

# The Ethnic Composition of US Inventors

William R. Kerr\*  
Harvard Business School  
Boston MA

HBS Working Paper 08-006  
23 August 2007

## 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 increase dramatically in the 1990s, before noticeably leveling off after 2000 and declining in the case of India. 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](mailto: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.<sup>1</sup> 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.<sup>2</sup>

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.<sup>3</sup> 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 April 2007.<sup>4</sup> Each patent record lists one or more inventors, with 7.5 million inventor names associated with the 4.3 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 1975-2007 period.

This study maps into these inventor names an ethnic-name database typically used for commercial applications.<sup>5</sup> 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, 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,

---

<sup>1</sup>For example, Stephan and Levin (2001), Burton and Wang (1999), Johnson (1998, 2001), and Streeter (1997).

<sup>2</sup>Representative papers are Lowell (2000), Borjas (2004), Saxenian (2002b), and Freeman (2005) respectively.

<sup>3</sup>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).

<sup>4</sup>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 early 2007. Some of the descriptive calculations have not been updated from their 1975-1999 values (noted in text).

<sup>5</sup>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.

cities, companies, and so on.<sup>6</sup>

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, 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.<sup>7</sup>

---

<sup>6</sup>This ethnic patenting database is employed by Kerr (2005, 2007a-d), Kerr and Lincoln (2007), and Foley and Kerr (2007) to study the role of ethnic scientists and entrepreneurs in technology formation and diffusion.

<sup>7</sup>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 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.<sup>8</sup> The match rate for this restricted procedure is 97% (98% US, 97% 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 88% (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 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

---

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.

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.

<sup>8</sup>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.

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 48% 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 but India (84%). In the Restricted Matching procedure, a matching rate of greater than 73% 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.<sup>9</sup>

---

<sup>9</sup>The main US SE ethnicity missing from the database is Iranian. Running the ethnic-name database on the few patents from Iran yields a 57%-76% match rate. Iran’s predicted composition does not favor any of the nine ethnicities studied, with the largest overlap being the English ethnicity at 55%.

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 fifty most common surnames of US-based inventors for each ethnicity, along with their relative contributions.<sup>10</sup> 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 scientists and engineers with previously unavailable detail: incorporating the major ethnicities working in the US SE community; separating out detailed technologies and manufacturing industries; providing MSA 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.

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

---

<sup>10</sup>The counts sum the ethnic contribution from 1975-1999 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).

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.<sup>11</sup> 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.<sup>12</sup> 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 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.<sup>13</sup> 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 72% 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 8% and 4%, respectively. As shown below, this Chinese and Indian growth is concentrated in high-tech sectors, where Chinese inventors supplant European researchers as

---

<sup>11</sup>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.

<sup>12</sup>The current patent data incorporate all patents granted by April 2007. 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 of algorithms to identify individual scientists with the USPTO data.

<sup>13</sup>The estimated European contribution in Table 2 is naturally higher than the immigrant contribution measured by foreign born.

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 5% 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 1.3% over the thirty years, while the Russian climbs from 1.3% 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 0.5% to 1.2%. Finally, while the Vietnamese contribution is the lowest throughout the sample, it does exhibit the strongest relative growth from 0.1% to 0.7%.

### 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 the most interesting and economically meaningful variation across technologies, as summarized in Figures 9 and 10.<sup>14</sup>

### 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

---

<sup>14</sup>The USPTO issues patents by technology categories rather than by industries. Combining the work of Johnson (1999), Silverman (1999), and Kerr (2007a), 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.

high-tech sectors.<sup>15</sup>

### 3.4 Contributions by Geography

This paper closes its descriptive statistics with an examination of the 1975-2004 ethnic inventor contributions by major MSAs.<sup>16</sup> The first three columns of Table 3 document each city's share of US patenting. Not surprisingly, these shares are highly correlated with city size, with the three largest patenting centers for 1995-2004 found in San Francisco (12%), New York (7%), and Los Angeles (6%). Comparing these total patenting percentages with the ethnic patenting shares, listed in the second set of three columns, reveals the more interesting fact that ethnic patenting concentration is higher than the concentration of general innovation. The 1995-2004 ethnic patent shares of San Francisco, New York, and Los Angeles are 19%, 10%, and 8%, respectively. Similarly, 80% of ethnic research occurs in the major MSAs listed in Table 3, compared to 72% of total patenting. The final three columns list the Chinese patenting share by MSA, highlighting the exceptional growth of San Francisco from 10% of 1975-1984 patenting to 28% in 1995-2004.

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 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.<sup>17</sup>

## 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

---

<sup>15</sup>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 representativeness 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.

<sup>16</sup>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.

<sup>17</sup>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.

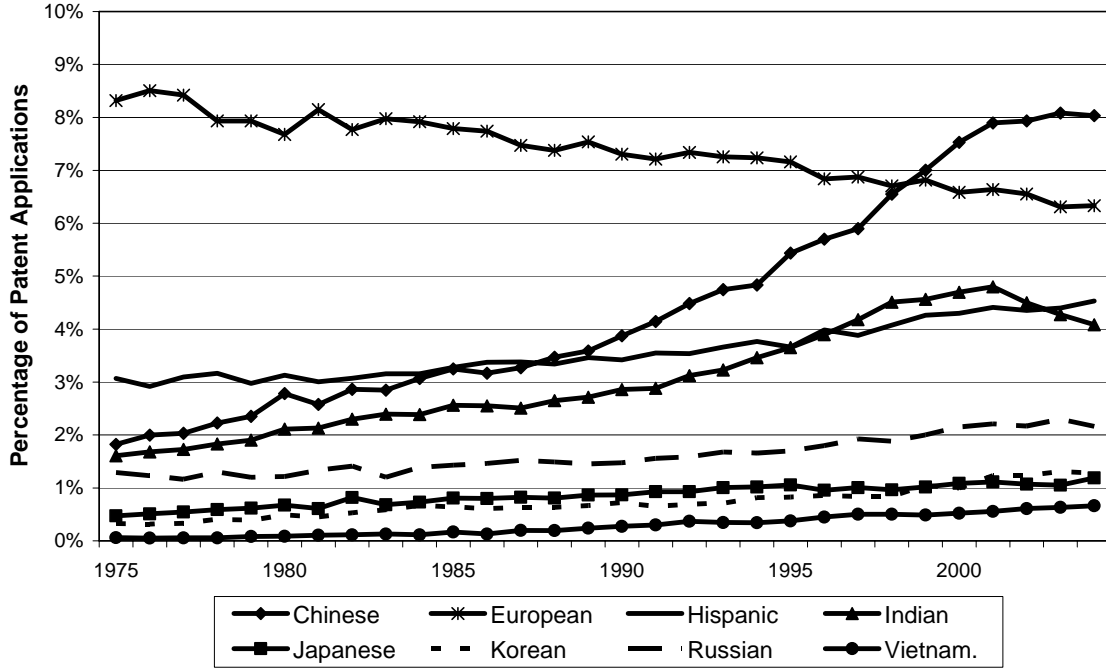
patent records. The ethnic-name database this study employs distinguishes nine ethnic groups, and the matched database describes the ethnic composition of US inventors with previously unavailable cross-sectional and longitudinal detail. This richer variation can support more detailed and informative empirical analyses than feasible otherwise.

## References

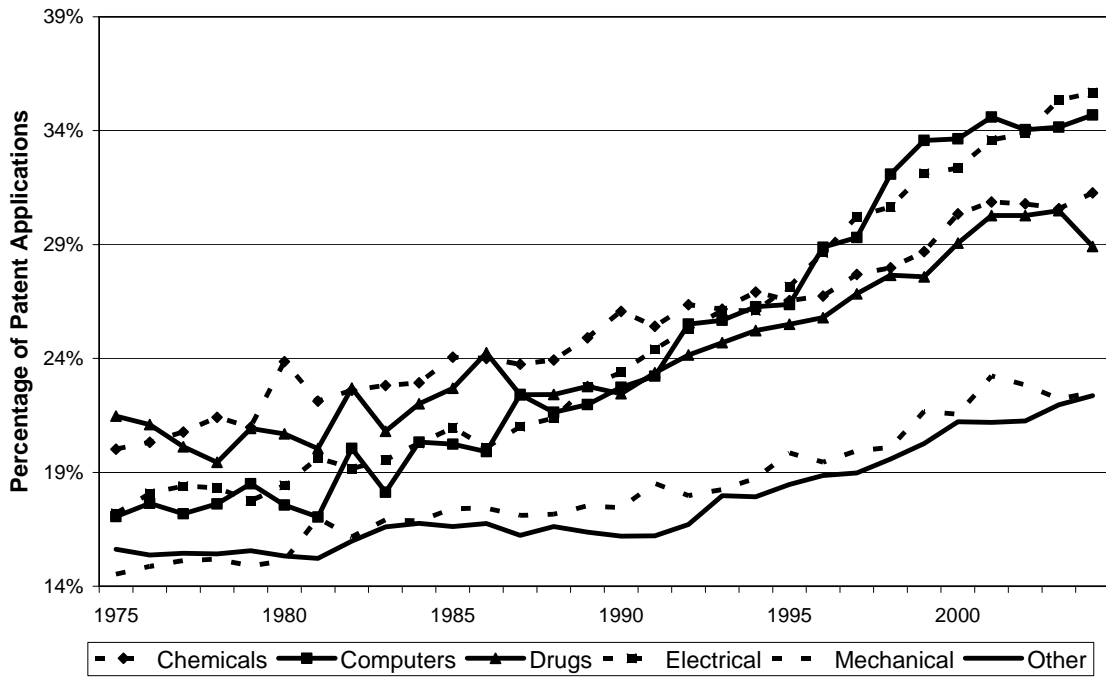
- [1] Borjas, George, "Do Foreign Students Crowd Out Native Students from Graduate Programs?", NBER Working Paper 10349 (2004).
- [2] Branstetter, Lee, and Yoshiaki Ogura, "Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations", NBER Working Paper 11561 (2005).
- [3] Burton, Lawrence, and Jack Wang, "How Much Does the U.S. Rely on Immigrant Engineers?", NSF SRS Issue Brief (1999).
- [4] Foley, C. Fritz, and William Kerr, "US Ethnic Scientists and Foreign Direct Investment Placement", Working Paper (2007).
- [5] Freeman, Richard, "Does Globalization of the Scientific/Engineering Workforce Threaten U.S. Economic Leadership?", NBER Working Paper 11457 (2005).
- [6] Griliches, Zvi, "Patent Statistics as Economic Indicators: A Survey", *Journal of Economic Literature* 28:4 (1990), 1661-1707.
- [7] Hall, Bronwyn, "Exploring the Patent Explosion", *Journal of Technology Transfer* 30 (2005), 35-48.
- [8] Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools", NBER Working Paper 8498 (2001).
- [9] Jaffe, Adam, and Joshua Lerner, *Innovation and Its Discontents* (Boston, MA: Harvard Business School Press, 2005).
- [10] 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>.
- [11] Johnson, Jean, "Statistical Profiles of Foreign Doctoral Recipients in Science and Engineering: Plans to Stay in the United States", NSF SRS Report (1998).
- [12] Johnson, Jean, "Human Resource Contribution to U.S. Science and Engineering From China", NSF SRS Issue Brief (2001).
- [13] Kannankutty, Nirmala, and R. Keith Wilkinson, "SESTAT: A Tool for Studying Scientists and Engineers in the United States", NSF SRS Report (1999).
- [14] Keller, Wolfgang, "Trade and the Transmission of Technology", *Journal of Economic Growth* 7 (2002), 5-24.
- [15] Kerr, William, "Ethnic Scientific Communities and International Technology Diffusion", *Review of Economics and Statistics* forthcoming (2007a).
- [16] Kerr, William, "Ethnic Scientific Networks and the Facilitation of Trade", Working Paper (2007b).
- [17] Kerr, William, "Heterogeneous Technology Diffusion and Ricardian Trade Patterns", Working Paper (2007c).
- [18] Kerr, William, "The Agglomeration of US Ethnic Inventors", Working Paper (2007d).

- [19] Kerr, William, "The Role of Immigrant Scientists and Entrepreneurs in International Technology Transfer", MIT Ph.D. Dissertation (2005).
- [20] Kerr, William, and William Lincoln, "The Impact of H1-B Visa Reforms on the Ethnic Composition of US Inventors", Working Paper (2007).
- [21] 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.
- [22] Kortum, Samuel, and Joshua Lerner, "Assessing the Contribution of Venture Capital to Innovation", *RAND Journal of Economics* 31:4 (2000), 674-692.
- [23] Lowell, B. Lindsay, "H1-B Temporary Workers: Estimating the Population", The Center for Comparative Immigration Studies Working Paper 12 (2000).
- [24] 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).
- [25] Saxenian, AnnaLee, "Silicon Valley's New Immigrant High-Growth Entrepreneurs", *Economic Development Quarterly* 16:1 (2002b), 20-31.
- [26] 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).
- [27] 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.
- [28] 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.
- [29] Streeter, Joanne, "Major Declines in Admissions of Immigrant Scientists and Engineers in Fiscal Year 1994", NSF SRS Issue Brief (1997).
- [30] Trajtenberg, Manuel, "The Mobility of Inventors and the Productivity of Research", Working Paper (2005).

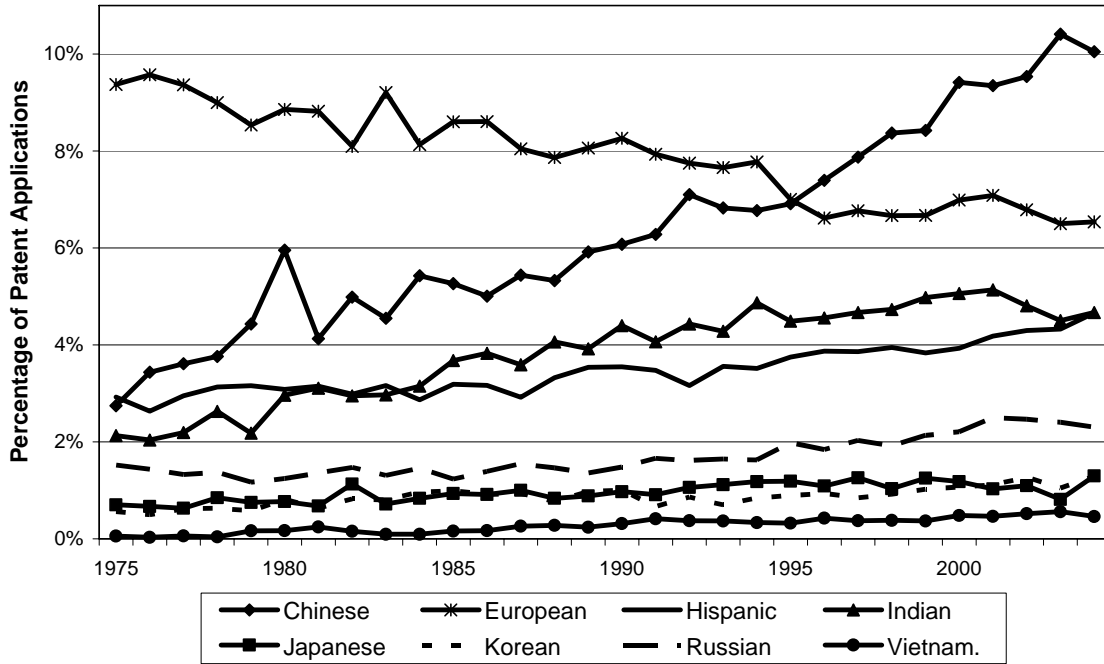
**Fig. 1: Ethnic Share of US Domestic Patents**



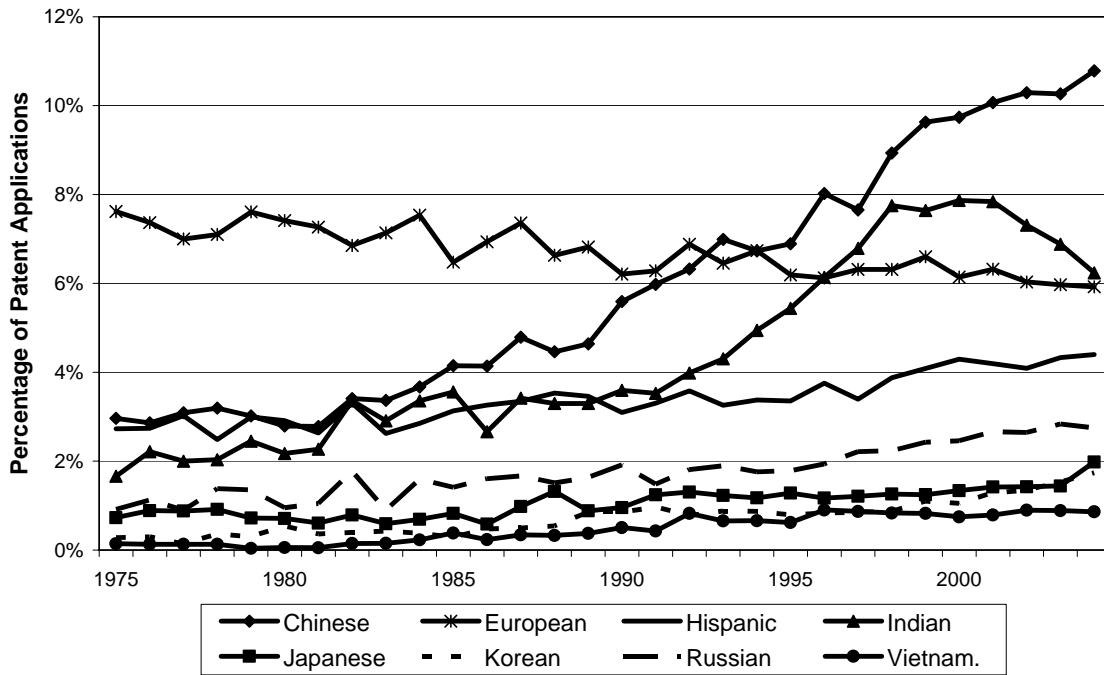
**Fig. 2: Total US Ethnic Share by Technology**



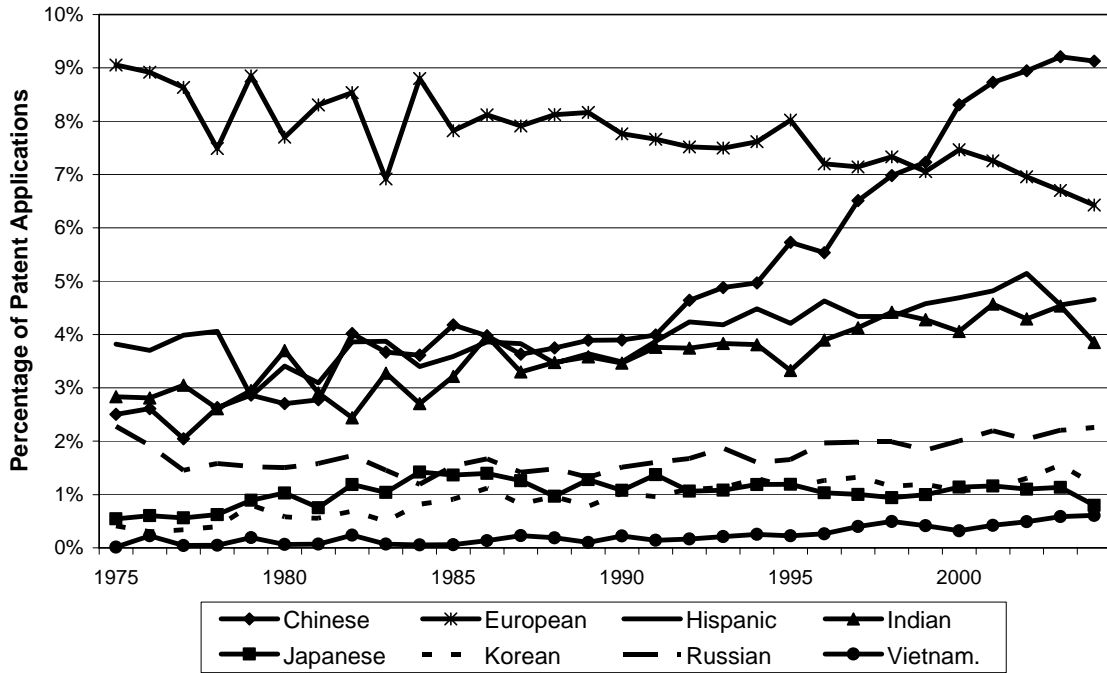
**Fig. 3: US Ethnic Patenting - Chemicals**



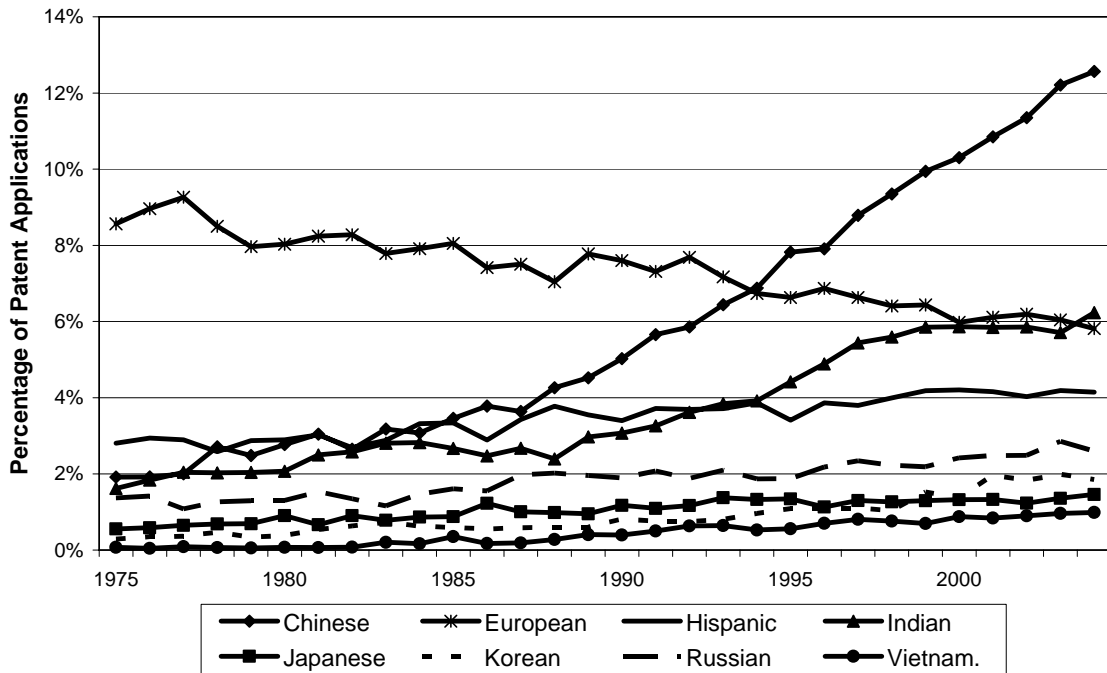
**Fig. 4: US Ethnic Patenting - Computers**



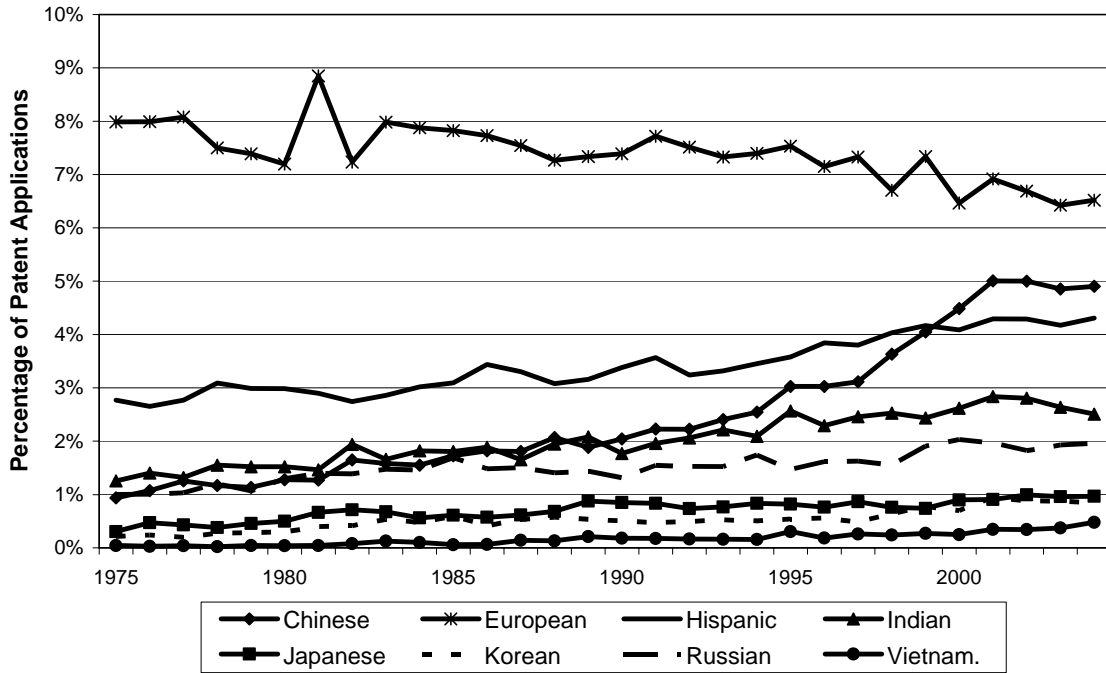
**Fig. 5: US Ethnic Patenting - Drugs**



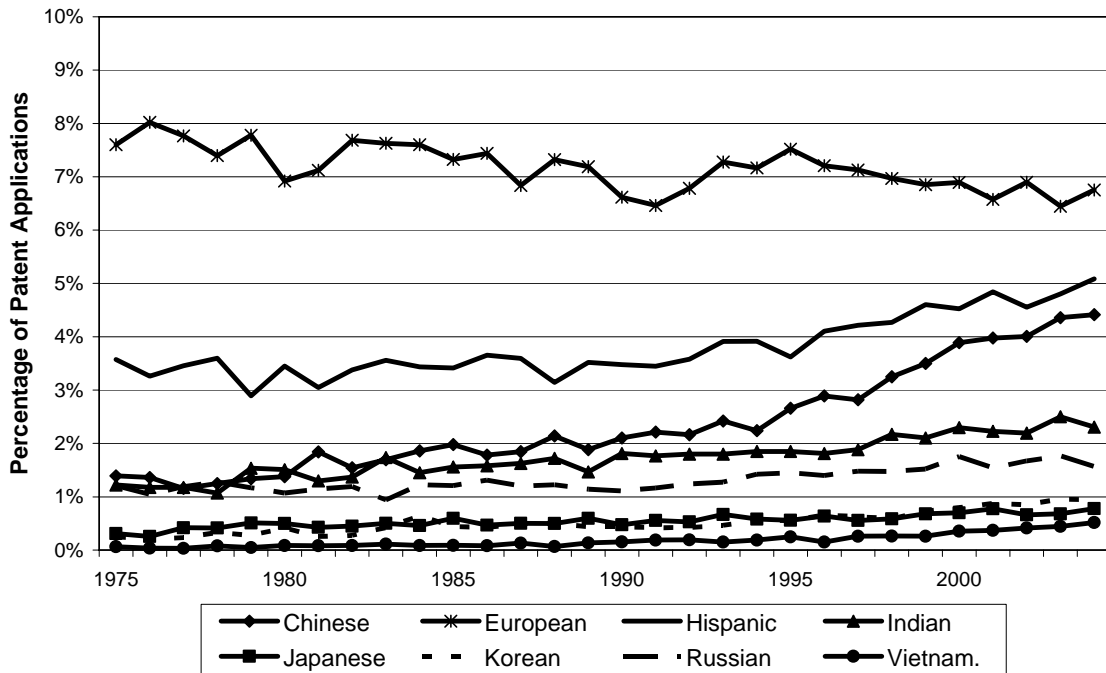
**Fig. 6: US Ethnic Patenting - Electrical**



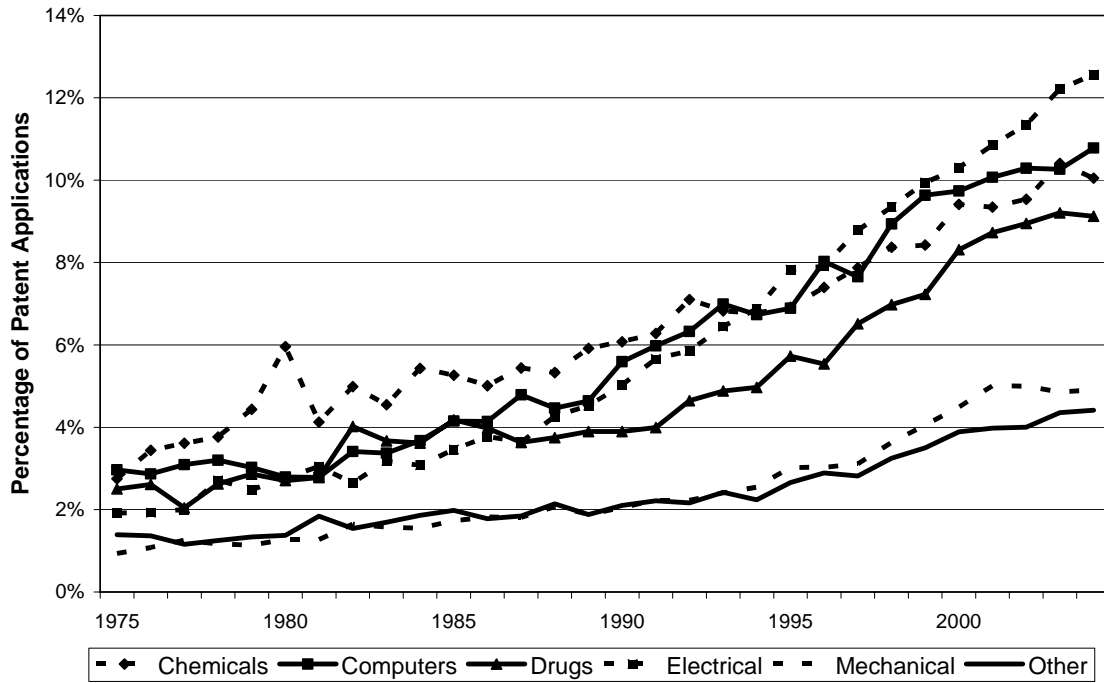
**Fig. 7: US Ethnic Patenting - Mechanical**



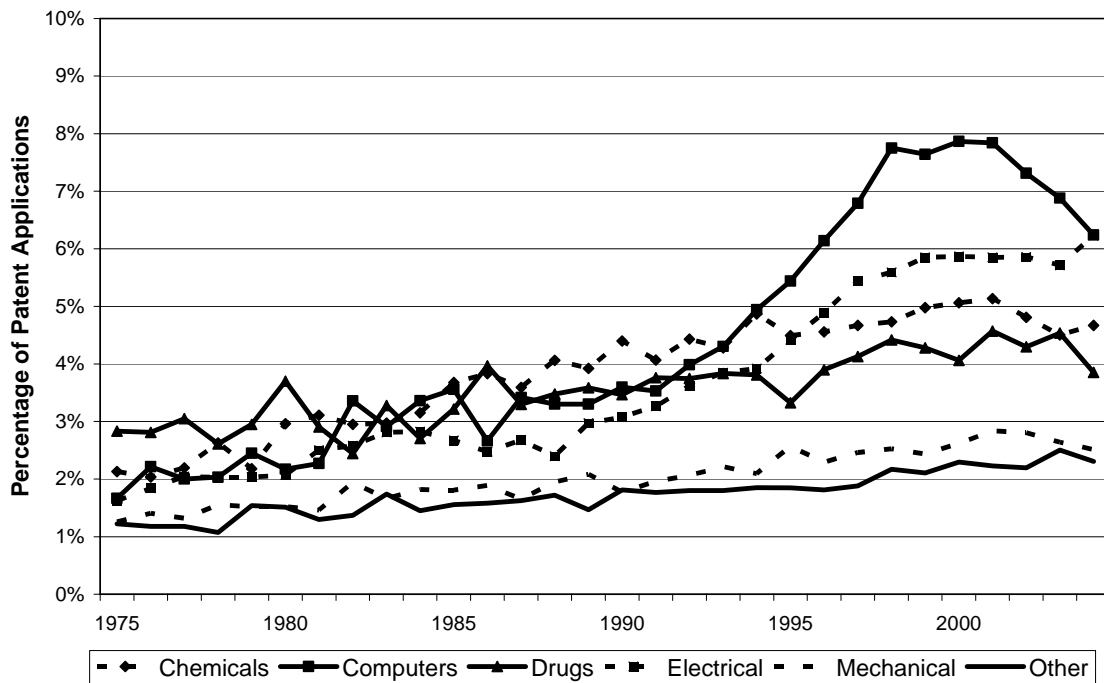
**Fig. 8: US Ethnic Patenting - Other**



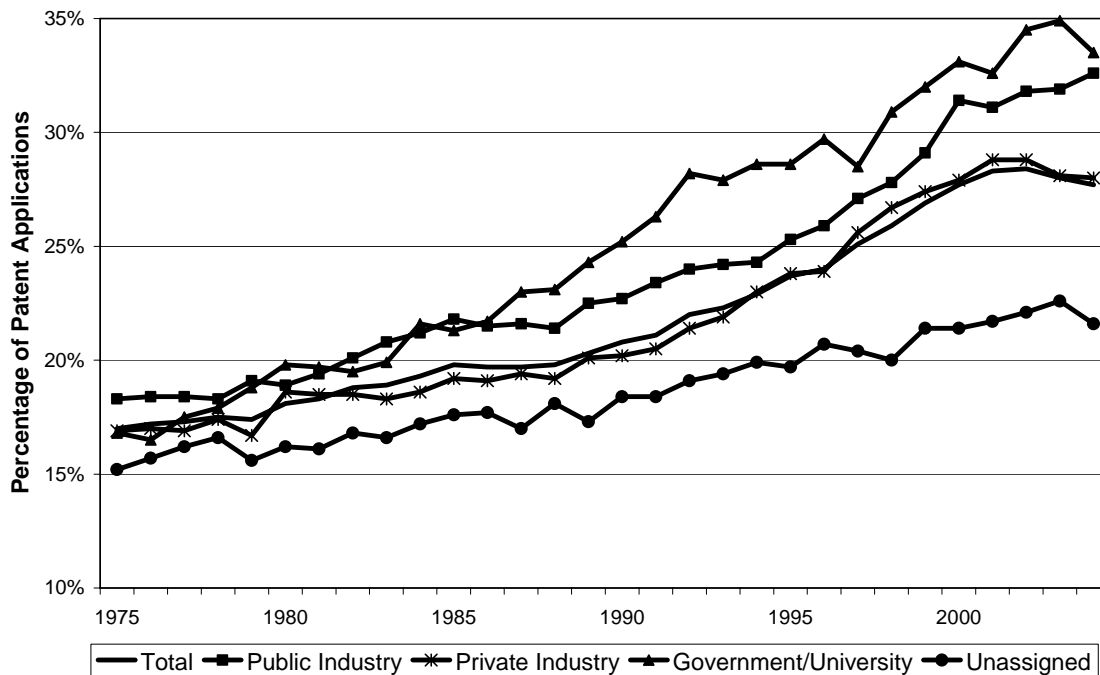
**Fig. 9: Chinese Contribution by Technology**



**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								
	Obs.	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)		
		Full	Restrict.	Full	Restrict.	Full	Restrict.	
United Kingdom	175,077	99%	95%	85%	83%	92%	91%	
China, Singapore	136,818	100%	98%	89%	89%	91%	91%	
Western Europe	1,137,751	97%	80%	66%	46%	73%	58%	
Hispanic Nations	24,793	98%	73%	74%	68%	93%	93%	
India	11,056	84%	77%	83%	88%	85%	89%	
Japan	1,691,337	98%	89%	100%	96%	100%	96%	
South Korea	100,140	100%	100%	83%	83%	87%	88%	
Russia	32,128	93%	80%	80%	84%	93%	94%	
Vietnam	34	100%	97%	42%	52%	50%	52%	

Complete Ethnic Composition of Region's Inventors (Full Matching)									
	English	Chinese	European	Hispanic	Indian	Japanese	Korean	Russian	Vietnam.
United Kingdom	<b>85%</b>	2%	5%	3%	2%	0%	0%	2%	0%
China, Singapore	2%	<b>89%</b>	1%	1%	1%	1%	5%	0%	1%
Western Europe	21%	1%	<b>66%</b>	8%	1%	0%	0%	3%	0%
Hispanic Nations	12%	1%	10%	<b>74%</b>	0%	1%	0%	2%	0%
India	5%	1%	2%	7%	<b>83%</b>	0%	0%	2%	0%
Japan	0%	0%	0%	0%	0%	<b>100%</b>	0%	0%	0%
South Korea	2%	12%	0%	0%	0%	1%	<b>83%</b>	1%	0%
Russia	5%	1%	3%	10%	0%	0%	0%	<b>80%</b>	0%
Vietnam	17%	24%	14%	0%	0%	3%	0%	0%	<b>42%</b>

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									
1975-1979	82.5%	2.2%	8.2%	3.0%	1.9%	0.6%	0.4%	1.2%	0.1%
1980-1984	81.1%	2.9%	7.9%	3.1%	2.4%	0.7%	0.6%	1.3%	0.1%
1985-1989	79.8%	3.6%	7.5%	3.3%	2.8%	0.8%	0.7%	1.4%	0.2%
1990-1994	77.6%	4.7%	7.2%	3.5%	3.4%	0.9%	0.8%	1.5%	0.4%
1995-1999	74.0%	6.6%	6.8%	3.9%	4.5%	0.9%	0.9%	1.8%	0.5%
2000-2004	71.0%	8.5%	6.4%	4.2%	4.8%	1.0%	1.2%	2.2%	0.6%
Chemicals	73.7%	7.1%	7.6%	3.6%	4.2%	0.9%	0.9%	1.7%	0.3%
Computers	71.3%	7.9%	6.3%	3.7%	6.1%	1.1%	1.0%	2.0%	0.7%
Pharmaceuticals	73.3%	6.9%	7.4%	4.3%	3.9%	1.1%	1.0%	1.8%	0.3%
Electrical	72.0%	8.0%	6.8%	3.7%	4.6%	1.1%	1.2%	2.0%	0.7%
Mechanical	80.6%	3.2%	7.2%	3.4%	2.4%	0.7%	0.6%	1.6%	0.2%
Miscellaneous	81.5%	2.9%	7.0%	3.8%	2.1%	0.6%	0.6%	1.4%	0.2%
Top MSAs as a	KC (89)	SF (14)	NOR (12)	MIA (16)	AUS (6)	SF (2)	BAL (2)	BOS (3)	AUS (2)
Percentage of MSA's	WS (88)	LA (8)	STL (11)	SA (9)	SF (6)	SD (2)	LA (2)	NYC (3)	SF (1)
Patents	NAS (88)	AUS (6)	NYC (11)	WPB (7)	BUF (5)	LA (2)	SF (2)	SF (3)	PRT (1)
B. Ethnic Scientist and Engineer Shares Estimated from 1990 US Census Records									
Bachelors Share	87.6%	2.7%	2.3%	2.4%	2.3%	0.6%	0.5%	0.4%	1.2%
Masters Share	78.9%	6.7%	3.4%	2.2%	5.4%	0.9%	0.7%	0.8%	1.0%
Doctorate Share	71.2%	13.2%	4.0%	1.7%	6.5%	0.9%	1.5%	0.5%	0.4%

Notes: MSAs - AUS (Austin), BAL (Baltimore), BOS (Boston), BUF (Buffalo), 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). 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 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 1; English provides a residual in the Census statistics.

**Table 3: Ethnic Inventor Contributions by MSA**

	Total Patenting Share			Ethnic Patenting Share			Chinese Patenting Share		
	1975-1984	1985-1994	1995-2004	1975-1984	1985-1994	1995-2004	1975-1984	1985-1994	1995-2004
Atlanta, GA	0.6%	1.0%	1.3%	0.3%	0.7%	1.0%	0.3%	0.6%	0.8%
Austin, TX	0.4%	0.9%	1.8%	0.5%	1.2%	1.9%	0.5%	1.6%	1.7%
Baltimore, MD	0.8%	0.8%	0.7%	0.7%	0.7%	0.7%	0.4%	0.5%	0.6%
Boston, MA	3.6%	3.8%	3.9%	3.9%	4.2%	4.1%	3.5%	4.1%	3.5%
Buffalo, NY	0.6%	0.5%	0.4%	0.8%	0.6%	0.4%	0.9%	0.5%	0.3%
Charlotte, SC	0.3%	0.3%	0.3%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%
Chicago, IL	6.0%	4.6%	3.5%	6.9%	5.1%	3.5%	5.0%	3.9%	2.9%
Cincinnati, OH	1.0%	1.1%	1.0%	0.9%	0.9%	0.7%	0.8%	0.8%	0.6%
Cleveland, OH	2.3%	1.7%	1.4%	2.5%	1.5%	1.0%	2.0%	1.2%	0.8%
Columbus, OH	0.7%	0.5%	0.5%	0.6%	0.6%	0.4%	0.6%	0.4%	0.3%
Dallas-Fort Worth, TX	1.6%	2.0%	2.2%	1.1%	1.9%	2.3%	1.4%	2.4%	2.5%
Denver, CO	1.1%	1.2%	1.3%	0.8%	1.0%	0.9%	0.9%	1.2%	0.5%
Detroit, MI	3.1%	3.3%	3.0%	3.1%	3.1%	2.7%	2.4%	1.9%	2.4%
Greensboro-W. Salem, NC	0.2%	0.3%	0.3%	0.1%	0.2%	0.2%	0.2%	0.2%	0.1%
Hartford, CT	0.9%	0.9%	0.6%	0.9%	0.8%	0.5%	0.5%	0.5%	0.3%
Houston, TX	2.3%	2.5%	1.9%	1.8%	2.3%	1.8%	2.0%	3.2%	1.7%
Indianapolis, IN	0.8%	0.7%	0.7%	0.6%	0.5%	0.4%	0.6%	0.4%	0.4%
Jacksonville, NC	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%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%
Las Vegas, NV	0.1%	0.1%	0.2%	0.1%	0.1%	0.2%	0.0%	0.0%	0.1%
Los Angeles, CA	6.6%	6.1%	6.1%	7.1%	7.2%	8.2%	7.4%	8.7%	10.0%
Memphis, TN	0.1%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Miami, FL	0.8%	0.9%	0.7%	1.0%	1.3%	1.0%	0.4%	0.5%	0.3%
Milwaukee, WI	1.0%	0.9%	0.8%	0.8%	0.8%	0.6%	0.4%	0.4%	0.5%
Minneapolis-St. Paul, MN	1.9%	2.4%	2.6%	1.6%	2.0%	1.9%	1.5%	1.6%	1.8%

**Table 3: Ethnic Inventor Contributions by MSA (continued)**

	Total Patenting Share			Ethnic Patenting Share			Chinese Patenting Share		
	1975-1984	1985-1994	1995-2004	1975-1984	1985-1994	1995-2004	1975-1984	1985-1994	1995-2004
Nashville, TN	0.1%	0.2%	0.2%	0.0%	0.1%	0.1%	0.0%	0.1%	0.1%
New Orleans, LA	0.3%	0.2%	0.2%	0.3%	0.3%	0.1%	0.2%	0.2%	0.0%
New York, NY	11.5%	8.9%	7.2%	16.6%	13.1%	9.9%	16.6%	12.5%	8.8%
Norfolk-VA Beach, VA	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Orlando, FL	0.2%	0.3%	0.3%	0.1%	0.2%	0.3%	0.1%	0.1%	0.3%
Philadelphia, PA	4.6%	3.9%	2.7%	5.7%	4.9%	2.8%	7.2%	6.1%	2.9%
Phoenix, AZ	1.0%	1.2%	1.4%	0.6%	1.1%	1.3%	0.4%	0.9%	1.3%
Pittsburgh, PA	2.0%	1.3%	0.8%	2.2%	1.4%	0.6%	2.1%	1.0%	0.5%
Portland, OR	0.5%	0.8%	1.4%	0.3%	0.6%	1.3%	0.1%	0.6%	1.4%
Providence, RI	0.3%	0.3%	0.3%	0.3%	0.4%	0.3%	0.1%	0.3%	0.2%
Raleigh-Durham, NC	0.3%	0.6%	1.1%	0.3%	0.6%	1.0%	0.3%	0.6%	0.8%
Richmond, VA	0.3%	0.3%	0.2%	0.3%	0.3%	0.2%	0.5%	0.4%	0.2%
Sacramento, CA	0.2%	0.4%	0.5%	0.2%	0.4%	0.5%	0.2%	0.3%	0.4%
Salt Lake City, UT	0.4%	0.5%	0.6%	0.2%	0.4%	0.3%	0.2%	0.3%	0.3%
San Antonio, TX	0.1%	0.2%	0.2%	0.1%	0.2%	0.2%	0.1%	0.1%	0.1%
San Diego, CA	1.1%	1.6%	2.1%	1.1%	1.6%	2.5%	1.1%	1.5%	2.4%
San Francisco, CA	4.8%	6.6%	11.8%	6.2%	9.4%	19.1%	10.1%	15.9%	28.1%
Seattle, WA	0.9%	1.3%	1.8%	0.8%	1.1%	1.7%	0.6%	1.0%	1.6%
St. Louis, MO	1.0%	0.9%	0.8%	0.9%	0.8%	0.7%	1.2%	0.9%	0.4%
Tallahassee, FL	0.4%	0.5%	0.4%	0.3%	0.4%	0.3%	0.1%	0.2%	0.1%
Washington, DC	1.5%	1.5%	1.4%	1.6%	1.6%	1.4%	1.5%	1.6%	1.5%
West Palm Beach, FL	0.3%	0.5%	0.4%	0.3%	0.5%	0.4%	0.3%	0.3%	0.2%
Other MSAs	21.8%	22.2%	20.9%	18.2%	18.1%	15.8%	19.5%	16.6%	13.3%
Not in a MSA	9.1%	8.2%	6.9%	6.3%	5.4%	4.0%	5.0%	3.6%	2.7%

**Table 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 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 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 A1: Most Common US Ethnic Surnames (continued)**

<b>Japanese</b>		<b>Korean</b>		<b>Russian</b>		<b>Vietnamese</b>	
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