23.3	24.3	26.2	25.4	26.7
USA	31.2	31.5	32.3	32.5	32.5	33.0	33.2	33.3	33.4	33.9	33.6	33.6	34.1	35.6	35.8	35.3	34.5	35.0	35.1	
ZAF											63.0					59.0				

Notes: Shading indicates years in which countries participated in the ISSP Role of the Government or Social Inequality modules; countries must participate in at least two surveys to be included. Other country-year gini observations are used with alternative ISSP modules and the WVS (these surveys are not highlighted). The target gini estimates are one-year lags (i.e., the square to the left of the shaded survey year), although contemporaneous or two-year or three-year lags are accepted when necessary. Survey responses are dropped if they do not meet these conditions. The first inequality series is employed where two are indicated (e.g., RUS1 and RUS2), but the results are robust to substituting the second. Appropriate gini estimates have not been identified for CYP, ICL, PRT, and TUR.

Table 4.A3: Minimum-Wage Instrument Descriptive Statistics

	Nominal	Real	Log Ratio	Expected Coverage Ratios				Log 80-20 Family Disposable Income			
Year	M. Wage	M. Wage	to 1970	Northeast	Midwest	South	West	Northeast	Midwest	South	West
1970	1.60	5.03	0.00	89.9	87.1	78.4	87.2	0.500	0.487	0.638	0.527
1971	1.60	4.81	0.04	89.9	87.1	78.3	87.2	0.509	0.495	0.649	0.555
1972	1.60	4.59	0.09	89.9	87.1	78.3	87.2	0.525	0.515	0.635	0.557
1973	1.60	4.46	0.12	90.2	87.5	78.9	87.6	0.532	0.504	0.649	0.597
1974	2.00	5.25	-0.04	90.3	87.7	79.2	87.8	0.525	0.503	0.628	0.578
1975	2.10	5.01	0.00	90.4	87.8	79.3	87.9	0.540	0.503	0.632	0.572
1976	2.30	5.07	-0.01	90.5	87.9	79.6	88.0	0.554	0.523	0.633	0.592
1977	2.30	4.80	0.05	90.6	88.0	79.6	88.1	0.553	0.528	0.634	0.576
1978	2.65	5.20	-0.03	90.7	88.2	79.8	88.2	0.569	0.536	0.640	0.592
1979	2.90	5.45	-0.08	90.7	88.2	79.8	88.2	0.565	0.523	0.646	0.589
1980	3.10	5.33	-0.06	90.8	88.4	80.1	88.4	0.574	0.538	0.646	0.593
1981	3.35	5.18	-0.03	90.8	88.4	80.1	88.4	0.577	0.550	0.659	0.605
1982	3.35	4.74	0.06	90.8	88.4	80.1	88.4	0.581	0.567	0.678	0.627
1983	3.35	4.48	0.12	90.8	88.4	80.1	88.4	0.601	0.587	0.690	0.652
1984	3.35	4.30	0.16	90.8	88.4	80.1	88.4	0.630	0.608	0.692	0.662
1985	3.35	4.13	0.20	90.9	88.4	80.1	88.4	0.631	0.601	0.685	0.644
1986	3.35	4.00	0.23	90.9	88.4	80.2	88.5	0.633	0.612	0.690	0.662
1987	3.35	3.93	0.25	90.9	88.5	80.2	88.5	0.617	0.609	0.706	0.670
1988	3.35	3.80	0.28	90.9	88.5	80.3	88.5	0.618	0.626	0.713	0.667
1989	3.35	3.66	0.32	90.9	88.5	80.3	88.6	0.619	0.612	0.713	0.674
1990	3.80	3.99	0.23	90.9	88.5	80.3	88.6	0.640	0.607	0.699	0.673
1991	4.25	4.25	0.17	90.9	88.4	80.2	88.5	0.638	0.627	0.687	0.682
1992	4.25	4.10	0.20	90.9	88.4	80.1	88.5	0.650	0.628	0.697	0.698
1993	4.25	4.00	0.23	90.9	88.4	80.2	88.5	0.650	0.637	0.719	0.698
1994	4.25	3.90	0.25	90.9	88.4	80.1	88.5	0.666	0.652	0.726	0.722
1995	4.25	3.82	0.27	90.9	88.4	80.2	88.5	0.678	0.630	0.705	0.745
1996	4.75	4.17	0.19	90.9	88.4	80.2	88.5	0.675	0.616	0.698	0.733
1997	5.15	4.40	0.13	90.9	88.4	80.2	88.5	0.689	0.617	0.702	0.739
1998	5.15	4.31	0.15	90.9	88.4	80.2	88.5	0.698	0.610	0.706	0.723
1999	5.15	4.25	0.17	90.9	88.4	80.2	88.5	0.702	0.621	0.702	0.731
2000	5.15	4.16	0.19	90.9	88.4	80.2	88.5	0.696	0.623	0.693	0.721