The Ethnic Composition of US Inventors

William R. Kerr*
Harvard Business School
Boston MA

HBS Working Paper 08-006 (revised)
December 2008

Abstract

The ethnic composition of US scientists and engineers is undergoing a significant 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 contributions of Chinese and Indian scientists to US technology formation increased dramatically in the 1990s, before noticeably leveling off after 2000 and declining in the case of Indian researchers. Growth in ethnic innovation is concentrated in high-tech sectors; the institutional and geographic dimensions are further characterized.

JEL Classification: F15, F22, J44, J61, O31.

Key Words: Innovation, Research and Development, Patents, Scientists, Engineers, Inventors, Ethnicity, Immigration.

*Comments are appreciated and can be sent to wkerr@hbs.edu. This permanent working paper is continually updated as additional patenting data are collected. The first version is included in Kerr (2005). I am grateful to William Lincoln and Debbie Strumsky for data assistance. This research is supported by the National Science Foundation, HBS Research, the Innovation Policy and the Economy Group, and the MIT George Schultz Fund.

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 make 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 in the 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, the brain-drain or brain-circulation effect on sending countries, and the future prospects for US technology leadership are just four 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 inventor names contained on the micro-records for all patents granted by the United States Patent and Trademark Office (USPTO) from January 1975 to May 2008.[4] Each patent record lists one or more inventors, with 8 million inventor names associated with the 4.5 million patents. The USPTO grants patents to inventors living within and outside of the US, with each group accounting for about half of patents over the 1975-2008 period.

This study maps into these inventor names an ethnic-name database typically used for commercial applications.[5] 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 92%-98% 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,

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

[2] Representative papers are Lowell (2000), Borjas (2004), Saxenian (2002b), and Freeman (2005) 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 data offer a better trade-off between the two dimensions but suffer important sampling biases with respect to immigrants (Kannankutty and Wilkinson 1999).

[4] The project initially employed the NBER Patent Data File, compiled by Hall et al. (2001), that includes patents granted by the USPTO from January 1975 to December 1999. The current version now employs an extended version developed by HBS Research that includes patents granted through mid 2008. Some of the descriptive calculations have not been updated from their 1975-1999 values (noted in text).

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

and Vietnamese. Moreover, because the matching is done at the micro-level, greater detail on the ethnic composition of inventors is available annually on multiple dimensions: technologies, cities, companies, etc.[6]

The next section details the ethnic-name matching strategy, outlines the strengths and weaknesses of the database selected, and offers some validation exercises using patent records filed by foreign inventors with the USPTO. Section 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 granted to Chinese and Indian inventors is particularly striking, as is the leveling off in these trends after 2000. The relative contributions from scientists of European ethnicity, however, decline somewhat from their levels in 1975. The institutional and geographic dimensions of ethnic innovation are further delineated. Section 4 concludes.

2 Ethnic-Name Matching Technique

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 name-matching strategy, which is described in detail. Descriptive statistics are then provided 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. The section concludes with a further discussion of the advantages and disadvantages for empirical estimations of the resulting dataset.

2.1 Melissa Ethnic-Name Database and Name-Matching Technique

Ethnic-name databases suffer from two inherent limitations — not all ethnicities are covered, and included ethnicities usually receive unequal treatment. The strength of the ethnic-name database obtained from the Melissa Data Corporation 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 the matching technique. In total, nine ethnicities are distinguished: Chinese, English, European, Hispanic/Filipino, Indian/Hindi,

[6] This ethnic patenting database is employed by Kerr (2005, 2008a-c), Kerr and Lincoln (2008), and Foley and Kerr (2008) to study the role of ethnic scientists and entrepreneurs in technology formation and diffusion.

Japanese, Korean, Russian, and Vietnamese. The largest 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.[7]

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. For example, the 1975 to 1999 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%).

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.[8] The match rate for this procedure is 98% (98% US, 98% foreign). This rate should be less than 100% with the Melissa database as not all ethnicities are included.
>
> *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% (92% US, 86% foreign).

For matching, names in both the patent and ethnic-name databases are capitalized and truncated to ten characters. Approximately 88% of the patent name records have a unique surname, first

[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 names), and Vietnamese.

The final matching procedure employs a joint Hispanic/Filipino ethnicity, while in earlier work they are kept separate. These two ethnic groups are combined due to extensive name overlaps (e.g., the common surnames Martinez and Ramirez are in both ethnic lists), but this choice is not a first-order concern.

The Bangladeshi and Pakistani name counts are extremely small (8 and 15 respectively) and are not distinct from the Indian/Hindi names. Their assignment does not materially affect the Indian/Hindi outcome, which represents in some ways a South Asian identifer.

Jewish ethnic names overlap extensively with other ethnic groupings and are not separately treated. A handful of names classified as Arab, Burmese, and Malay are also discarded.

[8] A simple rule is applied to take advantage of the information embedded in the patent database itself. If over 90% of the 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 during 1975 to 1999.

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 in Metropolitan Statistical Areas (MSAs). Ethnic probabilities for the remaining 3% of records (mostly foreign) are calculated as equal shares. MSA ethnic compositions are averages of the 1980 and 1990 US 5% Census files; they are kept constant through the sample period. 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 (e.g., 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.

2.2 Inventors Residing in Foreign Countries and Regions

The application of the ethnic-name database to the inventors residing outside of the US provides a natural quality-assurance exercise for the technique. Inventions originating outside the US account for just under half of USPTO patents, with applications from Japan comprising about half of this foreign total. The top panel of Table 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 93% for all countries. In the Restricted Matching procedure, a matching rate of greater than 74% holds for all regions.

Second, the estimated inventor compositions are reasonable. The own-ethnicity shares are summarized in the fourth and fifth columns. The weighted average is 86% 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.

The bottom panel of Table 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, Vietnam), small sample sizes (Vietnam), or overlaps of common names. Two prominent examples of common names are the surname Lee (Chinese, English, and Korean) and the first name Igor (Hispanic and Russian). The most frequent name overlap occurs between the European and Hispanic ethnicities.[9]

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 assigned as equal shares. The last two columns of Table 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.

2.3 Advantages and Disadvantages of Name-Matching Technique

Visual confirmation of the top 1000 surnames and first names in the USPTO records confirms the name-matching technique works well. Table A1 in the appendix lists the 100 most common surnames of US-based inventors for each ethnicity, along with their relative contributions. These counts sum the ethnic contribution from inventors with each surname. These counts include partial or split assignments. Moreover, they are not necessarily direct or exclusive matches (e.g., the ethnic match may have occurred through the first name). 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 negligible differences when the Restricted Matching dataset is employed instead.

The matched records describe the ethnic composition of US SEs with previously unavailable detail: incorporating the major ethnicities working in the US SE community; separating out detailed technologies and manufacturing industries; providing city and state statistics; and providing annual metrics. Moreover, the assignment of patents to corporations and institutions affords firm-level and university-level characterizations (e.g., the ethnic composition of IBM's inventors filing computer patents from San Francisco in 1985). Detailed econometrics require this level of cross-sectional and longitudinal variation, and the next section provides graphical descriptions along these various dimensions. These descriptive statistics highlight the advantages of name matching through individual patent records.

[9]The main US SE ethnicity missing from the database is Iranian. Running the ethnic-name database on the few patents from Iran yields a 55%-77% match rate. Iran's predicted composition does not favor any of the nine ethnicities studied, with the largest overlap being the English ethnicity at 52%. Ongoing work is attempting to develop better strength for Iranian names.

The ethnic-name procedure does, however, have two potential limitations for empirical work that should be highlighted. First, the approach does not distinguish foreign-born ethnic researchers in the US from later generations working as SEs. The procedure can only estimate total ethnic SE populations, and these levels are to some extent measured with time-invariant error due to the name-matching approach. The resulting data are very powerful, however, for panel econometrics employ changes in these ethnic SE populations for identification. Moreover, Census and INS records confirm these changes are primarily due to new SE immigration for this period, substantially weakening this overall concern.

The name-matching technique also does not distinguish finer divisions within the nine major ethnic groupings. For ethnic network analyses, it would be advantageous to separate Mexican from Chilean scientists within the Hispanic ethnicity, to distinguish Chinese engineers with ethnic ties to Taipei versus Beijing versus Shanghai, and so on. These distinctions are not possible with the Melissa database, and researchers should understand that measurement error from the broader ethnic divisions may bias their estimated coefficients downward depending upon the application.[10] Nevertheless, Section 3 demonstrates how the deep variation available with the ethnic patenting data provides a much richer description of US ethnic invention than previously available.

3 Ethnic Composition of US Inventors

Table 2 describes the ethnic composition of US inventors for 1975-2004.[11] The trends demonstrate a growing ethnic contribution to US technology development, especially among Chinese and Indian scientists. Ethnic inventors are more concentrated in high-tech industries like computers and pharmaceuticals and in gateway cities relatively closer to their home countries (e.g., Chinese in San Francisco, European in New York, and Hispanics in Miami). The final three rows demonstrate a close correspondence of the estimated ethnic composition to the country-of-birth

[10] When mapping the ethnic patenting data to country-level data for international diffusion estimations, researchers will also need to cluster their standard errors to reflect the multiple country-to-ethnicity mappings.

[11] The current patent data incorporate all patents granted by May 2008. 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 leads to significant attrition in the last two years — patents are only included in the database if they have been granted, but a smaller number of applications close to the cut-off have completed the review cycle.

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 (e.g., 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 Marschke 2004, Hall 2005, Jaffe and Lerner 2005, and Branstetter and Ogura 2005). 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) and HBS Research are working on algorithms to identify individual scientists with the USPTO data.

composition of the US SE workforce in the 1990 Census.[12] The next four subsections more closely examine each dimension of this data.

3.1 Contributions by Year

Figure 1 illustrates the evolving ethnic composition of US inventors from 1975-2004. The omitted English share declines from 83% to 70% during this period. Looking across all technology categories, the European ethnicity is initially the largest foreign contributor to US technology development. Like the English ethnicity, however, the European share of US domestic inventors declines steadily from 8% in 1975 to 6% in 2004. This declining share is partly due to the exceptional growth over the thirty years of the Chinese and Indian ethnicities, which increase from under 2% to over 8% and 5%, respectively. 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. The Indian ethnic contribution declines somewhat after 2000, mostly due to changes within the computer technology sector as seen below.

Among the other ethnicities, the Hispanic contribution grows from 3% to 4% from 1975 to 2004. 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. The Korean share increases dramatically from 0.3% to 1.1% over the thirty years, while the Russian climbs from 1.2% to 2.2%. Although difficult to see with Figure 1's scaling, much of the Russian increase occurs in the 1990s following the dissolution of the Soviet Union. The Japanese share steadily increases from 0.6% to 1.0%. Finally, while the Vietnamese contribution is the lowest throughout the sample, it does exhibit the strongest relative growth from 0.1% to 0.6%.

3.2 Contributions by Technology

Figure 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 3-8 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

[12] The estimated European contribution in Table 2 is naturally higher than the immigrant contribution measured by foreign born.

the most interesting and economically meaningful variation across technologies, as summarized in Figures 9 and 10.[13]

3.3 Contributions by Institution

Figure 11 demonstrates that intriguing differences in ethnic scientific contributions also exist by institution type. Over the 1975-2004 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. Part of this levels difference is certainly due to immigration visa sponsorships by larger institutions. Growth in ethnic shares are initially stronger in the government and university labs, but publicly listed companies appear to close the gap by 2004. The other interesting trend in Figure 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.[14]

3.4 Contributions by Geography

This paper closes its descriptive statistics with an examination of the 1975-2004 ethnic inventor contributions by major cities in Table 3. Cities are defined through 281 Metropolitan Statistical Areas.[15] Not surprisingly, total patenting shares are highly correlated with city size, with the three largest shares of US domestic patenting for 1995-2004 found in San Francisco (12%), New York (7%), and Los Angeles (6%). More interestingly, non-English patenting is more concentrated than general innovation. The 1995-2004 non-English patent shares of San Francisco, New York, and Los Angeles are 19%, 10%, and 8%, respectively. Similarly, 81% of non-English invention occurs in the top 47 patenting cities listed in Table 3, compared to 73% of total patenting. Indian and Chinese invention is even further agglomerated. San Francisco shows exceptional growth from an 8% share of total US Indian and Chinese patenting in

[13] The USPTO issues patents by technology categories rather than by industries. Combining the work of Johnson (1999), Silverman (1999), and Kerr (2008a), concordances can be developed to map the USPTO classification scheme to the three-digit industries in which new inventions are manufactured or used. Scherer (1984) and Keller (2002) further discuss the importance of inter-industry R&D flows.

[14] Publicly listed companies are identified from a 1989 mapping developed by Hall et al. (2001). This company list is not updated for delistings or new public offerings. This approach maintains a constant public grouping for reference, but it also weakens the respresentativeness of the public and private company groupings at the sample extremes for current companies.

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. Unassigned patents account for about 26% of patents granted.

[15] 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 99%. Manual coding further ensures all patents with more than 100 citations and all city names with more than 100 patents are identified.

1975-1984 to 25% in 1995-2004, while the combined shares of New York and Chicago decline from 22% to 13%. Agrawal et al. (2007a,b) and Kerr (2008c) further describe ethnic inventor agglomeration in the US using the ethnic name approach.

Not only are ethnic scientists disproportionately concentrated in major cities, but growth in a city's share of ethnic patenting is highly correlated with growth in its share of total US patenting. Across the whole sample and including all of the intervening years, an increase of 1% in a city's ethnic patenting share correlates with a 0.6% increase in the city'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.[16]

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 would be feasible otherwise.

[16] 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. See Kerr and Lincoln (2008) for recent work in this area.

References

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

[2] Agrawal, Ajay, Devesh Kapur, and John McHale, "Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation", Working Paper (2007b).

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

[4] Branstetter, Lee, and Yoshiaki Ogura, "Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations", NBER Working Paper 11561 (2005).

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

[6] Foley, C. Fritz, and William Kerr, "US Ethnic Scientists and Foreign Direct Investment Placement", Working Paper (2008).

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

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

[9] Hall, Bronwyn, "Exploring the Patent Explosion", *Journal of Technology Transfer* 30 (2005), 35-48.

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

[11] Jaffe, Adam, and Joshua Lerner, *Innovation and Its Discontents* (Boston, MA: Harvard Business School Press, 2005).

[12] 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.

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

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

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

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

[17] Kerr, William, "Ethnic Scientific Communities and International Technology Diffusion", *Review of Economics and Statistics* 90:3 (2008a), 518-537.

[18] Kerr, William, "Heterogeneous Technology Diffusion and Ricardian Trade Patterns", Working Paper (2008b).

[19] Kerr, William, "The Agglomeration of US Ethnic Inventors", HBS Working Paper (2008c).

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

[21] Kerr, William, and William Lincoln, "The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention", HBS Working Paper 09-005 (2008).

[22] 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.

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

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

[25] 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).

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

[27] 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).

[28] 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.

[29] 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.

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

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

[32] Wadhwa, Vivek, AnnaLee Saxenian, Ben Rissing, and Gary Gereffi, "America's New Immigrant Entrepreneurs I", Working Paper (2007).

Fig. 1: Ethnic Share of US Domestic Patents

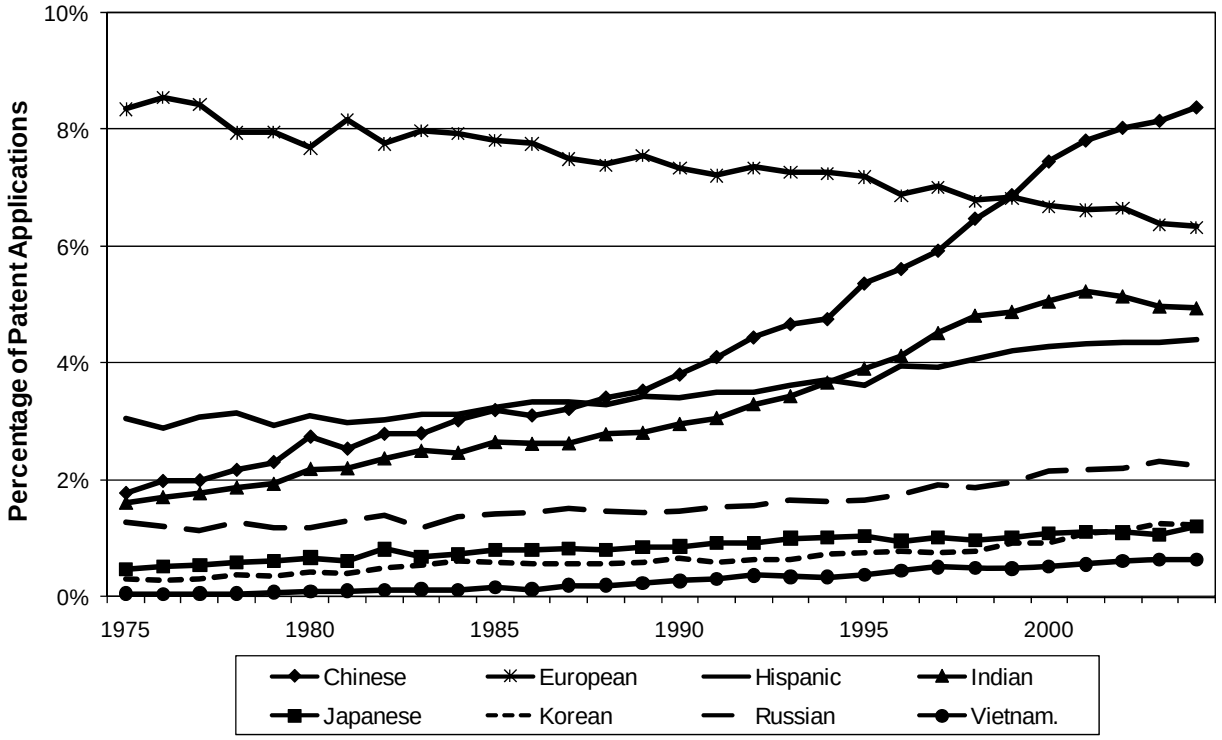

Fig. 2: Total US Ethnic Share by Technology

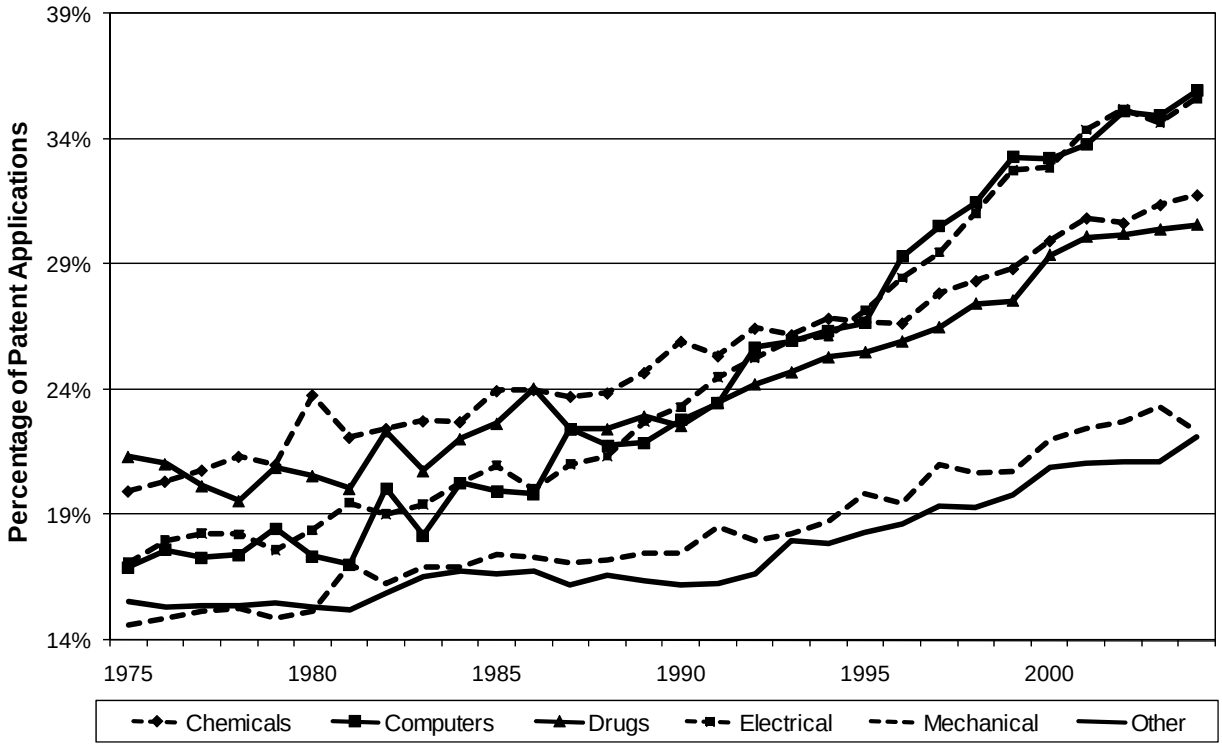

Fig. 3: US Ethnic Patenting - Chemicals

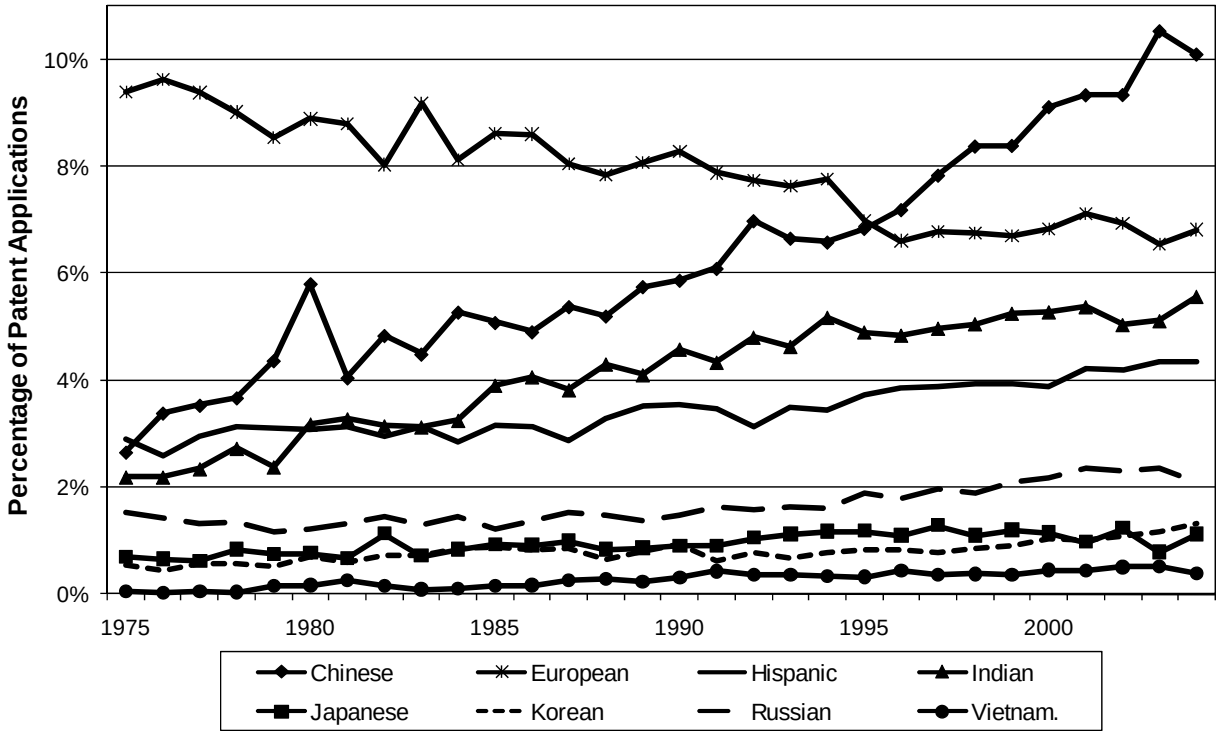

Fig. 4: US Ethnic Patenting - Computers

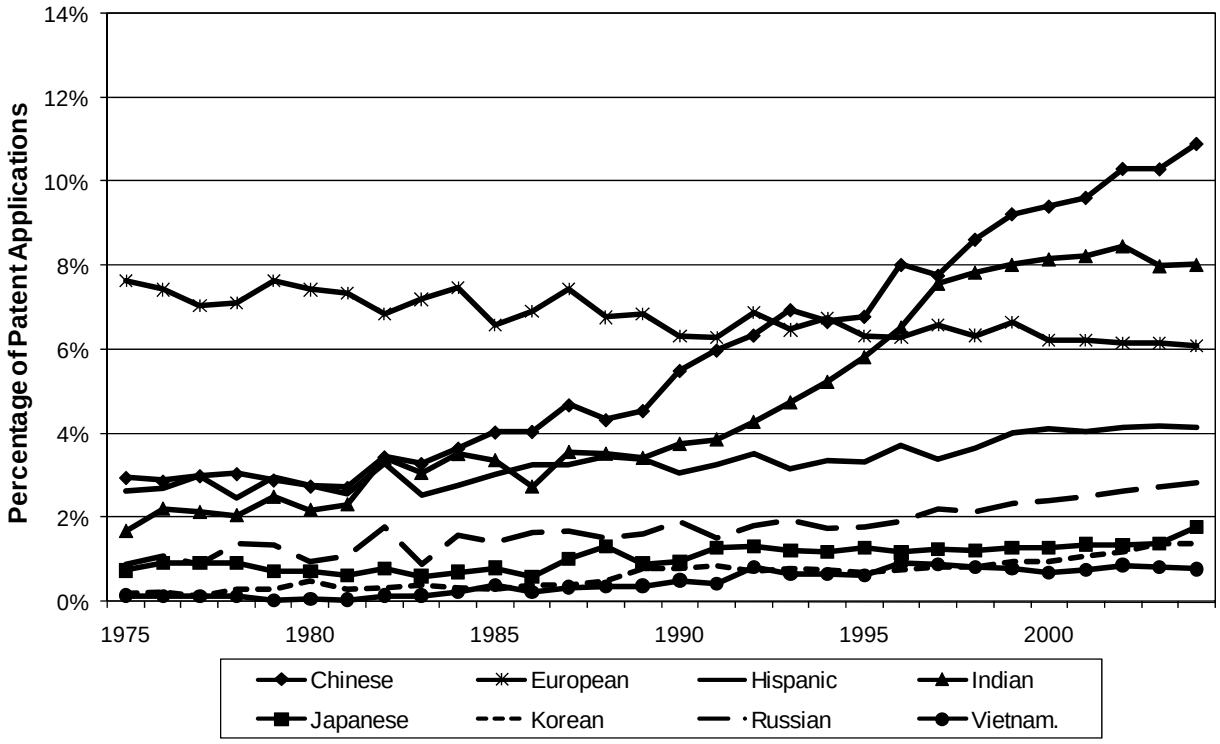

Fig. 5: US Ethnic Patenting - Drugs

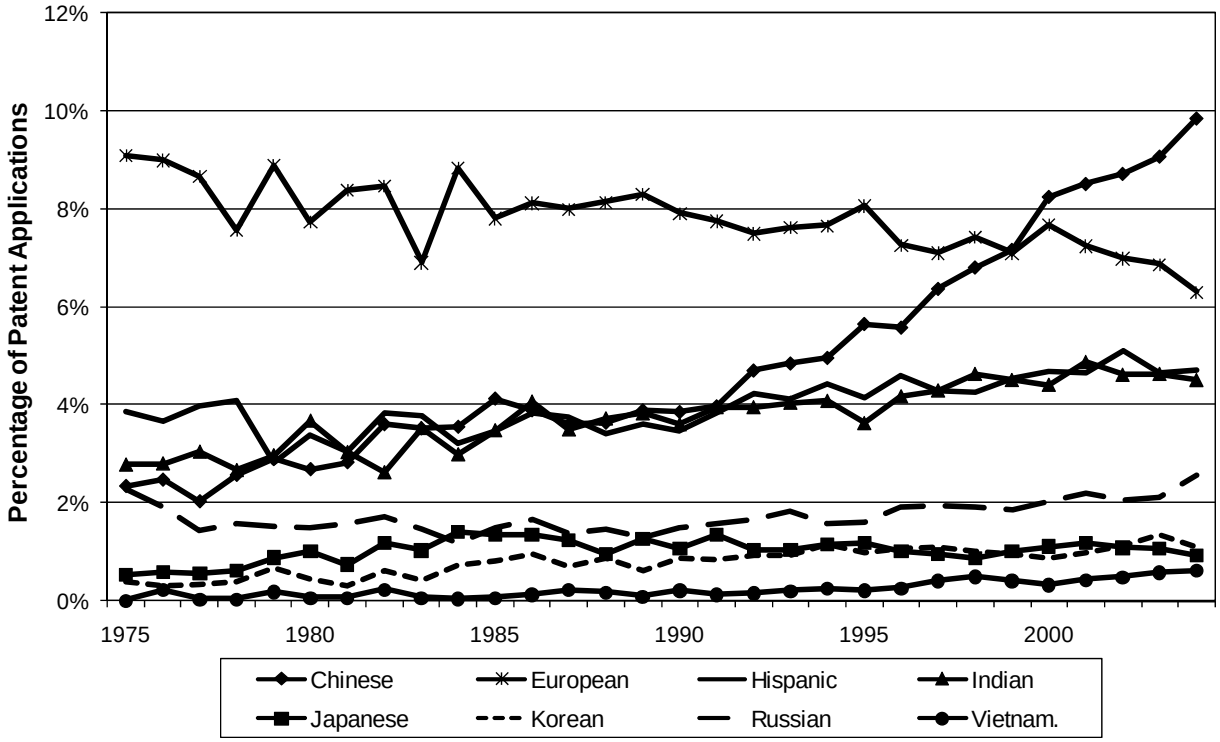

Fig. 6: US Ethnic Patenting - Electrical

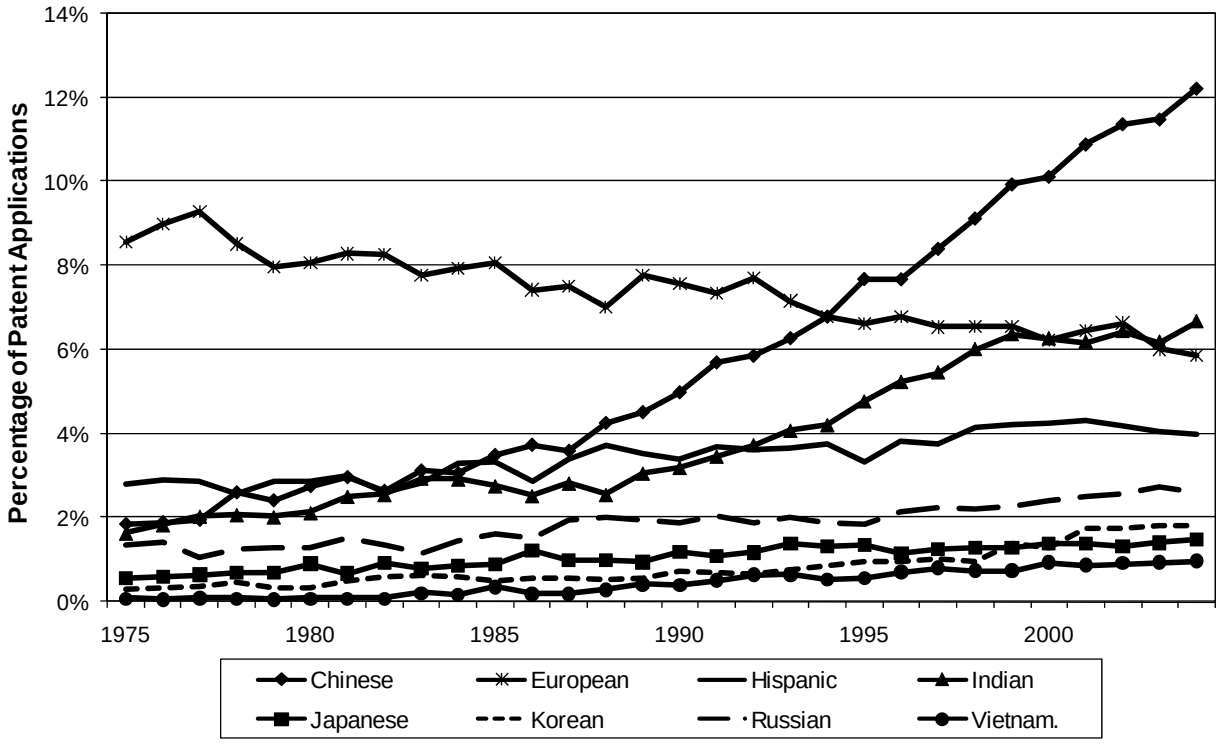

Fig. 7: US Ethnic Patenting - Mechanical

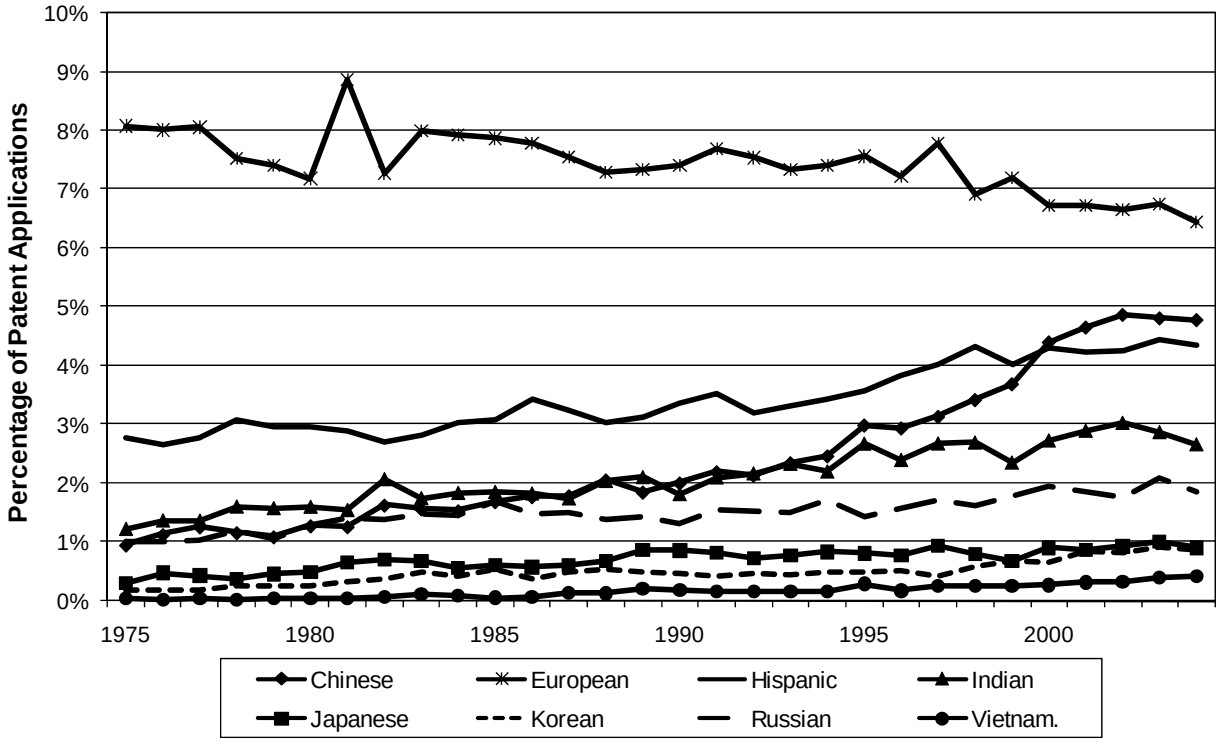

Fig. 8: US Ethnic Patenting - Other

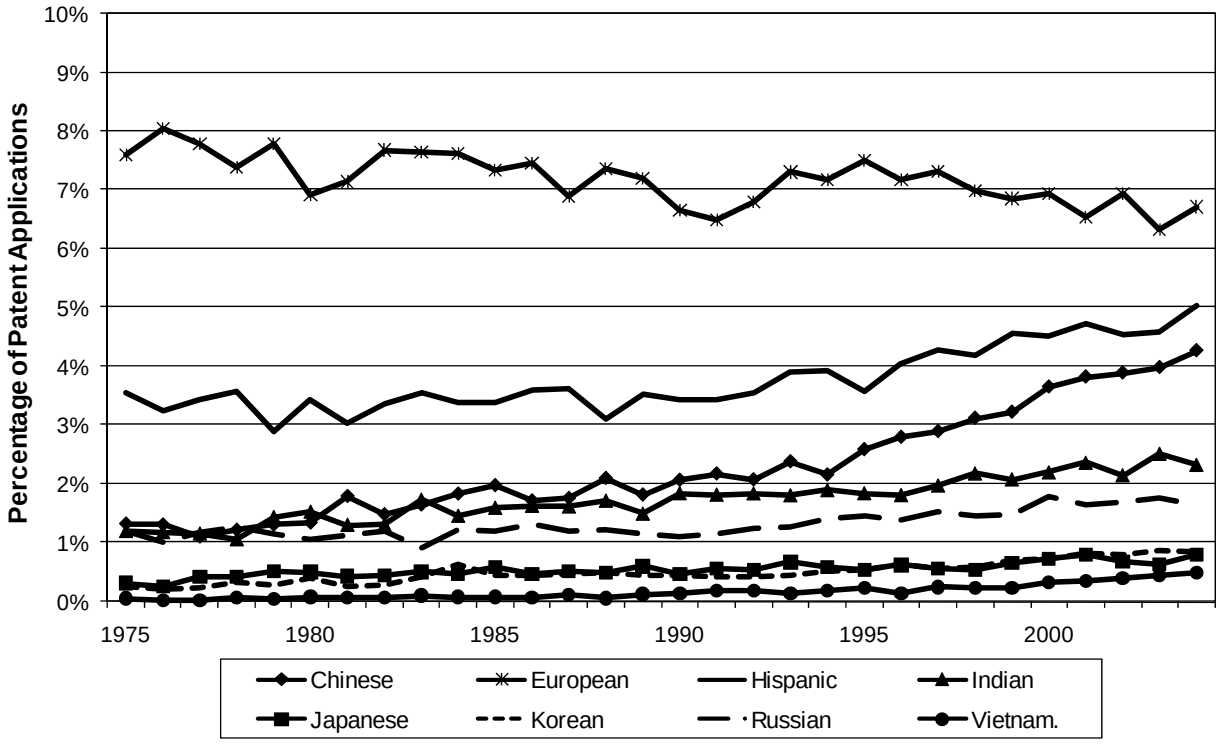

Fig. 9: Chinese Contribution by Technology

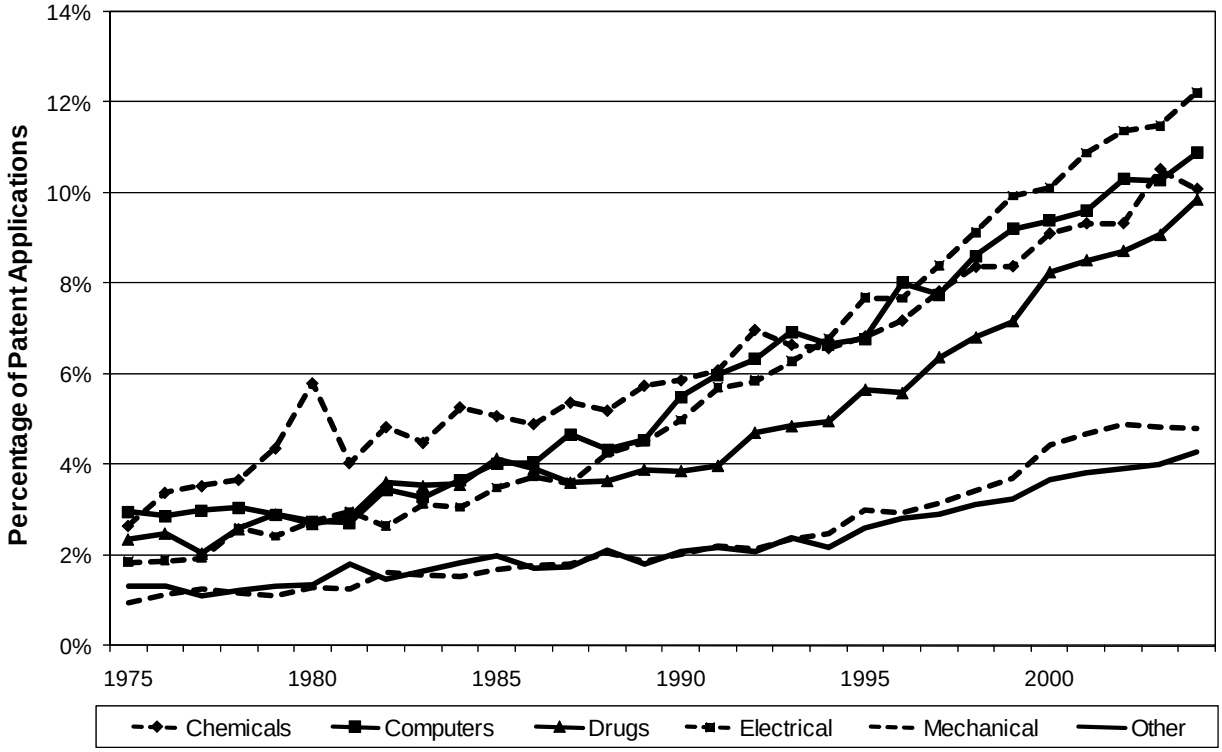

Fig. 10: Indian Contribution by Technology

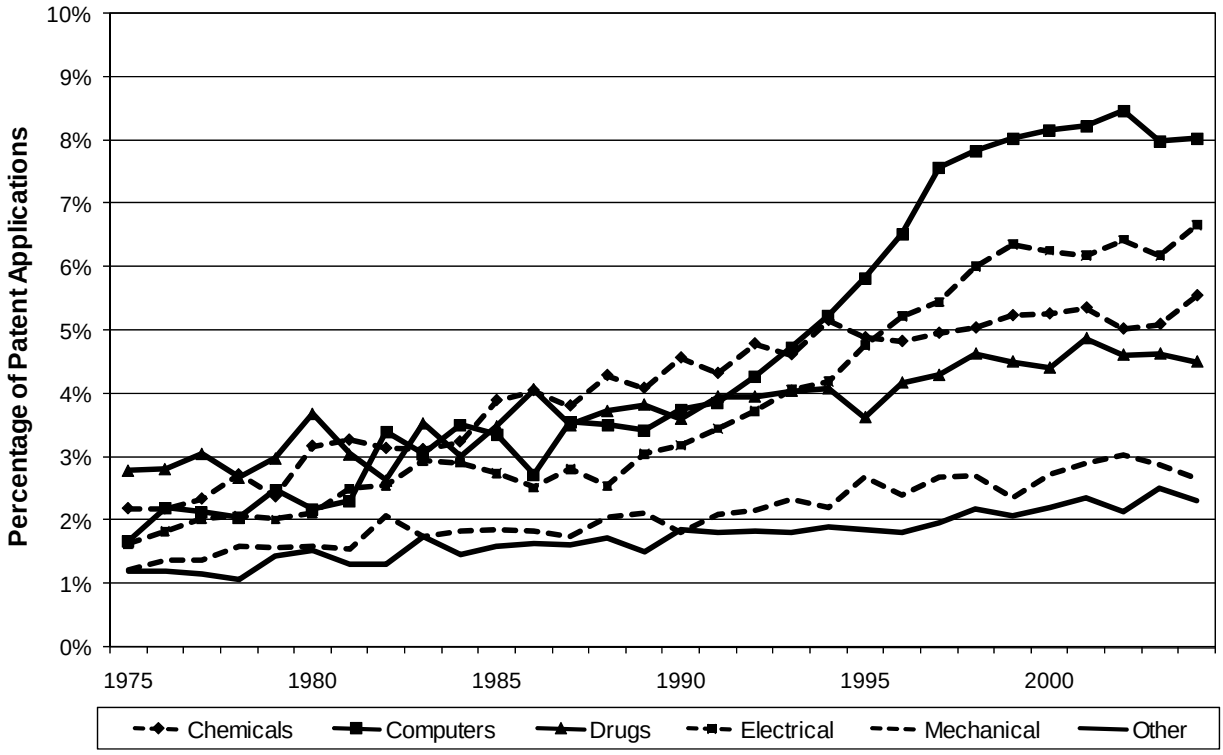

Fig. 11: Total US Ethnic Share by Institution

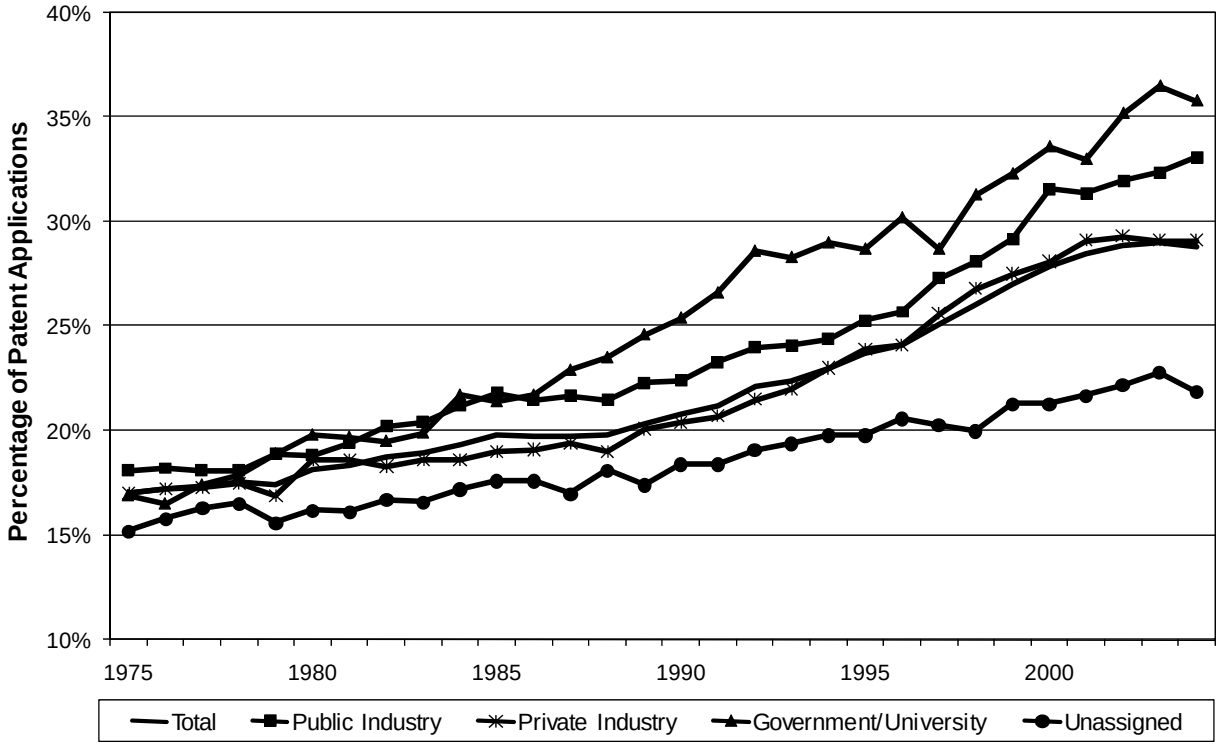

Table 1: Descriptive Statistics for Inventors Residing in Foreign Countries and Regions

Summary Statistics for Full and Restricted Matching Procedures

		Percentage of Region's Inventors Matched with Ethnic Database		Percentage of Region's Inventors Assigned Ethnicity of Their Region		Percentage of Region's Inventors Assigned Ethnicity of Region (Partial)	
	Obs.	Full	Restrict.	Full	Restrict.	Full	Restrict.
United Kingdom	187,266	99%	95%	85%	83%	92%	91%
China, Singapore	167,370	100%	98%	88%	89%	91%	91%
Western Europe	1,210,231	98%	79%	66%	46%	73%	58%
Hispanic Nations	27,298	99%	74%	74%	69%	93%	93%
India	13,582	93%	76%	88%	88%	90%	89%
Japan	1,822,253	100%	89%	100%	96%	100%	96%
South Korea	127,975	100%	100%	84%	83%	89%	88%
Russia	33,237	94%	78%	81%	84%	93%	94%
Vietnam	41	100%	98%	36%	43%	44%	43%

Complete Ethnic Composition of Region's Inventors (Full Matching)

	English	Chinese	European	Hispanic	Indian	Japanese	Korean	Russian	Vietnam.
United Kingdom	**85%**	2%	5%	3%	2%	0%	0%	2%	0%
China, Singapore	3%	**88%**	1%	1%	1%	1%	4%	1%	1%
Western Europe	21%	1%	**66%**	8%	1%	0%	0%	3%	0%
Hispanic Nations	11%	1%	10%	**74%**	0%	1%	0%	2%	0%
India	3%	1%	1%	5%	**88%**	0%	0%	2%	0%
Japan	0%	0%	0%	0%	0%	**100%**	0%	0%	0%
South Korea	2%	11%	0%	1%	0%	1%	**84%**	1%	0%
Russia	5%	1%	3%	9%	0%	0%	0%	**81%**	0%
Vietnam	17%	21%	12%	0%	0%	10%	2%	2%	**36%**

Notes: Matching is undertaken at inventor level using the Full and Restricted Matching procedures outlined in the text. The middle columns of the top panel summarize the share of each region's inventors assigned the ethnicity of that region; the complete composition for the Full Matching procedure is detailed in the bottom panel. The right-hand columns in the top panel document the percentage of the region's inventors assigned at least partially to their region's ethnicity.

Greater China includes Mainland China, Hong Kong, Macao, and Taiwan. Western Europe includes Austria, Belgium, Denmark, Finland, France, Germany, Italy, Luxembourg, Netherlands, 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 2: Descriptive Statistics for Inventors Residing in US

	Ethnicity of Inventor								
	English	Chinese	European	Hispanic	Indian	Japanese	Korean	Russian	Vietnam.
	A. Ethnic Inventor Shares Estimated from US Inventor Records, 1975-2004								
1975-1979	82.5%	2.2%	8.3%	2.9%	1.9%	0.6%	0.3%	1.2%	0.1%
1980-1984	81.1%	2.9%	7.9%	3.0%	2.4%	0.7%	0.5%	1.3%	0.1%
1985-1989	79.8%	3.6%	7.5%	3.2%	2.9%	0.8%	0.6%	1.4%	0.2%
1990-1994	77.6%	4.6%	7.2%	3.5%	3.6%	0.9%	0.7%	1.5%	0.4%
1995-1999	73.9%	6.5%	6.8%	3.9%	4.8%	0.9%	0.8%	1.8%	0.5%
2000-2004	70.4%	8.5%	6.4%	4.2%	5.4%	1.0%	1.1%	2.2%	0.6%
Chemicals	73.4%	7.2%	7.5%	3.6%	4.5%	1.0%	0.8%	1.7%	0.3%
Computers	70.1%	8.2%	6.3%	3.8%	6.9%	1.1%	0.9%	2.1%	0.7%
Pharmaceuticals	72.9%	7.1%	7.4%	4.3%	4.2%	1.1%	0.9%	1.8%	0.4%
Electrical	71.6%	8.0%	6.8%	3.7%	4.9%	1.1%	1.1%	2.1%	0.7%
Mechanical	80.4%	3.2%	7.1%	3.5%	2.6%	0.7%	0.6%	1.6%	0.2%
Miscellaneous	81.3%	2.9%	7.0%	3.8%	2.1%	0.6%	0.6%	1.4%	0.3%
Top Cities as a	KC (89)	SF (13)	NOR (12)	MIA (16)	SF (7)	SD (2)	BAL (2)	BOS (3)	AUS (2)
Percentage of	WS (88)	LA (8)	STL (11)	SA (9)	AUS (7)	SF (2)	LA (2)	NYC (3)	SF (1)
City's Patents	NAS (88)	AUS (6)	NYC (11)	WPB (7)	PRT (6)	LA (2)	SF (1)	SF (3)	LA (1)
	B. Ethnic Scientist and Engineer Shares Estimated from 1990 US Census Records								
Bachelors Share	87.6%	2.7%	2.3%	2.4%	2.3%	0.6%	0.5%	0.4%	1.2%
Masters Share	78.9%	6.7%	3.4%	2.2%	5.4%	0.9%	0.7%	0.8%	1.0%
Doctorate Share	71.2%	13.2%	4.0%	1.7%	6.5%	0.9%	1.5%	0.5%	0.4%

Notes: Panel A presents descriptive statistics for inventors residing in the US at the time of patent application. Inventor ethnicities are estimated through inventors' names using techniques described in the text. Patents are grouped by application years and major technology fields. Cities, defined through Metropolitan Statistical Areas, include AUS (Austin), BAL (Baltimore), BOS (Boston), KC (Kansas City), LA (Los Angeles), MIA (Miami), NAS (Nashville), NOR (New Orleans), NYC (New York City), PRT (Portland), SA (San Antonio), SD (San Diego), SF (San Francisco), STL (St. Louis), WPB (West Palm Beach), and WS (Winston-Salem). Cities 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 99%. Manual recoding further ensures all patents with more than 100 citations and all city names with more than 100 patents are identified. Panel B presents comparable statistics calculated from the 1990 Census using country of birth for scientists and engineers. Country groupings follow Table 1; English provides a residual in the Census statistics.

Table 3: Ethnic Inventor Contributions by City

	Total Patenting Share				non-English Ethnic Patenting Share				Indian and Chinese Patenting Share			
	1975-1984	1985-1994	1995-2004	2001-2006 (A)	1975-1984	1985-1994	1995-2004	2001-2006 (A)	1975-1984	1985-1994	1995-2004	2001-2006 (A)
Atlanta, GA	0.6%	1.0%	1.3%	1.5%	0.3%	0.7%	1.0%	1.1%	0.3%	0.7%	1.0%	1.2%
Austin, TX	0.4%	0.9%	1.8%	2.0%	0.5%	1.2%	1.9%	2.0%	0.4%	1.6%	2.3%	2.3%
Baltimore, MD	0.8%	0.8%	0.7%	0.7%	0.7%	0.7%	0.6%	0.5%	0.4%	0.5%	0.6%	0.5%
Boston, MA	3.6%	3.8%	3.9%	4.6%	3.9%	4.2%	4.1%	4.8%	4.0%	4.0%	3.6%	4.3%
Buffalo, NY	0.6%	0.5%	0.4%	0.3%	0.8%	0.6%	0.4%	0.3%	1.1%	0.7%	0.4%	0.3%
Charlotte, NC	0.3%	0.3%	0.3%	0.3%	0.2%	0.2%	0.2%	0.2%	0.1%	0.2%	0.1%	0.2%
Chicago, IL	6.0%	4.6%	3.5%	3.2%	6.9%	5.0%	3.5%	3.0%	5.6%	3.9%	2.9%	2.8%
Cincinnati, OH	1.0%	1.1%	1.0%	1.0%	0.9%	0.9%	0.7%	0.7%	0.7%	1.0%	0.6%	0.6%
Cleveland, OH	2.3%	1.7%	1.3%	1.1%	2.5%	1.5%	1.0%	0.8%	2.5%	1.4%	0.9%	0.6%
Columbus, OH	0.7%	0.5%	0.5%	0.4%	0.6%	0.6%	0.4%	0.3%	0.8%	0.7%	0.3%	0.3%
Dallas-Fort Worth, TX	1.6%	2.0%	2.3%	2.1%	1.1%	1.9%	2.3%	2.2%	1.5%	2.4%	2.9%	2.8%
Denver, CO	1.0%	1.2%	1.3%	1.3%	0.8%	1.0%	0.9%	0.8%	0.8%	1.0%	0.6%	0.5%
Detroit, MI	3.1%	3.3%	2.9%	2.8%	3.1%	3.1%	2.6%	2.6%	3.2%	2.8%	2.5%	2.5%
Greensboro-W.S., NC	0.2%	0.3%	0.3%	0.2%	0.1%	0.2%	0.2%	0.1%	0.2%	0.2%	0.1%	0.1%
Hartford, CT	0.9%	0.9%	0.6%	0.6%	1.0%	0.8%	0.5%	0.5%	0.8%	0.6%	0.3%	0.4%
Houston, TX	2.3%	2.5%	1.9%	2.0%	1.8%	2.3%	1.8%	1.9%	2.2%	2.8%	1.8%	1.9%
Indianapolis, IN	0.8%	0.7%	0.7%	0.5%	0.6%	0.4%	0.4%	0.3%	0.7%	0.5%	0.4%	0.3%
Jacksonville, NC	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Kansas City, MO	0.4%	0.3%	0.4%	0.3%	0.2%	0.2%	0.2%	0.2%	0.2%	0.1%	0.2%	0.2%
Las Vegas, NV	0.1%	0.1%	0.2%	0.3%	0.1%	0.1%	0.2%	0.2%	0.0%	0.1%	0.1%	0.1%
Los Angeles, CA	6.6%	6.1%	6.0%	5.7%	7.2%	7.2%	7.9%	7.3%	6.7%	6.9%	7.5%	7.0%
Memphis, TN	0.1%	0.2%	0.2%	0.3%	0.1%	0.1%	0.1%	0.2%	0.1%	0.1%	0.1%	0.1%
Miami, FL	0.8%	0.9%	0.7%	0.7%	1.0%	1.3%	1.0%	0.9%	0.5%	0.6%	0.5%	0.4%
Milwaukee, WI	1.0%	0.9%	0.8%	0.7%	0.8%	0.8%	0.6%	0.5%	0.5%	0.4%	0.5%	0.4%
Minneap.-St. Paul, MN	1.9%	2.4%	2.7%	2.8%	1.6%	2.0%	2.0%	2.0%	1.5%	1.7%	1.7%	1.8%

Table 3: Ethnic Inventor Contributions by City, continued

	Total Patenting Share				non-English Ethnic Patenting Share				Indian and Chinese Patenting Share			
	1975-1984	1985-1994	1995-2004	2001-2006 (A)	1975-1984	1985-1994	1995-2004	2001-2006 (A)	1975-1984	1985-1994	1995-2004	2001-2006 (A)
Nashville, TN	0.1%	0.2%	0.2%	0.2%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
New Orleans, LA	0.3%	0.2%	0.2%	0.1%	0.3%	0.3%	0.1%	0.1%	0.2%	0.2%	0.0%	0.0%
New York, NY	11.5%	8.9%	7.3%	6.9%	16.6%	13.1%	10.1%	8.9%	16.6%	13.3%	9.7%	9.0%
Norfolk-VA Beach, VA	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Orlando, FL	0.2%	0.3%	0.3%	0.3%	0.1%	0.2%	0.3%	0.3%	0.1%	0.2%	0.3%	0.3%
Philadelphia, PA	4.6%	4.0%	2.7%	2.8%	5.6%	4.9%	2.8%	2.9%	6.2%	5.8%	2.8%	3.0%
Phoenix, AZ	1.0%	1.2%	1.4%	1.3%	0.6%	1.1%	1.3%	1.2%	0.4%	1.0%	1.4%	1.3%
Pittsburgh, PA	2.0%	1.3%	0.8%	0.7%	2.2%	1.4%	0.6%	0.5%	2.2%	1.3%	0.5%	0.5%
Portland, OR	0.5%	0.8%	1.4%	1.6%	0.3%	0.6%	1.4%	1.6%	0.2%	0.6%	1.7%	2.0%
Providence, RI	0.3%	0.3%	0.3%	0.2%	0.3%	0.4%	0.3%	0.2%	0.2%	0.3%	0.2%	0.2%
Raleigh-Durham, NC	0.3%	0.6%	1.1%	1.5%	0.3%	0.6%	1.0%	1.3%	0.3%	0.8%	1.0%	1.2%
Richmond, VA	0.3%	0.3%	0.2%	0.2%	0.3%	0.3%	0.2%	0.2%	0.3%	0.4%	0.2%	0.2%
Sacramento, CA	0.2%	0.4%	0.5%	0.5%	0.2%	0.4%	0.5%	0.5%	0.2%	0.3%	0.5%	0.5%
Salt Lake City, UT	0.4%	0.5%	0.6%	0.6%	0.2%	0.4%	0.3%	0.3%	0.2%	0.3%	0.3%	0.3%
San Antonio, TX	0.1%	0.2%	0.2%	0.2%	0.1%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%
San Diego, CA	1.1%	1.6%	2.2%	2.8%	1.1%	1.6%	2.6%	3.6%	0.8%	1.4%	2.4%	3.9%
San Francisco, CA	4.8%	6.6%	12.1%	13.2%	6.2%	9.3%	19.3%	19.9%	8.4%	13.0%	25.4%	24.0%
Seattle, WA	0.9%	1.3%	1.9%	3.4%	0.8%	1.1%	1.8%	3.5%	0.6%	1.0%	1.8%	3.7%
St. Louis, MO	1.0%	0.9%	0.8%	0.8%	0.9%	0.8%	0.8%	0.7%	1.0%	0.8%	0.4%	0.4%
Tallahassee, FL	0.4%	0.5%	0.4%	0.4%	0.3%	0.4%	0.3%	0.3%	0.2%	0.2%	0.2%	0.2%
Washington, DC	1.5%	1.5%	1.4%	1.6%	1.6%	1.6%	1.5%	1.7%	1.6%	1.7%	1.5%	1.7%
West Palm Beach, FL	0.3%	0.5%	0.4%	0.4%	0.3%	0.5%	0.4%	0.4%	0.3%	0.3%	0.2%	0.2%
Other 234 Major Cities	21.8%	22.3%	20.7%	18.4%	18.1%	18.1%	15.6%	13.6%	19.7%	18.2%	14.6%	12.7%
Not in a Major City	9.0%	8.2%	6.6%	6.2%	6.3%	5.4%	3.7%	4.1%	5.2%	3.8%	2.5%	2.7%

Notes: See Table 1. The first three columns of each grouping are for granted patents. The fourth column, marked with (A), is for published patent applications.

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

Chinese		English		European		Hispanic / Filipino		Indian / Hindi	
CAI	585	ADAMS	4,490	ABEL	269	ACOSTA	171	ACHARYA	338
CAO	657	ALLEN	5,074	ALBRECHT	564	AGUILAR	138	AGARWAL	580
CHAN	3,096	ANDERSON	10,719	ANTOS	230	ALVAREZ	446	AGGARWAL	282
CHANG	3,842	BAILEY	2,431	AUERBACH	193	ANDREAS	128	AGRAWAL	797
CHAO	796	BAKER	4,671	BAER	422	AYER	166	AHMAD	355
CHAU	486	BELL	2,738	BAERLOCHER	252	AYRES	180	AHMED	652
CHEN	12,860	BENNETT	2,734	BAUER	1,470	BALES	240	AKRAM	640
CHENG	2,648	BROOKS	2,015	BECHTEL	179	BLANCO	141	ALI	559
CHEUNG	950	BROWN	11,662	BECK	1,712	BOLANOS	130	ARIMILLI	432
CHIANG	1,112	BURNS	2,098	BENDER	650	BOLES	118	ARORA	214
CHIEN	429	CAMPBELL	3,959	BERG	1,465	CABRAL	154	ASH	290
CHIN	423	CARLSON	2,745	BERGER	1,304	CABRERA	163	BALAKRISHNAN	228
CHIU	924	CARTER	2,658	BOEHM	256	CALDERON	124	BANERJEE	371
CHOU	1,144	CHANG	2,032	BOUTAGHOU	266	CASTANEDA	116	BASU	233
CHOW	1,139	CLARK	5,493	CARON	290	CASTILLO	124	BHAT	224
CHU	2,353	COHEN	2,626	CERAMI	172	CASTRO	119	BHATIA	411
DENG	439	COLE	2,143	CHANDRARATNA	229	CHAVEZ	194	BHATT	242
DING	589	COLLINS	2,992	CHEVALLIER	204	CONTRERAS	137	BHATTACHARYA	216
DONG	492	COOK	3,556	DIETRICH	312	CRUZ	319	BHATTACHARYYA	265
FAN	1,036	COOPER	3,045	DIETZ	496	CUEVAS	123	BOSE	238
FANG	846	COX	2,407	EBERHARDT	192	DAS	213	CHANDRA	221
FENG	658	DAVIS	8,848	EHRLICH	311	DELGADO	216	CHATTERJEE	647
FONG	727	EDWARDS	3,375	ERRICO	190	DIAS	174	DAOUD	305
FU	767	EVANS	4,082	FARKAS	169	DIAZ	584	DAS	522
FUNG	455	FISCHER	2,081	FERRARI	177	DOMINGUEZ	195	DATTA	424
GAO	785	FISHER	2,748	FISCHELL	280	DURAN	142	DE	234
GUO	921	FOSTER	2,616	FUCHS	394	ELIAS	230	DESAI	974
HAN	777	FOX	1,990	GAISER	193	ESTRADA	142	DIXIT	256
HE	1,159	GARDNER	2,412	GELARDI	176	FERNANDES	152	DUTTA	338
HO	1,282	GORDON	2,315	GRILLIOT	201	FERNANDEZ	546	GANDHI	228
HSIEH	980	GRAHAM	2,042	GUEGLER	179	FIGUEROA	146	GARG	345
HSU	3,034	GRAY	2,626	GUNTER	177	FLORES	191	GHOSH	661
HU	1,695	GREEN	3,540	GUNTHER	247	FREITAS	132	GOEL	279
HUANG	4,605	HALL	4,907	HAAS	843	GAGNON	265	GUPTA	1,935
HUI	451	HAMILTON	1,991	HAMPEL	187	GARCIA	1,310	HASSAN	217
HUNG	562	HANSON	2,148	HANSEN	2,947	GARZA	167	HUSSAIN	233
HWANG	800	HARRIS	4,793	HARTMAN	1,214	GOMES	199	HUSSAINI	299
JIANG	1,399	HAYES	2,031	HARTMANN	385	GOMEZ	413	ISLAM	266
KAO	714	HILL	3,590	HAUSE	266	GONSALVES	141	IYER	601
KUO	1,157	HOFFMAN	2,387	HECHT	245	GONZALES	281	JAIN	912
LAI	1,134	HOWARD	2,160	HEINZ	168	GONZALEZ	1,055	JOSHI	886
LAM	1,336	HUGHES	2,198	HORODYSKY	230	GUTIERREZ	601	KAMATH	219
LAU	1,320	JACKSON	3,980	HORVATH	387	GUZMAN	139	KAPOOR	222
LEE	4,006	JENSEN	2,361	IACOVELLI	287	HALASA	202	KHANNA	378
LEUNG	1,165	JOHNSON	17,960	JACOBS	1,962	HERNANDEZ	703	KRISHNAMURTHY	369
LEW	460	JONES	10,630	KARR	196	HERRERA	171	KRISHNAN	512
LI	6,863	KELLER	2,041	KASPER	227	HERRON	450	KULKARNI	299
LIANG	1,173	KELLY	2,775	KEMPF	228	HIDALGO	186	KUMAR	2,005
LIAO	553	KENNEDY	2,208	KNAPP	833	JIMENEZ	246	LAL	366
LIM	485	KING	4,686	KNIFTON	206	LEE	237	MALIK	532
LIN	5,770	KLEIN	2,347	KOENIG	521	LOPEZ	738	MATHUR	306
LING	521	LARSON	2,537	KRESGE	179	MACHADO	135	MEHROTRA	265

Table A1: Most Common US Ethnic Surnames (continued)

Chinese		English		European		Hispanic / Filipino		Indian / Hindi	
LIU	6,406	LEE	9,490	LANGE	757	MARIN	177	MEHTA	925
LO	1,053	LEWIS	4,732	LASKARIS	192	MARQUEZ	117	MENON	325
LU	2,289	LONG	2,392	LEMELSON	324	MARTIN	183	MISHRA	348
LUO	815	MARSHALL	2,088	LIOTTA	171	MARTINEZ	1,112	MISRA	282
MA	1,708	MARTIN	6,773	LORENZ	341	MATIS	249	MOOKHERJEE	272
MAO	545	MILLER	14,942	LUDWIG	500	MEDINA	192	MUKHERJEE	327
NG	1,132	MITCHELL	3,075	LUTZ	679	MENARD	149	MURTHY	236
ONG	473	MOORE	6,459	MAIER	492	MENDOZA	173	NAGARAJAN	270
PAN	1,435	MORGAN	2,824	MARTIN	223	MIRANDA	140	NAIR	560
PENG	530	MORRIS	3,223	MAYER	1,097	MOLINA	129	NARASIMHAN	225
SHEN	1,480	MURPHY	3,609	MEYER	3,004	MORALES	146	NARAYAN	312
SHI	964	MURRAY	2,207	MOLNAR	335	MORENO	128	NARAYANAN	419
SHIH	938	MYERS	2,625	MORIN	320	MUNOZ	177	NATARAJAN	301
SONG	636	NELSON	6,444	MUELLER	2,242	NUNEZ	207	PAREKH	301
SU	1,025	OLSON	3,140	MULLER	985	ORTEGA	206	PARIKH	286
SUN	2,521	PARKER	3,181	NAGEL	383	ORTIZ	362	PATEL	3,879
TAI	463	PETERSON	4,912	NATHAN	171	PADILLA	116	PATIL	352
TAM	589	PHILLIPS	3,875	NILSSEN	234	PAZ DE ARAUJO	148	PRAKASH	326
TAN	1,105	PRICE	2,062	NOVAK	788	PEREIRA	280	PRASAD	549
TANG	2,277	REED	2,645	PAGANO	177	PEREZ	675	PURI	233
TENG	437	RICHARDSON	2,114	PALERMO	177	QUINTANA	126	RAGHAVAN	378
TONG	677	ROBERTS	4,352	PASTOR	238	RAMIREZ	345	RAHMAN	367
TSAI	1,244	ROBINSON	3,741	POPP	202	RAMOS	226	RAJAGOPALAN	396
TSANG	499	ROGERS	2,974	RAO	343	REGNIER	137	RAMACHANDRAN	388
TSENG	538	ROSS	2,377	REITZ	248	REIS	168	RAMAKRISHNAN	270
TUNG	565	RUSSELL	2,611	ROHRBACH	246	REYES	150	RAMAN	222
WANG	11,905	RYAN	2,404	ROMAN	362	RIVERA	489	RAMASWAMY	244
WEI	1,317	SCOTT	3,583	ROSTOKER	245	RODRIGUES	188	RAMESH	364
WEN	455	SHAW	2,369	SCHMIDT	3,753	RODRIGUEZ	1,314	RANGARAJAN	244
WONG	4,811	SIMPSON	2,014	SCHNEIDER	2,246	ROMERO	292	RAO	1,196
WOO	710	SMITH	24,173	SCHULTZ	2,273	RUIZ	297	REDDY	459
WU	5,521	SNYDER	2,335	SCHULZ	921	SALAZAR	179	ROY	279
XIE	609	STEVENS	2,221	SCHWARTZ	2,394	SANCHEZ	717	SANDHU	878
XU	2,249	STEWART	2,924	SCHWARZ	633	SANTIAGO	158	SAXENA	213
YAN	826	SULLIVAN	2,933	SPERANZA	215	SERRANO	172	SHAH	2,467
YANG	4,584	TAYLOR	6,659	SPIEGEL	177	SILVA	457	SHARMA	1,249
YAO	699	THOMAS	5,312	STRAETER	454	SOTO	158	SINGH	2,412
YE	525	THOMPSON	6,424	THEEUWES	247	SOUZA	145	SINGHAL	245
YEE	729	TURNER	2,855	TROKHAN	167	SUAREZ	150	SINHA	463
YEH	928	WALKER	4,887	VOCK	423	TORRES	352	SIRCAR	225
YEN	467	WALLACE	1,963	WACHTER	199	VALDEZ	127	SRINIVASAN	876
YIN	617	WARD	2,913	WAGNER	2,499	VARGA	130	SRIVASTAVA	498
YU	2,293	WATSON	2,139	WEBER	3,003	VASQUEZ	153	SUBRAMANIAN	702
YUAN	825	WHITE	6,190	WEDER	1,067	VAZQUEZ	260	THAKUR	381
ZHANG	4,532	WILLIAMS	10,442	WEISS	1,533	VELAZQUEZ	134	TRIVEDI	383
ZHAO	1,337	WILSON	7,677	WOLF	1,604	VINALS	220	VENKATESAN	281
ZHENG	1,037	WOOD	4,525	WRISTERS	185	YU	140	VERMA	262
ZHOU	1,517	WRIGHT	4,521	ZIMMERMAN	1,542	ZAMORA	120	VISWANATHAN	218
ZHU	1,749	YOUNG	5,957	ZIMMERMANN	226	ZUNIGA	128	VORA	223

Table A1: Most Common US Ethnic Surnames (continued)

Japanese		Korean		Russian		Vietnamese	
AOKI	141	AHN	610	AGHAJANIAN	77	ABOU-GHARBIA	22
AOYAMA	66	BAE	122	ALPEROVICH	64	BAHN	15
ASATO	73	BAEK	77	ALTSHULER	71	BANH	21
CHEN	88	BAK	68	ANDREEV	94	BI	158
DOI	90	BANG	91	ANSCHER	95	BICH	18
FUJII	92	BARK	39	BABICH	79	BIEN	91
FUJIMOTO	98	BYUN	87	BABLER	73	BUI	309
FUKUDA	84	CHA	45	BARINAGA	72	CAN	19
FURUKAWA	218	CHAE	33	BARNA	96	CONG	41
HANAWA	69	CHANG	289	BELOPOLSKY	71	DANG	23
HARADA	90	CHIN	33	BERCHENKO	94	DIEM	24
HASEGAWA	171	CHO	977	BLASKO	79	DIEP	52
HASHIMOTO	110	CHOE	193	BLONDER	82	DINH	232
HAYASHI	148	CHOI	1,081	BONIN	97	DIP	11
HEY	75	CHON	33	CODILIAN	90	DO	13
HIGASHI	98	CHOO	94	COMISKEY	74	DOAN	616
HIGUCHI	81	CHUN	330	DAMADIAN	118	DOMINH	33
HONDA	102	CHUNG	1,499	DANKO	69	DONLAN	21
IDE	136	DROZD	45	DAYAN	143	DOVAN	26
IKEDA	98	EYUBOGLU	36	DERDERIAN	169	DUAN	241
IMAI	129	GANG	34	DOMBROSKI	66	DUE	20
INOUE	90	GU	533	ELKO	81	DUONG	153
IRICK	86	HAHM	42	FETCENKO	62	DUONG-VAN	13
ISHIDA	93	HAHN	1,016	FISHKIN	82	ESKEW	12
ISHII	82	HAM	45	FOMENKOV	73	GRAN	20
ISHIKAWA	208	HAN	145	FRENKEL	71	HAC	20
ITO	260	HANSELL	39	FRIDMAN	67	HAUGAN	16
IWAMOTO	78	HOGLE	43	FROLOV	68	HO	35
KANEKO	157	HONE	78	GARABEDIAN	104	HOANG	277
KATO	113	HONG	907	GELFAND	139	HOPPING	15
KAUTZ	87	HOSKING	63	GINZBURG	73	HUYNH	317
KAWAMURA	87	HUH	32	GITLIN	73	HUYNH-BA	19
KAWASAKI	104	HWANG	108	GLUSCHENKOV	73	KHA	13
KAYA	78	HYUN	54	GORALSKI	69	KHAW	20
KIMURA	108	IM	80	GORDIN	65	KHIEU	35
KINO	74	JANG	46	GORIN	99	KHU	13
KINOSHITA	93	JEON	134	GRINBERG	104	KHUC	15
KIRIHATA	107	JEONG	122	GROCHOWSKI	77	LAHUE	17
KISHI	65	JI	268	GUREVICH	107	LAURSEN	72
KIWALA	132	JIN	673	GURSKY	89	LAVAN	18
KOBAYASHI	296	JO	41	GUZIK	79	LE	1,263
LI	75	JOO	68	HABA	96	LE ROY	29
LIU	84	JU	55	HYNECEK	82	LEEN	75
MAKI	167	JUNG	582	IBRAHIM	229	LEMINH	17
MATSUMOTO	147	KANG	809	IVANOV	165	LUONG	107
MIYANO	70	KIANI	74	IVERS	66	LY	118
MIZUHARA	87	KIM	5,455	JOVANOVIC	65	MINH	41
MORI	128	KO	595	JU	126	NELLUMS	17
MORITA	64	KOO	214	JUHASZ	71	NGO	735
MOSLEHI	165	KUN	63	KAHLE	173	NGUY	12
MOTOYAMA	130	KWAK	96	KAMINSKI	393	NGUYEN	4,720
MURAKAMI	67	KWON	298	KAMINSKY	150	NHO	12

Table A1: Most Common US Ethnic Surnames (continued)

Japanese		Korean		Russian		Vietnamese	
NAJJAR	81	LEE	1,032	KANEVSKY	114	NIEH	69
NAKAGAWA	125	LIM	135	KAPLINSKY	69	NIM	14
NAKAJIMA	99	MENNIE	96	KAPOSI	72	PHAM	901
NAKAMURA	187	MIN	242	KHAN	104	PHAN	27
NAKANISHI	64	NA	34	KHANDROS	161	PHANG	11
NAKANO	104	NAM	68	KHOVAYLO	69	PHY	19
NEMOTO	70	NEVINS	42	KOLMANOVSKY	70	POSTMAN	12
NISHIBORI	88	NYCE	56	KORSUNSKY	153	QUACH	95
NISHIMURA	131	OH	461	KOWAL	74	QUI	11
NODA	107	PAEK	41	LAPIDUS	63	QUY	13
OGAWA	74	PAIK	144	LEE	113	ROCH	26
OGURA	209	PAK	116	LOPATA	113	TA	91
OHARA	269	PARK	2,145	MESSING	74	TAKACH	30
OHKAWA	89	QUAY	107	METLITSKY	95	TAU	23
OKADA	87	RHEE	191	MIKHAIL	115	THACH	33
OKAMOTO	103	RIM	57	MIRKIN	66	THAI	86
ONO	148	RYANG	38	MOGHADAM	72	THAO	21
OVSHINSKY	314	RYU	99	NADELSON	65	THI	13
SAITO	136	SAHM	45	NAZARIAN	75	THIEN	15
SAKAI	79	SAHOO	58	NEMIROVSKY	73	THUT	28
SASAKI	209	SEO	47	NIE	72	TIEDT	14
SATO	231	SHIM	162	OGG	125	TIEP	12
SETO	73	SHIN	399	PAPADOPOULOS	132	TIETJEN	59
SHIMIZU	103	SHINN	96	PAPATHOMAS	67	TO	76
SUZUKI	306	SIN	62	PETROV	102	TON-THAT	16
TAKAHASHI	245	SJOSTROM	39	PINARBASI	131	TRAN	2,050
TAKEUCHI	242	SO	332	PINCHUK	123	TRANDAI	14
TAMURA	83	SOHN	78	POPOV	81	TRANG	34
TANAKA	328	SON	147	PROKOP	86	TRANK	11
THOR	66	SONG	105	RABER	78	TRIEU	49
TSUJI	92	SUE	64	RABINOVICH	123	TRONG	12
TSUKAMOTO	89	SUH	311	ROBICHAUX	65	TRUC	27
UCHIDA	72	SUK	75	RUBSAMEN	69	TU	545
UEDA	72	SUNG	41	SAHATJIAN	66	TUTEN	23
WADA	153	SUR	38	SARKISIAN	65	TUY	16
WANG	81	TOOHEY	33	SARRAF	82	TY	27
WATANABE	416	UM	36	SCHREIER	62	VAN	58
WU	67	WHANG	175	SCHWAN	81	VAN CLEVE	40
YAMADA	180	WON	108	SIMKO	77	VAN DAM	20
YAMAGUCHI	102	YI	237	SMETANA	69	VAN LE	17
YAMAMOTO	432	YIM	145	SOFRANKO	66	VAN NGUYEN	29
YAMASAKI	67	YOHN	32	SOKOLOV	91	VAN PHAN	26
YAMASHITA	105	YOO	290	SORKIN	111	VAN TRAN	15
YAMAZAKI	91	YOON	614	TABAK	85	VIET	11
YANG	65	YOUN	38	TEPMAN	80	VO	269
YASUDA	75	YU	198	TERZIAN	87	VO-DINH	32
YOSHIDA	178	YUH	96	VASHCHENKO	96	VOVAN	20
YUAN	112	YUM	78	WASILEWSKI	80	VU	502
ZHAO	81	YUN	222	ZEMEL	126	VUONG	107