Social Capital and Business Development in High-Technology Clusters

INTERNATIONAL STUDIES IN ENTREPRENEURSHIP

Series Editors:

Zoltan J. Acs
George Mason University
Fairfax, VA USA

David B. Audretsch
Max Planck Institute of Economics
Jena, Germany

Other books in the series:

Black, G.
The Geography of Small Firm Innovation
Tubke, A.
Success Factors of Corporate Spin-Offs
Corbetta, G., Huse, M., Ravasi, D.
Crossroads of Entrepreneurship
Hansen, T., Solgaard, H.S.
New Perspectives in Retailing and Store Patronage Behavior
Davidsson, P.
Researching Entrepreneurship
Fornahl, D., Audretsch D., Zellner, C.
The Role of Labour Mobility and Informal Networks for Knowledge Transfer
Audretsch D., Grimm, H., Wessner, C.
Local Heroes in the Global Village
Landström, H.
Pioneers in Entrepreneurship and Small Business Research
Lundström, A., Stevenson, L.
Entrepreneurship Policy: Theory and Practice
Elfring, T.
Corporate Entrepreneurship
van Stel, A.
Empirical Analysis of Entrepreneurship and Economic Growth
Fritsch, M., Schmude, J.
Entrepreneurship in the Region
Reynolds, P. D.
Entrepreneurship in the United States
Congregado, E.
Measuring Entrepreneurship: Building a Statistical System
Acs, Z., Stough, R.
Public Policy in an Entrepreneurial Economy: Creating the Conditions for Business Growth
Aydogan, N., Chen, Y.P.
Social Capital and Business Development in High-Technology Clusters: An Analysis of Contemporary U.S. Agglomerations

Neslihan Aydogan · Yiu Por Chen

Social Capital and Business Development in High-Technology Clusters

An Analysis of Contemporary U.S. Agglomerations

Springer

Neslihan Aydogan
Cankaya University
Department of Economics
Ankara
Turkey
aydongan@cankaya.edu.tr

Yiu Por Chen
DePaul University
Management of Public Services Program
Chicago, IL
USA
ychen16@depaul.edu

ISBN: 978-0-387-71910-8 e-ISBN: 978-0-387-71911-5
DOI: 10.1007/978-0-387-71911-5

Library of Congress Control Number: 2008926738

Printed on acid-free paper

springer.com

To my mother, Yucel Aydogan, and my father, Muammer Aydogan, for their constant support and love, and to the joy of our lives,
M. Akin Aydogan

Neslihan Aydogan

To my mother, Chow Yee-har, in grateful memory

Yiu Por Chen

Acknowledgments

I gratefully acknowledge my advisor, Kenneth A. Small, at the University of California, Irvine, who provided me with invaluable teachings, support and guidance throughout my years at the graduate school. With great patience, Professor Small supported me through my venture into a very different topic than his own by sending me to conferences, helping me meet with the pathbreakers in the field and with access to data to complete my thesis. I also would like to thank Joshua Lerner at Harvard University for his invaluable advice and support, without which I would not have excelled in academia and publishing. Ashish Arora taught me a lot during my visit to the Carnegie Mellon University, Heinz School of Economics and Public Policy. David B. Audretsch provided me with invaluable advice and teachings during my appointment as a Visiting Assistant Professor at Indiana University, Kelley School of Business, which greatly helped me to mature as an academic. I also owe greatly to my Dean, Ahmet Yalniz, whose creativity, encouragement and support helped to pave my way to the light. It was an honor to have met these scholars. I would also like to thank my family. Since I was a little girl my dad taught me that sky is the limit as long as I really want something and work for it. He put me through the best schools in Turkey and supported me with his great optimism over the years, while I was in graduate school in the United States. He is one man who deserves the credit for everything I have become. My mom has been my inspirational angel, my greatest supporter who believed in me at every step of the way. Her strength and courage I inherited have been the light that showed me the way. Finally, the joy of our family, my nephew, with his great intellect, sense of humor and kindness, has taught me things that are quite impressive for a ten-year-old and which supported me during the writing process. I would particularly like to thank to our editor, Nicholas Philipson, without whose encouragement, optimism, support, and teaching this book would not have published.

Neslihan Aydogan

I gratefully acknowledge Elliott Sclar at Columbia University and Richard Freeman at Harvard University, who provided me with years of invaluable guidance, advice, and support during graduate school. Without their support, I would not have finished my graduate study, not to mention this book. During the course of this writing, the help from two prestigious organizations made my portion of the book possible. Through the National Bureau of Economic Research, I received numerous opportunities to explore my academic curiosity in an intellectually stimulating environment; also through NBER's support I was granted the license to utilize the individual level 2001 Survey of Doctorate Recipients (SDR) from National Science Foundation. Certainly, the permission to use the individual-level data provided by the National Science Foundation was indispensable. During the academic year 2003–2004, an Andrew W. Mellon Post-doctoral Fellowship on Asian Urbanization at the Population Studies and Training Center of Brown University provided an excellent environment in which to begin writing this book project. During the course of this project, my thinking was stimulated by discussions with and support from many colleagues; they include Wood Bowman, Ronald Fernandes, Zai Liang, Carl Riskin, Victor Shih, and Michael White. I would also like to acknowledge my family. Perhaps the most important person in my life is my mother, who almost forfeited her precious life to her first baby, Yiu Por. Since then, her health declined and she eventually passed away when I was eight. Left with three little sons, my father bravely took up the grave responsibility of raising three children. Twenty-some years past, my younger brothers grew up and supported the family, which freed me from financial worries during my graduate studies. Last, but not the least, while the book project carries over bit by bit, without the patience, support, and encouragement from our book editor, Nicholas Philipson, this book would not be published.

Yiu Por Chen

Contents

Introduction

Neslihan Aydogan

It is now widely accepted that geographical proximity matters to economic and social life. Not only does it provide externalities that reduce transaction costs, but it also helps transacting parties form networks from which they can benefit greatly. More than ever, economists have been trying to bridge the development gap between poor and rich regions. Agglomerations are investigated toward that target because economists think that if certain regions can mix up that special formula to continuously innovate and produce, then understanding and forming such agglomerations could be one way to go about that growth path. We have designed the chapters of this book to work out the mechanics of geographical agglomerations in the United States with the focus of identifying the characteristics of such special formula

Chapters 1–3 are designed to investigate the high-tech clusters that have sprung up in the United States due to their innovative capacity to engage in high-value-added activities. The first question we ask is, What promotes the productivity of high-tech firms? We ask this question by taking into account the region in which a firm is located and the spillover effects of the region on the firm. In particular, we ask if the presence of a variety of industries or of similar industries promotes the productivity of high-tech firms. In this regard, we are interested in distinguishing the high- and low-tech firms in terms of their driving factors. Next, we focus on the way the primary output of high-tech production, that is, knowledge, gets exchanged among the firms embedded in a region. Knowing that innovation and high-tech production is hardly the sole responsibility of a single firm, we know that investigating the efficiency of such exchange is critical for a high-tech region's success. We take into account the importance of social networks in promoting such exchange by preventing firms from cheating on their partners in q contract. We acknowledge that spatial proximity is a necessary but not sufficient condition for the existence of social networks and the build up of social capital in a region. Its development is partly historical and partly cultural. By investigating the effects of social capital on the Silicon Valley firms' ability to exchange knowledge with efficiency, we also understand the performance-enhancing ability of social capital. Next, we concentrate on the knowledge exchange between a consortium's partners, which compete in a downstream market. It is important to understand the effects of proximity, the complementarities of knowledge, and repeated interactions on the efficiency of such transactions. Such work provides us with the mechanics of the actual knowledge exchange.

The last two chapters investigate the engine behind U.S. agglomerations, and that is labor mobility. If one understands how and by what factors a region can retain high-tech labor, one can then design policies to achieve this. These two chapters are fundamental in particular to the U.S. economy, which is ethnically very diverse and in which a select group of foreigners receive their doctorate degrees and form an important resource of productivity and innovation. If one understands the motivating and forcing factors for first- and second-generation immigrants and work visa holders that influence a region's ability to retain such a labor pool, one can then work on promoting future high-tech clusters.

Chapter 1 is focused on understanding the separating effects of the diversity of a variety of industries versus the concentration of a particular industry on the productivity of a region. We specifically disentangle the low- from the high-tech industries in doing this. We know that high-tech firms are more likely than other firms to engage in R&D activities and to innovate, and hence, that the impact of these industries on the entire economy is very critical. This American way of advancing the economy via high-tech firms has been accepted as a model of development for other countries. Hence, the analysis in this chapter attempts to learn what drives the productivity differences across the different states for low- and high-technology industries. It appears that our analysis, which is an aggregation from county to state levels, shows that for high-tech industries it is the variety of economic activity that drives productivity. This is a significant finding; from this observation, one can deduce the requirements of an environment in which high-tech firms can flourish and innovate. Hence, this explains why a region like Silicon Valley, a hotbed of high-tech firms from a variety of industries, is more conducive to increased productivity than a place like Rochester, N.Y., which is a highly concentrated region.

Chapter 2 takes us a step further by forcing us to understand the efficient ways of knowledge exchange among the frequently contracting firms in Santa Clara County, Calif. It appears that the type of intangible knowledge exchange via contracts would involve cheating and hence be problematic; however, Valley firms appear to have cracked the code for executing these contracts with success. Therefore, in this chapter we intend to find the ways in which Valley firms choose a contract type, for example, choosing between producing in house or via a licensing agreement. We find that the location of the partner and the intangibility of the knowledge that is exchanged promote one type of contracting as being more efficient than another. In particular, if the contract involves a lot of intangible skills to be exchanged and the partnering firm is located on the East Coast, it appears to be more likely that the firms will form a joint venture rather than choosing to contract via licensing, which is a much looser type of agreement. We also discover that social networks among the contracting parties are likely to curb cheating, enabling Valley firms to engage in contracts that are cheaper to administer, that is, the loose type of contract.

Finally, in Chapter 3 we provide the ways in which knowledge is exchanged among competing firms. It turns out that although a consortium type of construct reduces the possibility of cheating, it can also promote cheating if the parties rely less on reciprocity in benefiting from such exchange because if such is the case cheating is encouraged.

Chapter 4 investigates a very critical concept, namely, a region's ability to retain the high-skilled workers with whom it has provided science and engineering doctoral degrees. Such a labor mobility approach to clusters is critical as regions' inability to understand ways in which they can "bank" their valuable brainpower would mean that we would not be able to have innovative, productive clusters. This chapter focuses on individuals' decision to stay in the region in which they received their science or engineering degree. To this end, we investigate the different patterns between native-born and naturalized Americans. These two groups appear to be influenced by different factors. For example, as compared with native-born Americans, naturalized Americans seem to have weaker social ties to the region in which they received their graduate degree, affecting their likelihood of staying there after graduation.

Chapter 5 investigates the effects of technical and ethnic clustering, along with other factors in the decision of high-tech workers to stay in the region in which their doctoral degree was granted. This chapter compares native-born Americans with work visa holders, particularly those with Indian and Chinese citizenship, as these two groups make up an important supply of high-tech labor in the United States. Work visa holders are significant for the U.S. economy, as many foreign high-skilled employees work under this visa.

The book provides a comprehensive analysis of the regional factors for high-tech development in the United States. Given the entrepreneurial and innovative potential of high-tech companies, this book provides a detailed account of their success, which is imperative to understand if we are to advance such success and/or replicate it.

Chapter 1
How High-Tech Industries Benefit from the Economies of Agglomeration

Neslihan Aydogan

1.1 Introduction

Economic agents choose to concentrate at discrete locations across space. Industry clusters and cities are good examples of such concentration. The investigation of this behavior has been an important focus of essays in some subdisciplines of economic theory such as new growth theory and economic geography. The general interest is centered around measuring concentration and diversification of economic activity and their effect on regional productivity. We argue that in doing such investigation, we must achieve two critical steps: the first is to find statistically and empirically reasonable measures of these variables, and the second is to disentangle industries in terms of measuring the effect of these variables.

There exist several attempts at developing spatial measures of the concentration and diversification of economic activity. Some of these measures, such as Herfindahl (concentration) and entropy (diversification), are intuitively straightforward and relatively simple mathematically. Some other indices, such as the Ellison-Glaeser index, are more complex and require access to more detailed information such as plant-level employment data. In this study, we adopt and expand an agglomeration measure, developed by Ciccone and Hall (1996). The main difference of the agglomeration measure as compared to others is that Ciccone and Hall employ the number of employees *per land area* rather than merely using the number of employees. We further disentangle this measure so that we account for both the concentration and diversification of economic activity.

We first need to establish the driving forces behind agglomerations before we move on to explaining agglomeration measures. We know that the equilibrium location of economic agents is determined by the opposing forces of clustering and dispersing effects (refer to Anas et al. 1998 for an excellent review of this literature). In general, clustering forces are classified by drawing from Marshall's (1920) seminal work. Briefly, the benefits of agglomeration are summarized as reduction in

N. Aydogan
Department of Economics, Çankaya University

N. Aydogan, Y.P. Chen, *Social Capital and Business Development in High-Technology Clusters*, DOI: 10.1007/978-0-387-71911-5_1,

transportation and job search costs and increased skill transfer. Dispersing effects, on the other hand, are generally listed as air pollution and congestion.

Furthering this analysis calls for an investigation that tackles the question of whether the concentration of a single activity (industry in the broad sense), as in the case of some industry clusters, or a variety of activities in a region promotes productivity. This issue is of great interest to the understanding of the formation and growth of cities, where diversification of economic activity is the key to regional productivity.

There exist a variety of studies that intend to capture the separate effects of concentration and diversification on regional productivity (e.g., Rosenthal and Strange 2003; Henderson 1986, 1998; Glaeser et al. 1992; Abdel-Rahman and Whitney 1998). Some of these studies take the productivity of economic activity as aggregate, and some conduct this analysis on individual industries. In this study, we aggregate across low- and high-tech industries, at 2-digit SIC (Standard Industrial Classification) detail, aiming to disentangle their behavioral characteristics; by employing an approach similar to that of Ciccone and Hall (1996), we develop concentration and diversification measures accounting for the land area. Hence, unlike that in Ciccone and Hall, our productivity shifter is a composite multiplicative function of concentration and diversification terms. Despite the increased interest in high-technology industries, to our knowledge, such separate industry grouping has not been employed previously. We claim that such a study is significant to promotion of regional policies directed at increasing the productivity of highly innovative high-tech firms. Obviously, accounting for both the concentration and diversification of economic activity is necessary to tailor such policies.

In the next section, we summarize some of the extant studies in the literature and lay out the construction of the productivity equation. In the third section, we explain the data and sources, and display the regression results. In the last section, we conclude the chapter with some implications for policy and future work.

1.2 Static Externalities and Industry Productivity

Extant literature on geographic externalities is centered on the question, Why are some cities large and diversified while others are small and specialized? Such a question naturally leads us to investigate which industries are likely to locate in metropolitan areas as opposed to medium- and small-sized cities. That is, what type of industries benefit more from the specialization of local economic activity than from variety and size? A common approach to obtaining stylized facts on this matter is to estimate productivity equations for each industry that is classified under the SIC for a specific region. A typical such equation involves a productivity shifter that is some function of concentration and diversification measures.

Hence, most work in this area is parametric and requires a particular functional form to be employed for estimation. The joint work by Abdel-Rahman and Whitney (1998) is an exception. Specifically, the authors employ the Tornqvist index to measure productivity, a procedure that does not entail any parametric restriction aside from the magnitude of returns to scale parameter. However, the regression results show some obvious sensitivity to the choice of this parameter.

Several authors choose to define and express agglomeration economies in several ways. As for the concentration measure, which is by definition concentration of an economic activity, some authors argue that industry scale (number of employees) should be used as a measure. Subsequently, some authors have chosen instead to employ fraction of employees in a particular industry as an appropriate measure of localization economies.

A similar pattern is observed in measuring the diversification of economic activity whose effect is often called *urbanization economies*. For example, Henderson (1988) argues that it is the size (population or number of employees) in a region, not the industry composition, that affects industry productivity. Subsequently, such authors as Glaeser et al. (1992) and Henderson (1998) himself found it appropriate to separate size effects from the variety of economic activity. The variety effects are commonly cited with the name of Jacobs (1969), who stresses diversity-related benefits such as the cross-fertilization of ideas inside a region.

For most studies in this area, Marshall's book (1920) stands as the prototype for explaining benefits to agglomeration economies in general. These benefits can be briefed enumerated as follows:

- Firms locate in close proximity to decrease transportation costs.
- Firms locate near one another to enable employees to find jobs in case they might be laid off in response to firm-specific shocks.
- Firms locate next to each other because frequency of face-to-face contacts makes it possible for ideas or skills to be exchanged easily.

None of the studies in this area explicitly focuses on distinguishing high- and low-technology industry characteristics. As in Glaeser et al. (1992), for example, the effect of agglomeration economies on innovation is explained without distinguishing among industries. Specifically, it is implicitly assumed, for instance, that the same arguments apply for Italian ceramics and electronics industries.

However, we claim that high-technology industries show different characteristics and that detecting this difference is thus significant to understanding innovation and growth. For example, Premus (1984) notes that high-technology industries employ a considerably higher percentage of scientists, engineers, and technicians than other manufacturing firms. In addition, research and development (R&D) inputs are much more important to the productivity of these firms. If one could uncover the different characteristics across low- and high-tech industries in terms of the differing effects of localization and urbanization economies, we believe, it would be more feasible to form targeted regional policies.

In this chapter, we estimate a nonlinear industry-productivity equation where the nonlinearity originates in the adoption of the Cobb-Douglas production function a la Ciccone and Hall (1996). We argue that such specification is worth accounting for as it is a very general functional form and any linear version is just an approximation. Further, authors construct a state-specific productivity equation, which they derive from a county-specific firm production function. This technique has an obvious advantage as the finer the geographic unit of measurement the more possible it becomes to avoid accounting for land area where no economic activity takes place at all. We further employ county "urban" land area in order to make this measure more vigilant in accounting for such effects.

In particular, in this work we ask the following question: To a what extent do overall density and localization of economic activity in a locale account for the productivity differences across high- and low-technology industry groups in larger regions? We discuss the construction of the productivity equation in the next section.

1.3 Industry Productivity Equations

In this chapter, we construct the productivity equation with a methodology quite different from Ciccone and Hall's (1996). The major differences are in the specification of the productivity shifter, disaggregation of the total economic activity into industry groups, and construction of the productivity equation. Ciccone and Hall specify the shifter as a general density term and hypothesize it to be Hicks neutral. We also assume Hicks neutrality and construct a multiplicative productivity shifter that utilizes the measure by Ciccone and Hall (1996).

We start our analysis with a production function that has three factors of production: land, capital, and labor. We hypothesize firm production function with constant returns to scale technology and Cobb-Douglas functional form.

Let the firm f production function for a typical industry i be represented as the following [we initially suppress the industry county subscript c]:

$$q_{i,\,f} = B_i\, a_{i,\,f}[N_{i,\,f}{}^{\phi} k_{i,\,f}{}^{1-\phi}]^{\alpha} L_{i,\,f}{}^{1-\alpha} \tag{1.1}$$

The parameter B_i is the industry-specific constant; N, k, and L, respectively, represent the number of employees, the amount of capital, and the amount of land employed at each firm f. We account for capital only parametrically, following Ciccone and Hall (1996). In other words, we hypothesize that each firm employs the optimum level of capital, and we hold the rental rate r as constant across firms in a particular industry. Some other studies in this subject area, such as Glaeser et al. (1992), ignore accounting for capital entirely. Although we do not explicitly measure capital, we find this approach more complete for mathematical exposition.

The factor a_{ic} represents the productivity shifter, which is constructed by utilizing the widely used measure of localization economies, that is, own industry employment per land area and density (number of employees per urban land area). The latter measure is a la Ciccone and Hall (1996), as described previously:

$$a_{i,\,c} = \left(\frac{N_{i,\,c}}{L_c}\right)^{\sigma} \left(\frac{N_c}{L_c}\right)^{\delta} \tag{1.2}$$

where $N_{i,c}$ represents the employment fraction of a specific industry in a county. Specifically, $N_{i,c}/L_c$ represents the number of employees per urban land area in industry i at county c, and N_c represents the number of employees across all industries. Following Ciccone and Hall (1996), we employ the density measure for urbanization in a locale; L_c represents the urban land area in a locale.

We insert the optimal value of capital k in Equation (1.1). Employing the characteristics of Cobb-Douglas specification, therefore, k can be expressed as follows:

$$k_{i,f} = \frac{(1-\phi)\alpha}{r} q_{i_f}$$
$$\Rightarrow q_{i,f} = B_i \left[a_i N_{i,f}{}^{\phi} \left(\frac{(1-\phi)\alpha}{r} q_{i,f} \right)^{1-\phi} \right]^{\alpha} L_{i,f}{}^{1-\alpha} \tag{1.3}$$

Let

$$\kappa = \left[\frac{(1-\phi)\alpha}{r} \right]^{\alpha(1-\phi)}$$

and

$$\eta = \frac{\alpha\phi}{1 - [\alpha(1-\phi)]}$$

where $0 < \eta < 1$.

Therefore, the firm f production function for industry i county c can be represented as follows:

$$q_{i,f,c} = \theta_i a_{i,c} N^{\eta}_{i,f,c} L^{1-\eta}_{i,f,c} \tag{1.4}$$

where $\theta_i = B_i \kappa_i$.

Hence, we assume that the rental rate r is identical across all firms and counties in a given industry i in state s.

Next, we aggregate (1.4) across all firms in an industry to obtain the industry production function. Our limited access to data imposes some restrictions in construction. Specifically, we hypothesize that the number of employees per urban land area is the same across all firms in an industry i and county c, that is,

$$\frac{N_{i,f,c}}{L_{i,f,c}} = \frac{N_{i,c}}{L_{i,c}} \qquad \left[\text{And let } \frac{N_{i,c}}{L_{i,c}} \equiv d_{i,c} \right] \tag{1.5}$$

This assumption follows from Equation (1.3) and is true if every firm that belongs to a particular industry i in county c faces identical wage and land rental rates. This assumption is necessary as we do not have data on land area used by each firm at each industry and county in the United States.

In addition, we assume that the number of employees per urban land area across all industries in a county is proportional to county density. And that the proportionality factor is specific to each county.

$$\frac{N_{i,\,c}}{L_{i,\,c}} = v_i \frac{N_c}{L_c} \tag{1.6}$$

where $v_i \equiv$ number of establishments in industry (i) and v_i is the industry-specific constant.

This assumption is made necessary by the lack of data on industry-county urban land area $L_{i,c}$.Therefore following (1.5) and (1.6) and aggregating Equation (1.4), we obtain the following equations leading to Equation (1.7):

$$\begin{aligned} q_{i,\,c} &= \theta_i a_{i,\,c} \sum_f N^{\eta}_{i,\,f,\,c} \left[\frac{N_{i,\,f,\,c}}{d_{i,\,c}}\right]^{1-\eta} \\ &\Rightarrow q_{i,\,c} = \theta_i a_{i,\,c} \left[\sum_f N_{i,\,f,\,c}\right] d^{\eta-1}_{i,\,c} \\ &\Rightarrow q_{i,\,c} = \theta_i a_{i,\,c} N^{\eta}_{i,\,c} L^{(1-\eta)}_{i,\,c} \end{aligned} \tag{1.7}$$

Next, we aggregate (1.7) to the state level and divide both sides of the equation by the number of employees in industry i in state s to get the productivity equation (1.8):

$$\Rightarrow \frac{\sum\limits_{c \in C_s} q_{i,\,c}}{N_{i,\,s}} = \frac{\theta_i v_i^{\eta-1} \sum\limits_{c \in C_S} \left(\frac{N_{i,\,c}}{L_c}\right)^{\sigma} N_{i,\,c} \left[\frac{N_c}{L_c}\right]^{\eta+\delta-1}}{N_{i,\,s}} \tag{1.8}$$

In this study, we use gross state product (GSP) as a total production measure, which is provided by the U.S. Bureau of Economic Analysis. The left-hand side of the aggregation term is made necessary as data on GSP is provided only at the state level. This requires the productivity term to be calculated as shown in Equation (1.8), which solves the problem on the lack of data at the county level.

The final estimation equation can be expressed as follows after taking the logarithm of both sides of Equation (1.8):

$$\Rightarrow \quad \log \frac{q_{i,\,s}}{N_{i,\,s}} = \left[\psi_i + \log \sum_{c \in C_s} \left(\frac{N_c}{L_c}\right)^{\eta+\delta-1} \left(\frac{N_{i,\,c}}{L_c}\right)^{\sigma} N_{i,\,c}\right] - \log N_{i,\,s} + u_{i,\,s} \tag{1.9}$$

where $\psi_i = \log \theta_i + \log v_i^{\eta-1}$.

In this study, we do not estimate this equation but rather display the preliminary observations. Our aim here is to construct and display an alternative and theoretically more reasonable functional expression. Preliminary facts, which follow, reveal the nonlinearity between the gross state product and productivity shifters. These results also display the distinguishing characteristics of low- versus high-tech industries.

1.4 Empirical Observations

1.4.1 Data

The data for this study was obtained mainly from the databases of the U.S. Census Bureau and the Bureau of Economic Analysis for the year 1991. The land area data is for the year 1990, based on availability.

Data on industry Gross State Product (GSP) is obtained from the Bureau of Economic Analysis Web site. The data is available for all industries at 2-digit SIC detail, and GSP is measured in million current dollars. We exclude Alaska, the District of Columbia, Hawaii, and Wyoming from analysis because either they are too small or they have large reserves of natural resources, which are likely to contaminate the results.

We obtained the industry employment data from the County Business Patterns database of the U.S. Census Bureau. Because of confidentiality concerns, some data on industry employment is flagged at the county level. Each flag represents an employment range provided by the Bureau. To fill these data gaps, we simply choose the midpoint of the flag range, following Glaeser et al. (1992).

We obtained the urban land area data from U.S. Census Bureau sources, and it is measured in square kilometers. The value zero is assigned to land area that does not fit the definition of urban place provided by the Bureau.

In this study, we mainly use the subsectors from the "Manufacturing Industry" group (U.S. Census Bureau). Unlike that of Ciccone and Hall (1996), our interest is in individual industries rather than the aggregate economic activity at a location. Hence, grouping industries across main sectors (such as manufacturing; services; transportation; communications; and electric, gas, and sanitary services) into high- and low-tech categories may not be such a good idea. Specifically separate groups face different government regulations, which might affect the behavioral response of separate industry groups.

Specifically, the high-technology industry group includes the following sectors:

- Primary Metal Industries (SIC 3300)
- Fabricated Metal Products except Machinery and Transportation Equipment (SIC 3400)
- Industrial and Commercial Machinery and Computer Equipment (SIC 3500)
- Electronic and Other Electrical Equipment and Components except Computer Equipment (SIC 3600)
- Chemicals and Allied Products (SIC 2800)

The low-technology industry group includes the following sectors:

- Food and Kindred Products (SIC 2000)
- Textile Mill Products (SIC 2200)
- Rubber and Miscellaneous Plastics Products (SIC 3000)
- Stone, Clay, Glass, and Concrete Products (SIC 3200)
- Apparel and Other Finished Products Made from Fabrics and Similar Materials (SIC 2300)

The rest of the subsectors under the Manufacturing Industry group—Tobacco Products; Lumber and Wood Products except Furniture; Furniture and Fixtures; Paper and Allied Products; Printing, Publishing, and Allied Industries; Petroleum

Table 1.1 Bottom and top 15 states, ranked by Gross State Product (GSP)

State	GSP (million $)	Urbanization no employees/ urban land area km^2
Bottom 15		
North Dakota	11,490	61.47
Vermont	11,543	94.07
South Dakota	13,947	86.18
Montana	13,994	148.91
Idaho	18,400	149.75
Rhode Island	21,577	20.56
Delaware	22,479	30.26
Maine	23,250	66.00
New Hampshire	24,704	41.56
West Virginia	29,084	252.08
New Mexico	30,202	51.01
Nevada	33,195	34.06
Utah	33,283	57.12
Nebraska	35,074	2128.49
Arkansas	40,641	165.33
Average	14,6857	251
Top 15		
	11,2937	551.84
Maryland	11,5917	133.59
Washington	12,1085	146.04
Georgia	14,7448	480.10
North Carolina	14,8713	611.21
Virginia	15,3449	557.86
Massachusetts	15,9671	91.83
Michigan	18,9876	26999
New Jersey	22,1255	151.97
Ohio	23,2337	581.15
Pennsylvania	25,5766	762.71
Florida	26,5677	183.92
Illinois	28,1930	533.68
Texas	40,5080	502.10
New York	49,9854	900.52
California	80,7789	278.07
Average	27,4586	2231

Note: Alaska, the District of Columbia, Hawaii, and Wyoming are excluded.

Refining and Related Industries; Leather and Leather Products; Measuring, Analyzing, and Controlling Instruments: Photographic, Medical, and Optical Goods; Watches and Clocks—are excluded from the analysis. This is because all these sectors except the Measuring, Analyzing, and Controlling Instruments industry depend heavily on the existence of natural resources in a region. Hence, the productivity and location decision of firms in these industries cannot be based solely on the labor mix of the area. As for the Watches and Clocks sector, we were not sure whether one might classify it as high or low technology, and hence we decided to exclude it from the analysis.

1.4.2 Preliminary Facts

In Table 1.1, we display the bottom and top 15 states, ranked by their GSP values along with their corresponding density values. (We excluded Alaska, the District of Columbia, Hawaii, and Wyoming from the entire analysis proves to be quite insignificant for the industries discussed in this chapter.) As expected, the average GSP value for the bottom 15 states is $146,857 million, with a corresponding average density value of 251. The top 15 states, on the other hand, have an average GSP value of $274,586 million, with an average density value of 2,231. Hence, urbanization or diversification appears to increase productivity. Next, we pooled industries (listed in Sect. 1.4.1) into low- and high-tech industry groups and listed the top and bottom 15 states by GSP and corresponding density (urbanization measure) for each group. The results are summarized in Tables 1.2 and 1.3.

Table 1.2 Bottom and top 15 states ranked by Gross State Product (GSP): High-Technology group

State	GSP (million $)	Urbanization no employees/ urban land area km^2
Bottom 15		
Montana	33.6	65.46
North Dakota	75.5	56.77
Nevada	76.8	31.73
South Dakota	85.4	273.96
Maine	160.4	56.92
Idaho	227.6	135.94
Rhode Island	256.2	196.49
Vermont	260.6	494.66
Nebraska	383.4	187.76
Utah	384.8	82.10
New Mexico	451.6	48.85
New Hampshire	516	38.19
Kansas	551.2	262.04
Mississippi	595.6	169.82
Delaware	604	26.33
Average	311	142
Top 15		
Missouri	1729.4	345.35
Florida	1817.2	177.35

Table 1.2 (continued)

State	GSP (million $)	Urbanization no employees/ urban land area km^2
Tennessee	1994.2	473.13
Wisconsin	2524.6	66.40
Massachusetts	2584.4	80.20
North Carolina	3142	537.70
	3420.2	494.01
Michigan	3918.6	417.08
New Jersey	3925.4	1.6278
New York	4462	731.69
Pennsylvania	4774.6	153.46
Illinois	5359.6	477.81
Texas	6209.8	53.63
Ohio	6340.2	18.31
California	9013	248.43
Average	4194	287.4

Note: Alaska, Hawaii, the District of Columbia, Hawaii, and Wyoming are excluded.

Table 1.3 Bottom and top 15 states ranked by Gross State Product: Low-Technology group

State	GSP (million $)	Urbanization no employees/ urban land area km^2
Bottom 15		
Montana	34.2	65.46
Nevada	53.6	31.73
North Dakota	53.8	56.77
Vermont	65	412.14
New Mexico	69	48.85
South Dakota	83.6	235.09
West Virginia	125.2	295.31
Maine	130.8	56.92
Rhode Island	142	196.49
Delaware	147.8	26.326
Utah	151.2	76.40
Idaho	159.6	135.94
New Hampshire	179.4	38.101
Arizona	196.4	25.47
Average	109.24	118
Top 15		
Virginia	1042.6	204.98
Missouri	1085.6	345.35
Wisconsin	1087.4	115.74
South Carolina	1213.2	102.99
Tennessee	1348.2	433.30
New Jersey	1375	1.63
Michigan	1380.2	417.08
Georgia	1974	436.13
Illinois	2037	477.81
Pennsylvania	2067.4	153.46
Texas	2123.4	137.53
Ohio	2212.2	18.31
North Carolina	2519.4	537.70
New York	2583	731.69
California	4455	248.43
Average	1900	291

Note: Alaska, the District of Columbia, Hawaii, and Wyoming are excluded.

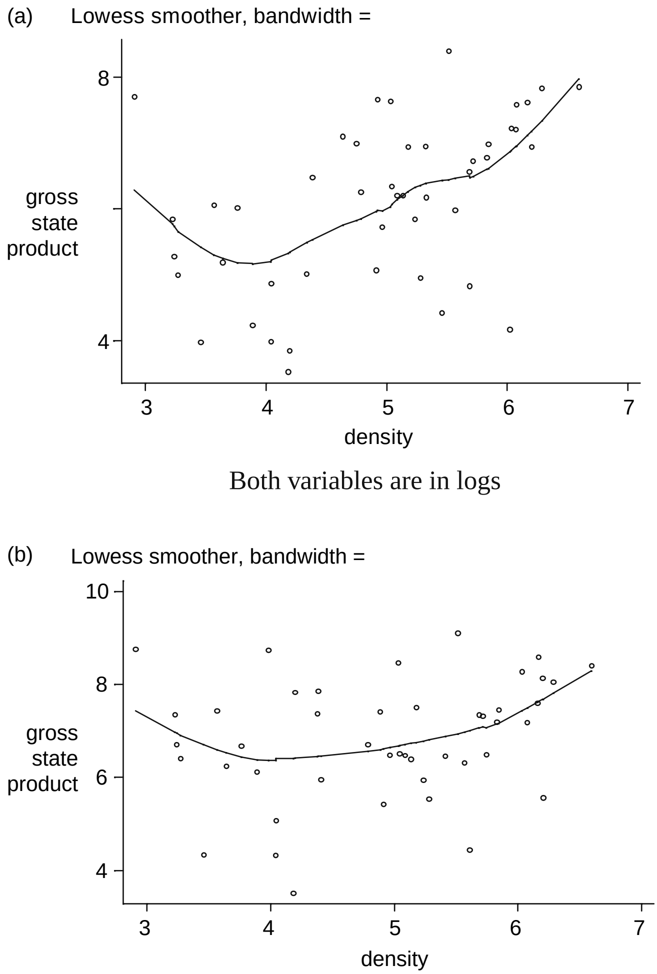

Fig. 1.1 Relation between the GSP and density (number of employees/urban land area) variables. (a) Low-technology industry group; (b) high-technology industry group.

The results in Tables 1.2 and 1.3 indicate that there is a larger difference between the GSP values of the top and bottom 15 states for the high-technology group than for the low-technology industry group.

In Figs. 1.1a and b, we fit a curve to the urbanization data by using the Lowess methodology (STATA Statistical Software Manual). Figure 1.1a indicates that at low levels of density, GSP is moderately high and that as the density values increases, GSP hits a minimum and starts increasing linearly. Figure 1.1b indicates a similar but much less distinct pattern. Specifically the relationship between GSP and density displays a flatter curve.

In Tables 1.4 and 1.5, we rank the GSP for the localization measure and list the bottom and top states for high- and low-tech industries. It appears that there is a

Table 1.4 Bottom and top states ranked by Gross State Product: High-Technology group

State	GSP (million $)	Localization (no of employees)
Bottom 15		
Nevada	76.8	2608.6
South Dakota	85.4	1532.2
Maine	160.4	2665.6
Idaho	227.6	1813.2
Rhode Island	256.2	4861.8
Vermont	260.6	2229.1
Nebraska	383.4	4890
Utah	384.8	6294.2
New Mexico	451.6	1849
New Hampshire	516	7075.2
Kansas	551.2	8498
Mississippi	595.6	6682.4
Delaware	604	1959.4
Average	350.27	3408.5
Top 15		
Florida	1817.2	27063
Tennessee	1994.2	20276.6
Wisconsin	2524.6	41272
Massachusetts	2584.4	36498.4
North Carolina	3142	28041
Indiana	3420.2	46096.8
Michigan	3918.6	58141.9
New Jersey	3925.4	39630.2
Pennsylvania	4774.6	64553.8
Illinois	5359.6	76807.5
Texas	6209.8	67360.4
Ohio	6340.2	75905.8
California	9013	135063.2
Average	3539.3	55131.58

Note: Alaska, the District of Columbia, Hawaii, and Wyoming are excluded.

Table 1.5 Bottom and top 15 states ranked by Gross State Product: Low-Technology group

State	GSP (million$)	Localization (no of employees)
Bottom 15		
Montana	34.2	500.8
Nevada	53.6	1173.4
North Dakota	53.8	671.4
Vermont	65	2661.4
New Mexico	69	1283.4
South Dakota	83.6	11015.6
West Virginia	125.2	2117.8
Maine	130.8	3021.4
Rhode Island	139.2	4512.2
Delaware	147.8	2138.4
Utah	151.2	3747.2
Idaho	159.6	3310.6
New Hampshire	179.4	12376.6
Arizona	196.4	5411
Oregon	307.6	12441.3
Average	126.42	4425.5

Table 1.5 (continued)

State	GSP (million$)	Localization (no of employees)
Top 15		
Virginia	1042.6	14447
Missouri	1085.6	32083.8
Wisconsin	1087.4	19658.75
South Carolina	1213.2	25510.6
Tennessee	1348.2	24287.4
New Jersey	1375	16078.2
Michigan	1380.2	18516.5
Georgia	1974	37041
Illinois	2037	30753.8
Pennsylvania	2067.4	24819.2
Texas	2123.4	23824.6
Ohio	2212.2	24260.4
North Carolina	2519.4	64365.2
New York	2583	43692.4
California	4455	89236.4
Average	1900.24	577811.65

Note: Alaska, the District of Columbia, Hawaii, and Wyoming are excluded.

much larger gap between the GSP values for the top and bottom states in low-tech than in high-tech industries.

1.5 Conclusions

Our quest in this chapter was to theoretically and empirically distinguish the industry concentration and diversification measures and to disentangle low- from high-tech industries in relation to these variables. The preliminary results show that one can indeed develop theoretically sound measures, despite the mathematical and conceptual challenges and that high- and low-tech industry groups display distinctive characteristics. In particular, it appears that an urbanization or diversification variable seems, relative to the localization or concentration measure, to affect high-tech industry productivity much more than low-tech industry productivity. This result supports the idea that innovative activity is fed in metropolitan areas more than in more localized regions. Silicon Valley is a great example of this observation as the Valley itself contains all kinds of supporting industries that are networked tightly. In the next chapter, we will concentrate on the matters of skill transfer among these industries.

We empirically investigate the type of contractual arrangements through which skills are transferred based on the intangible nature of the skills transferred and the location of the partner firm. Since the Silicon Valley firms empirically are found to be notoriously able to effectively exchange knowledge via inter-firm contracts across different industries, we choose to investigate the behavior of these firms in the following chapter.

References

Abdel-Rahman HM, Whitney G (1998) A nonparametric approach to measuring productivity difference between cities. Working Paper No. 15-98, University of New Orleans

Anas A, Arnott R, Small KA (1998) Urban spatial structure. J Econ Lit 36: 1426–1464

Bureau of Economic Analysis, Gross State Product Data, 1991

Ciccone A, Hall RE (1996) Productivity and the density of economic activity. Am Econ Rev 86: 54–70

Glaeser EL, Kallal HD, Scheinkman JA, Shleifer A (1992) Growth in cities. J Polit Econ 100: 1126–1152

Henderson JV (1986) Efficiency of resource usage and city size. J Urb Econ 18: 47–70

Henderson JV (1988) Urban development: Theory, fact, and illusion. Oxford University Press, New York

Henderson JV (1998) Evidence of scale economies on agglomeration. Working Paper No. 98-22, Brown University

Henderson V (1998) Urban Development: Theory, Fact, and Illusion, Oxford University Press, Paperback edition, Oxford University Press, 1991.

Jacobs J (1969) The economy of cities. Random House, New York

Marshall A (1920) Industry and trade. MacMillan, London

Premus, R. (1984). Urban growth and technological innovation. Urb Econ Devel: Urb Aff Ann Rev 27: 47–61

Rosenthal SS, Strange WC (2003) Geography, industrial organization, and agglomeration. Rev Econ Stat 85: 477–494

U.S. Census Bureau County Business Patterns Data File, 1991

U.S. Census Bureau, urban land area Data File, 1990

Chapter 2
Tacit Knowledge Transfer, Geographical Proximity, and Inter-Firm Contracts: The Silicon Valley Case

Neslihan Aydogan

2.1 Introduction

The economics of regional clusters has been a popular research area among academic and business authors as well as policymakers (e.g., see Porter1990; Porter and Stern 2001). Several studies on regional clusters are based on issues related to knowledge transfer among individuals and firms and to the mechanics of increasing returns in relation to agglomeration. As evidence of the effectiveness of knowledge transfer among firms in clusters, Jaffe et al. (1993) show that patent citations come from the same U.S. geographical area. In other words, patent applicants cite patents whose holders are located in the same city, state or statistical metropolitan area. Hence, it appears that geographical proximity facilitates effective knowledge transfer among firms and individuals. In addition, several authors after Marshall (1920) argue that agglomeration reduces transportation costs, facilitates skill transfer, and provides access to a labor pool for the area firms. One issue that is often overlooked in the literature is the exact mechanism behind the knowledge-transfer strategy of firms in agglomerations. This is the particular issue that we tackle in this chapter.

Recent research has focused on regional clusters, as they are observed to be critical for economic growth. In this chapter, we tailor a transaction costs type of framework to investigate the effects of geographical proximity, inter-firm networks, and the type of knowledge involved in a contract on the contract form (i.e., whether the contract is, e.g., unilateral as in the case of a licensing agreement or bilateral as in the case of a cross-licensing agreement or it has any other possible legal structure). The inter-firm contract is critical, as it is the primary mechanism by which firms learn and bring new products to the market or diversify into other markets. This is particularly important, as the main premise of the transaction costs theory is that firms select contracts based on cost. In other words, a firm chooses to, for example, firmsengage in a joint venture rather than a cross-licensing agreement only if the

N. Aydogan
Department of Economics, Çankaya University

N. Aydogan, Y.P. Chan, *Social Capital and Business Development in High-Technology Clusters*, DOI: 10.1007/978-0-387-71911-5_2,

former is a cheaper contract. By contract, we argue that, in the context of clusters three main factors determine the contractual costs: (1) geographical proximity, (2) inter-firm networks, and (3) the type of knowledge that is exchanged among firms. In particular, we empirically study the contract choice of Silicon Valley firms based on this premise. The motivating observation behind this work is the persistent empirical and anecdotal evidence on the systematic selection of particular contracts by the Silicon Valley firms. In this work, we explain the inter-firm contract behavior of firms located in a successful cluster (in this case, Silicon Valley) and the way they exchange knowledge successfully with their contract partners.

The significance of this work also rests on the observed frequency of the signing of inter-firm contracts after the mid 1990s in the United States and elsewhere in the world. Such observation has enriched the theoretical backbone of transaction cost economics, which initially had dismissed the different types of inter-firm contracts by lumping them under hybrid forms, between spot exchange (i.e., when exchange is done via markets) and vertical integration (i.e., when parties merge to form an independent firm; e.g., see Williamson 1991). The proliferation of such forms as licensing, cross-licensing agreements, supplier-customer networks, and others has proven that each different form has distinctive contractual implications and hence should be analyzed separately.

Inter-firm contracts are particularly important for high-tech firms that are in a hurry to bring new products into the market. In this work, we claim that the relevance and observed frequency of different types of inter-firm agreements in high-tech production calls for a more detailed study of different contract types. We particularly focus on the successful execution of the transfer of tacit knowledge among contracting parties.

Following Arora (1996), we observe that in high-tech industries even contracts that involve codified knowledge exchange, such as patent exchange, also involve tacit knowledge transfer as a critical component. Hence, the successful execution of such transfer carries a paramount significance for the success of the contractual goal.

One has to stress that exchanging tacit knowledge across firms involves some contractual difficulties. Specifically, in contracts that involve tacit knowledge exchange it may be impossible to specify in advance the nature of the tacit knowledge to be exchanged or to verify, ex-post, whether the promised knowledge has in fact been delivered. Hence, partner monitoring and moral hazard become key issues for the effective execution of knowledge transfer in inter-firm agreements for these industries. In other words, in such a context firms are likely to cheat on their contract partners, that is, engage in moral hazard (e.g., by sending their less skilled researchers to a joint research lab). This means that it is very important to monitor partner behavior closely in such cases.

In this study, following Oxley (1997), as mentioned earlier, we employ the transaction costs paradigm to analyze the inter-firm agreements in a market-hierarchy continuum by forming a ranking from markets to hierarchies. Although we rest on Williamson's (1991) framework, as does Oxley, we follow Grossman and Hart (1986) and view the firm as being composed of assets, rather than employees, and use the main thrust of the resource-based view to motivate our hypotheses.

In particular, as explained by Mowery et al. (1997), the resource-based literature involves extensive discussions of the key characteristics of the technology-related capabilities that are often based on tacit knowledge. In this study, we hypothesize that each inter-firm agreement involves activities that require the transfer of such capabilities and that the contracts closest to the hierarchy end are the least vulnerable to contractual cheating. Based on this theoretical background, in this study we group and rank agreements from markets to hierarchies. We employ several different groupings to verify the robustness of our empirical results.

In addition to applying transaction cost analysis to high-tech contracts, we incorporate geographical space into our study. We claim, in fact, that this is inevitable, given the characteristics of the tacit knowledge exchange. It is well established that such exchange demands face-to-face interaction and hence requires parties to bear travel costs. In addition, geographical proximity helps contracting parties to monitor each other's actions, which is likely to restore the correct incentives and alter the costs of different contracts (on the monitoring effects of proximity (e.g., see Lerner 1998).

Previous empirical work does discuss the significance of geographical proximity on the tacit knowledge exchange. However, there is no systematic work that integrates the two in a formal way. This is what we hope to achieve in this chapter. For example, Doz and Shuen (1995) describe spatial proximity as a significant factor that facilitates informal exchange, and they identifythe holding of personal meetings as a key factor in tacit skill transfer. In another context, Murphy (1991) describes the site-selection process of the Microelectronicsand Computer Technology Corporation (MCC) research lab as an example of the importance of proximity. Apparently, according to Murphy, the MCC had to choose a location that would not create an advantage for any member over the others in accessing skills from MCC. Hence, given these observations, we argue that any analysis that involves tacit knowledge transfer has to account for spatial proximity.

We also argue that the density of economic activity in a geographical cluster affects the way tacit skills are exchanged. This goes beyond the concept of a simple monitoring story. It is more in the spirit of regional coalitions, as described in Storper (1995), and supplier-customer networks, as formalized by Kranton and Minehart (2000). The hypothesized effect of regional coalitions and of supplier and customer networks is based on the repetitive nature of formal and informal agreements among firms in a cluster and on the influence of this activity on the efficiency of inter-firm agreements. Hence, in this study we separate the networks within clusters from those that are outside. However, our treatment of this issue is rather simplistic in that we do not attempt to specify the exact nature of the repetitive agreements other than claiming and statistically testing a likely change in behavior when all partner firms are located in Silicon Valleythat does not happen in other configurations. In other words we expect to see a change in partnership form if a firm is collaborating with another firm in the Valley versus otherwise.

Hence, in this study, we employ a ranking of different forms of inter-firm agreements, using the transaction cost theory by focusing on the transfer of technology-related capabilities, and we incorporate spatial proximity into the analysis. In addition, we employ a novel measure of "tacitness" per inter-firm agreement. As mentioned previously, Oxley employed a similar uniform ranking of hybrid forms, in which she also used transaction cost theory to conduct her analysis. However, her characterization of tacitness, as will be explained in detail here, involves a more general measure. The tacitness measure we employ does not require us to make a generalized claim, as Oxley did in hypothesizing that process design–related work is likely to involve more tacit activities than do production activities. Ours is a more direct measure that accounts for the varying degree of employee skills used for each mutual alliance activity of each agreement in the data set. This measure, we believe, equips us with much more accuracy for testing the likelihood of moral hazard and helps us avoid imposing strong assumptions.

To conduct this analysis, we employ an inter-firm alliance data set provided by Thomson's Financial (Thomson Financial is an arm of *The Thomson Corporation*, one of the world's leading information companies, focused on providing integrated information solutions to business and professional customers (Wikipedia)) to investigate the inter-firm alliance choice of a typical firm that was physically located in Silicon Valley between 1991 and 1995. Specifically, we ask how such a choice is affected by the location of the firm's partner(s) and the complexity of the skills employed to conduct the mutual alliance activities. The sample choice, as we will explain, is deliberate in that Valley firms are notorious for their ability to profit from inter-firm agreements that allow them to transfer tacit skills frequently and they are observed to frequently engage in inter-firm agreements both within the Valley cluster and outside.

In summary, the novelty of this work rests with three factors: (1) the introduction of a proxy for measuring intangible assets, (2) alternative rankings of different inter-firm agreements, and (3) accounting for the effect of geographical proximity on the partnership form.

2.2 Silicon Valley Firms and Firm Organization

Silicon Valley is a region where innovative activity is densely located and firms engage in various types of inter-firm agreements through which they actively exchange knowledge. Steve Kitrosser, vice president of operations of the disk drive producer Maxtor in Silicon Valley, for example, explains the intense relationship between suppliers and customers in the Valley: "In the really good relationships," Kitrosser emphasizes, "we are sharing process technology back and forth with our suppliers, just like we try to share information across functional groups within the company" (Saxenian 1994).

Tom Furlong, another Maxtor Corporation employee, who is former manager of DEC's workstation group in Palo Alto, describes the importance of spatial proximity: "An engineering team simply can not work with another engineering team that is three thousand miles away, unless the task is incredibly explicit and well defined –which they rarely are" (Saxenian 1994).

This anecdotal evidence, among other examples, provides support to Saxenian's careful work on the organizational structure of Silicon Valley firms. Comparing clusters in different regions, she notes that not all are as successful as those in Silicon Valley. There are, of course, many possible reasons for such discrepancy. Saxenian, however, focuses on firm organization and gives us a good insight into the possible institutional causes of the different levels of innovation across clusters. Specifically, she compares and contrasts the two almost-identical regions of Silicon Valley and Boston's Route 128, as both are occupied by firms from similar industries and first-rate universities, and they offer similar living conditions and access to venture capital. Despite such similarities, Saxenian argues that Silicon Valley's network structure, where firms frequently engage in formal and informal inter-firm agreements, is very much in contrast with Route 128's independent-establishment type. She concludes that the observed underperformance of Route 128 firms is likely to be due to this difference in the region's institutional structure.

2.3 Ranking the Inter-Firm Contracts

As described in the introduction to this chapter, in addition to incorporating the incentive-altering effects of geographical proximity we group the inter-firm agreements and rank them from markets to hierarchies. We construct our groupings based on observations from the literature (e.g., Oxley1997) and some definitions from contract law. We claim that such ranking is relevant because it has been observed that in certain industries, firms consistently choose some agreement types over others, such as cross-licensing over licensing agreements or equity over market-end agreements (Pisano 1991). This observation is also backed by theoretical arguments. For example, Morasch (1995) claims that despite the larger governance costs that joint venture agreements involve, they might be preferred over cross-licensing agreements, which are likely to be less efficient in the presence of monitoring problems, such as the case of contracts that involve tacit skill transfer, for example, teaching someone how to correct situations where laboratory equipment malfunctions.

Following Oxley (1997), we group and rank the inter-firm agreements in three alternative ways: (1) unilateral, bilateral, joint venture agreements; (2) unilateral, bilateral, equity agreements (this group includes equity, cross-equity, and joint venture agreements); and (3) joint venture and all other types of agreements. In this third case, we hypothesize that when companies get together and form a separate company, the incentive structure would be different from all the other cases, where the stakes of the parties are less dependent on each other. The overall rationale behind such dis-aggregation is the increased level of partner monitoring, administrative controls, and firm interdependency as one moves from unilateral to bilateral and joint venture or equity-type agreements, while the contract administration costs increase. In this study, we assume ex-post technology transfer in all contracts, even when the agreement is not classified as technology transfer. We approach this by carefully scrutinizing alliance activities and identifying the relevance of tacit skills that are involved in conducting them.

Contract law describes a unilateral contract as a contract where one party's obligation is due only upon the other party's performing first. Hagedoorn (2001), for example, describes licensing agreements as contracts that provide unilateral

technology access in return for a fee. An example of such a contract in the data set would be a technology transfer agreement between Apple Computer and Cirrus Logic, where the deal is described as the transfer of Apple Computer's Newton Systems
Technology to Cirrus Logic. As described in Wada (1999), when subsequent exchange of tacit knowledge is required or in case of cumulative innovation, licensing agreements fail to provide the partners with the correct incentives. Morasch (1995) describes these agreements as cheaper to administer, but claims that they involve large risks of partner cheating. Wada argues that cross-licensing agreements, based on reciprocity, establish the necessary incentives. Williamson (1984) also emphasizes the importance of bilateral dependence for avoiding the costs related to cheating.

In contract law, a bilateral contract is described as a contract in which both parties have obligations to each other simultaneously throughout the contract term. Therefore, even though firms might not be formally recontracting, in this type of contract they engage in a repeated, reciprocal relationship. For example, Hagedoorn (2001) describes the cross-licensing agreement as a bilateral form of licensing. In addition, Fehr et al. (1997) demonstrate experimental evidence of the effects of reciprocity in maintaining cooperative behavior. An example of such a contract in our data set would be a cross-technology transfer agreement between Apple and Symantec where Apple'sC-Plus Plus compiler technology is transferred in return for Symantec's PowerPC compiler technology.

In this study, we describe the equity and joint venture agreements as being the closest to integration meaning two firms getting together and forming a new firm. In addition, we treat these two types of agreements separately, where the former is an equity ownership of one firmby another. The joint venture agreement, on the other hand, involves increased integration, where two or more separate organizations create an independent business entity for strategic purposes allocating ownership, operational responsibilities, financial risks, and rewards to each member while preserving each member's identity/autonomy.

2.4 Hypotheses and Data

2.4.1 Hypotheses

In this section, we lay out the hypotheses that we test in this study:

Hypothesis 1 The higher the proportion of employees with tacit, difficult-to-verify/-transfer skills who are required to conduct mutual-alliance activities, the larger the risks of contractual hazards and the more likely that firms will engage in integration-type alliances.

In this study, we confine the term *difficult-to-transfer skills* to the skills of scientists, technicians, and engineers. Winter (1987) describes *skills* as a specified, defined set of capabilities applied to a particular task. In addition, Polanyi (1958) classifies skills as being explicit and tacit, describing *tacitness* as the situation in which the person who possesses this type of skill is not able to provide a useful explanation of the rules required to achieve a task. For example, explaining a chip design activity in words would be a rather formidable task as teaching such a skill would require on-site training. In this study, we claim that engineers, scientists, and technicians own skills that are more tacit than those required in other occupations. This is based on previous work, such as that by Almeida and Kogut (1999) and Arora (1996), and the description of skills provided by the O*NET database. O*NET is a comprehensive database, sponsored by the U.S. Department of Labor's Employment and Training Administration, of U.S. worker attributes and job characteristics. O*NET, among other things, provides a classification of occupations by required skills in each of the U.S. states, with the intention of matching the skill requirements of employers with the skills of job seekers. In addition, activities such as design, research, development, setting up, operating, and maintaining laboratory instruments and equipment have been described as tacit; for example, Almeida and Kogut contrast knowledge on electronic data forming with designing chips where the former is described as pushing the limits of physics, experience, intuition, creativity, and problem solving and that is difficult to articulate. These skills are also described as the primary skills of engineers, scientists, and technicians in the O*NET database.

To measure tacit skills we construct what we call a skill quotient variable. The construction of this variable will be described in detail. We basically claim here that the more tacit the skills that are required to be transferred between the partners, the more likely the partners are to cheat on the agreement. Hence, inter-firm agreements that are close to integration in form are likely to be more efficient than market-end agreements.

Hypothesis 2 The farther the location of its partner is from a firm the more likely it is that the firm will engage in integration-type alliances. We expect a pronounced change in the contract form choice when both/all partners are located in Santa Clara County. We also expect this effect to be accentuated as the tacitness of skills increases for achieving the mutual-alliance tasks.

This hypothesis is based on the anecdotal evidence and theoretical claims that suggest that reduction in monitoring, transportation, and information transfer costs is due to geographical proximity. Hence, if partners in an alliance are farther apart, the more likely it is that they will engage in agreements that are close to integration in form, in an attempt to reduce those costs. If both/all partners are located inside the Valley, we further hypothesize, such effect is likely to be magnified. In addition, we claim that increased tacitness of the skills required to achieve the mutual-alliance activities is likelyto accentuate the monitoring costs, increasing the likelihood of an integration-type agreement.

Hypothesis 3 Alliances that include research activities are more prone to partnership hazards than are those with manufacturing and/or marketing activities. We, therefore, expect such activities to be governed by integration-type alliances. This hypothesis is in the spirit of Oxley (1997) and based on observations on the distinguishing characteristics of research-related activities in relation to the increased involvement of tacit skills.

2.4.2 Data and Methodology

In this study, we use the only available commercial database, SDC-Platinum, which is provided by Thomson's Financial. This database contains substantive information on inter-firm agreements. In this data set, each data point is an inter-firm agreement. The data set contains information on the partners of an inter-firm agreement and activities that are involved. Details on the data can be found in the datatables at the end of the chapter. We supplement this data with the publicly available LexisNexis database (®LexisNexis® **ZipFind 2.0** from **Bridger Systems**, Inc.,) to find the subsidiary firm locations, ZIP Code Finder software to calculate the physical distance between the partnering firms, and the industry-occupation matrix for Santa Clara County, which is provided by the State of California, Employment Development Division (1995). This last data set is used to construct the skill quotient variable, which is our measure of tacitness in this study.

SDC-Platinum obtains alliance information from publicly available sources such as SEC filings, trade publications, news, and wire sources. As reported in Anand and Khanna (2000), the data goes as far back as 1986; however, the data prior to 1990 is not equally comprehensive. Hence, our sample runs from 1991 to 1995. Also, as Anand and Khanna point out, the data cannot be expected to include all the inter-firm agreements in which the Silicon Valley firms engage. This is because there exists no standard practice for firms to report every agreement that they sign.

A careful review of the sample reveals that the data is widely diversified with different firm sizes and types (public, private, and subsidiary) and that it includes both two- and multimember alliances. This is important, as a data set with a small number of Silicon Valley firms and a large number of partnerships would result in an analysis where the behavior of only a handful of Silicon Valley firms would be picked up, resulting in a sample-selection bias.

In terms of industry detail, a large portion of the sample includes five high-tech industries: (1) Industrial Machinery and Equipment, (2) Electrical and Electronic Equipment, (3) Communications, (4) Business Services (Computer Programming and Data Processing, Computer Programming Services, Prepackaged Software, Computer Integrated Systems Design, Information Retrieval Services, Computer-Related Services, Not Elsewhere Classified), and (5) Engineering and Management Services (Commercial Physical and Biological Research).

The SDC-Platinum database, as explained in detail by Anand and Khanna (2000), presents some problems for quality assurance. Specifically, we were obliged to verify whether each claimed agreement in the data set did actually take place. To this end, we compared each alliance in the database with the information in SEC filings in situations where at least one firm was publicly held. Fortunately, only 2.29% of the partnerships involved firms that were all private. Hence, we are confident that the data is factual.

In addition to assuring quality in the empirical work, we employ different groupings of alliance types. In addition to some theoretical reasons, we do this to check robustness and avoid coding errors; such errors may occur as a function of multiple classification of an alliance in the data set, for example, when an agreement is described as both a technology transfer and a joint venture. A scrutiny of such possible double counting across different results indicates no obvious bias.

The data set includes information about alliances between firms located in Silicon Valley and the rest of the United States. The information on partner location is provided by the ZIP code and corresponding city detail.

In Table 2.1, we display the names of the 131Silicon Valley companies used in the regressio analysis. In this study, we exclude alliances with firms in foreign countries. Alliance decisions with firms located in foreign countries are likely to include aspects unrelated to the scope of this work. For example, political and economic stability of the country in which the partner is located often plays a significant role in identifying the partner firm. In addition, for this particular sample, the general characteristics of alliances where all partners are located in the United States and alliances where at least one partner is located in a foreign country are quite similar. Hence, excluding the data on alliances where at least one partner is located in a foreign country is unlikely to bias the results (data and descriptive statistics on these alliances are available upon request from the author).

The total number of data points is 1,074 in the period between 1991 and 1995. The variable names and definitions are displayed in Table 2.2. For 489 of these alliances, at least one firm was located outside the United States. The data set lacks information on contract forms for 41 observations. Another 50 observations lack information on the location of the contracting firms, and 14 observations lack information on the skill quotient variable. Hence, the number of usable observations is 480.

The skill quotient variable is constructed by employing the Santa Clara County industry occupation matrix, which is put together, upon request, by the Employment Development Department (EDD) of the State of California for the year 1995. The industry occupation matrix provides information on the fraction of employees for each occupation under each 4-digit Standard Industrial Classification code (SIC). Occupation classifications are provided by the U.S. Bureau of Labor Statistics. Some information is suppressed due to confidentiality (2.4% of the 585 observations). The EDD obtains this information from survey data on the 113,000 California employers by region. The survey contains 830 occupations, and employers report the number of individuals they employ in each occupation. We calculate the skill quotient values by summing the fraction of scientists, engineers, and technicians for each alliance activity that is classified under the 4-digit SIC in the data set. For example when calculating the skill quotient value for the Computer Storage Devices Industry, the following proportions of employees from the relevant occupations are added:

Chemical Engineers = 0.015+
Electric and Electronic Engineers = 0.034+
Computer Engineers = 0.046+
Industrial Engineers, except Safety = 0.031+

Table 2.1 Company names in the data set

3Com Corp	Bus Logic, Inc.	Hewlett-Packard Corp.	MTI Technology Corp.
ADAC Laboratories	Cadence Design Systems, Inc.	Hybrid Networks, Inc.	NEC Electronics, Inc.
	Caere Corp.	IBM Corp.	Net Frame Systems, Inc.
ASK Group, Inc.	Calpine Corp.	ICTV Cox Communications Corp.	Net Manage, Inc.
AT&T Network Systems	Catalyst Semiconductor, Inc.	Indigo Medical, Inc.	Netscape Communications Corp.
Adaptec, Inc	C-Cube Micro-Systems	Insite Peripherals, Inc.	NETSYS Communications Corp.
Adobe Systems, Inc.	CEH, Inc.	Intel Corp.	Neuron Data, Inc.
Advanced Micro Devices. Inc.	Chips and Technologies, Inc.	International Imaging. Inc.	Novell, Inc.
Amdahl Corp.	Chronological Simulation Corp.	Internet Media Services	Oak Technology, Inc.
America Online, Inc.	Cisco Systems. Inc.	Intuit, Inc.	Octel Communications Group
Appian Technology, Inc.	Co-circuit Zilog, Inc.	Kalok Corp.	OIS Optical Imaging Systems Operations Group. Inc., Pacific Western Branch
Ariel PSS Corp	Commerce Net	Kubata Pacific Computer, Inc.	Opti, Inc.
Atari Corp	Compaq Devices, Inc.	Lightpost Publishing	Parallan Computer, Inc.
Atari Games Corp.	Compass Design Automation, Inc.	Logic Modelling Systems, Inc.	Pionex, Inc.
Atmel Corp.	Cypress Semiconductor Corp.	LSI Logic Corp.	Pen Magic Software
Apple Computer, Inc.	Dazix Corp.	Measurex Corp.	Philips Electronics NV
Ariel PSS Corp	DSP Group Inc.	Mediametics, Inc.	Photonics Corp.
Baxter International, Inc.	DSP Semiconductors U.S.A.	Memorex Computer Supplies	QD Technology
BCT, Telus Communications	Electronic Data Systems Corp.	Meta Software, Inc.	Quantum Corp.

Mechanical Engineers = 0.005+
Engineers, NEC = 0.189+
Electric and Electronic Engineering Technicians = 0.021+
Engineering-Related Technicians, NEC = 0.046
Engineering-Related Technicians, NEC = 0.046

Table 2.1 (continued)

Bell Microproducts, Inc.	EnaTech Software Systems, Inc.	Metricom, Inc.	Rambus, Inc.
Blockbuster Entertainment Corp	FDX Corp.	MicroModule Systems, Inc.	Rae Technology, Inc.
Boole and Babbage, Inc.	HAL Computer Systems, Inc.	MIPS Technologies, Inc.	Rational Software Corp.
Boorland International	Headway Technologies	Motorola Computer Group	Recognition Systems, Inc.
Reply Corp.	Stanford Telecom	Tandem Computers, Inc.	Verity, Inc.
Rocket Science Games, Inc.	Storage Dimensions, Inc.	Trident Microsystems	Xicor, Inc.
Rolm Corp.	Storm Technology, Inc.	Trilogy, Inc.	Zilog, Inc.
Seagate Software, Inc.	Stor Media, Inc	Trimble Navigation	
San Jose Aream San Jose Sharks	Symantec Corp.	Trinzic Corp.	
Secure Computing Corp	Synopsis, Inc.	Tyecin Systems, Inc.	
Semaphore Communications	Sysgen, Inc.	UB Networks	
Sharebase Corp.	Sun Microsystems	VLSI Technology, Inc.	
Sigma Designs, Inc.	Sunsoft, Inc.	Web TV Netwoirks, Inc.	
Silicon Storage Technology, Inc.	SuperMac Technology, Inc	Western Digitial Corp.	
Silicon Graphics, Inc.	Taligent, Inc.	Widham Hills Products	

Chemical Technicians = 0.001
Systems Analysts = 0.002
Support Specialist = 0.002

This gives us a sum of 0.392.

Table 2.3 displays the distributionof skill quotient values across different industries classified by the 4-digit SIC. Some alliance activities are provided at 3-digit SIC detail. In such cases, we took the average of the skill quotient values over their 4-digit SIC values. Among all industries, Semiconductors and Related Devices (19%) and Prepackaged Software (35%) are disproportionately represented. The large proportion of these two industry groups might prevent variation in the skill quotient variable, and hence the results might not be reliable. However, the results prove robust as we drop these industries from the regressions.

The top five industries that employ the largest proportion of scientists, engineers, and technicians are (1) Commercial Physical and Biological Research,

Table 2.2 Variable names and definitions

Variable Name	Variable Definition
Date of announcement	Date when the alliance is announced in the press
Participant name	Name of the participants involved in the alliance
Participant ultimate parent name	Name of the participants' parent company
Participant city	City of participants' headquarters
Participant ZIP code	ZIP code of the participants
Participant public status	Public status of participants (i.e., public, private, or a subsidiary)
Alliance number of participants	Total number of participants involved in the alliance
Alliance application and technology	Textual summary of the activity/activities in which the alliance is engaged. If one or more participants in an alliance transfer technology, the text also includes which technology was transferred and who transferred the technology*
Alliance activity description	Activities in which the alliance is engaged
Alliance industry and participant ***i***ndustry	4-Digit SIC code of the alliance and participants' primary business activity

**For example*: Develop Fiber Optic Network, Integrated Circuits; Dolby's Proprietary AC-3 audio algorithm, Speed up a time-consuming step in developing computer programs; Echo Logic's Flashport Technology.

(2) Computer-Related Services, Not Elsewhere Classified, (3) Electronic Components, (4) Prepackaged Software, and (5) Computer Storage Devices.

2.4.3 Variable Definitions

In this section, we describe the variables in the study to test the described hypotheses.

2.4.3.1 Dependent Variable

Partnership Type The dependent variable is formed by ranking and grouping the inter-firm agreements in three different ways, as described in Section 2.2.

Alliance Grouping Type (1) Unilateral alliances take the value (1), bilateral alliances take the value (2), and joint venture agreements take the value (3).

Alliance Grouping Type (2) Unilateral alliances take the value (1), bilateral alliances take the value (2), and equity alliances take the value (3). Although equity alliances can be described as unilateral, they are included under this last category of alliances. The reason for such grouping is to determine the separating incentive structure that equity-type alliances might provide.

Alliance Grouping Type (3) The joint venture agreements take the value (1) and all other agreements take the value (0).

Alliances grouped as unilateral are as follows: technology transfer agreements, licensing with royalty, equity, and supply agreements. Therefore any agreement that does not include any type of reciprocity among the partners is considered unilateral. In this data set, the agreements coded under licensing are technology transfer agreements. Taking this into account, we do not account for licensing agreements separately.

The alliances grouped as bilateral are as follows: cross-technology transfer and mixed agreements that contain more than one form. This particular sample does not include information on cross-licensing agreements.

2.4.3.2 Independent Variables

Skill Quotient Variable This variable is formed to test the first hypothesis. We expect larger values of this variable to increase the probability of integration-type alliances.

Table 2.3 Skill quotient distribution across industries

Industry	Pecentage of alliances (%)	Skill quotient
SIC2759 Commercial Printing, not elsewhere classified (NEC)	0.002	0.001
SIC3500 Industrial and Commercial Machinery and Computer Equipment	0.002	0.97
SIC3571 Electronic Components	0.075	0.441
SIC3572 Computer Storage Devices	0.031	0.392
SIC3575 Computer Terminals	0.002	0.095
SIC3577 Computer Peripheral Equipment, NEC	0.029	0.174
SIC3600 Electronic and Other Electrical Equipment and Components	0.002	0.18
SIC 3661 Telephone and Telegraph Apparatus	0.045	0.20
SIC3663 Radio and Television Broadcasting and Communications Equipment	0.016	0.294
SIC3669 Communications Equipment, NEC	0.002	0.34
SIC 3674 Semiconductors and Related Devices	0.19	0.337
SIC 3679 Electronic Components, NEC	0.0125	0.131
SIC 3695 Magnetic and Optical Recording Media	0.0125	0.088
SIC 4812 Radiotelephone Communications	0.0042	0.011
SIC 4813 Telephone Communications, Except Radio-telephone	0.0042	0.034
SIC4841 Cable and Other Pay Television Services	0.0083	0.02
SIC4899 Communications Services, NEC	0.002	0.017
SIC5045 Computers and Computer Peripheral Equipment and Software	0.00104	0.165
SIC5084 Industrial Machinery and Equipment	0.042	0.034
SIC5230 Paint, Glass, and Wallpaper Stores	0.002	0
SIC6531 Real Estate Agents and Managers	0.002	0
SIC6794 Patent Owners and Lessors	0.062	0.095
SIC7370 Computer Programming, Data Processing and Consulting Services	0.033	0.31
SIC7371 Computer Programming Services	0.042	0.41
SIC7372 Prepackaged Software	0.35	0.395
SIC7373 Computer Integrated Systems Design	0.025	0.388
SIC7375 Information Retrieval Services	0.0020	0.133
SIC7379 Computer Related Services, NEC	0.035	0.464
SIC8731 Commercial Physical and Biological Research	0.002	0.548

Note: The proportions are obtained from the State of California, Employment Development Department (1995).

Location Variables We construct three separate measures to determine the impact of proximity between firms. The data set has information on the location of firm headquarters. It also includes the participant firm name along with the parent company name and specifies whether a partnering firm is public, private, or subsidiary. Information on subsidiary firm locations is extracted from the online LexisNexis academic universe database. The location variables are formed to test the second hypothesis.

1. ***External location dummy***: This variable takes the value (0) if all partners of a Silicon Valley firm are located in Silicon Valley and (1) otherwise. We expect larger values of this variable to increase the likelihood of alliances that are close to the integration end. We also employ a more flexible version of this variable where it takes the value (0) if at least one member is located in Silicon Valley and (1) otherwise.
2. ***Regional location***: This variable takes the value (3) if the partner is located outside the Western United States; (2) if in a Western state; (1) if in California outside Silicon Valley; and (0) if in Silicon Valley. We expect larger values of this variable to increase the likelihood of alliances that are close to the integration end.
3. ***Distance***: This variable is formed by calculating the mileage between the partners by using the partners' ZIP codes. For multimember alliances, we calculate the distance between the Silicon Valley firm and all others and employ the shortest distance. We expect larger values of this variable to increase the probability of integration-type inter-firm agreements. We display the estimated distance coefficients in units of thousands. This is done for exposition.

Alliance Research Activity Variable This variable is formed to test the third hypothesis. It takes the value (1) if the partnership is a research alliance and (0) if it is manufacturing or a marketing alliance. We expect larger values of this variable to induce integration-type agreements.

Alliance Activity Variable This variable takes the value (2) if the partnership is a research alliance, (1) if it is manufacturing, and (0) if it is marketing. We form this variable as an alternative to the "alliance research activity variable." This version presumes that manufacturing involves more research than marketing activities.

2.4.3.3 Control Variables

Trend This variable is defined as (year– 1990) where "year" is the date when the alliance was established.

Same SIC Dummy This variable takes the value (0) if both firms belong to the same industry and (1) otherwise. This is a measure of technological distance in the spirit of Jaffe et al. (1993). We expect that firms belonging to the same industry might run a larger risk of partner cheating. Therefore, we expect larger values of this variable to encourage agreements that are close to the market end.

Alliance Size Dummy This variable takes the value (1) if the alliance has two members and (0) otherwise. We hypothesize that it might get costlier for firms to monitor their partners when the alliance is large. Therefore, we expect larger values of this variable to induce agreements that are close to the market end.

2.5 Results

2.5.1 The Ordered Probit Specification

The dependent variable is formed by applying an ordinalranking from markets to integration. Hence, an ordered probit model is used for the econometric analysis. We test the fitness of the ordered model in comparison to the unordered model. The results favor the ordered model in comparison to the unorderedspecification.[1]

The model is specified as follows:

$$Y_i^* = bX_i + u_i \tag{2.1}$$

where X is a set of explanatory variables and u is the random error term with normalized mean 0 and variance1. We hypothesize that each Silicon Valley firm chooses among three alternative forms of inter-firm agreement, where the observed choices can be stated as follows:

$$\text{If } Y_i < \mu_1 \qquad \text{then Partnership type} = 1 \tag{2.2}$$

$$\text{If } \mu_1 < Y_i <_1 \mu_2 \qquad \text{then Partnership type} = 2 \tag{2.3}$$

$$\text{If } \mu_2 < Y_i \qquad \text{then Partnership type} = 3 \tag{2.4}$$

In addition, (μ) is the threshold parameter that is estimated along with the other coefficients in the model. The probabilities of the dependent variable (Y) can be expressed as follows:

[1] Alternatively, we also estimate a multinomial logit equation and construct a test to compare the ordered versus unordered model. Since, these two alternatives are nontested, the appropriate test for such comparison in this case is not the conventional likelihood-ratio test. Instead, we employ an asymptotic likelihood-ratio test, developed by Vuong (1989), which enables us to compare these two models. As reported in Small and Song (1992), this test computes the difference in fitted log-likelihood values between the two models and compares it to a theoretical distribution that Vuong derives. In this case, the value of the test statistic, under the null hypothesis this becomes, is 0.083, and hence the test is inconclusive. There is an obvious weakness to Vuong's test, as the test statistic does not take the number of parameters into account. A rough comparison of the results shows that the coefficients from the unordered model have invariably larger standard errors. Further, we observe that the signs of the coefficient estimates for the critical variables are the same from these two regressions and the magnitudes are quite similar.

$$\Pr ob(Y_i = 1) = prob[bX_i + u < \mu_1] = F(\mu_1 - bX_i) \tag{2.5}$$

$$\begin{aligned}\Pr ob(Y_i = 2) &= \Pr ob[\mu_1 < bX_i + u < \mu_2] \\ &= F(\mu_1 - bX_i) - F(\mu_2 - bX_i)\end{aligned} \tag{2.6}$$

$$\Pr ob(Y_i = 3) = \Pr ob[\mu_2 < bX_i + u] = 1 - F(\mu_2 - bX_i) \tag{2.7}$$

The estimation procedure of these probabilities involves the estimation of the parameters, that is, $(b)s$ and the $(\mu)s$, which are obtained by maximizing the following function:

$$L = \prod_i \prod_j \Pr ob(Y_i = j|X_i)^{d_{ij}} \tag{2.8}$$

where $d_{ij} = 1$ if $(Y_i = j)$ and (0) otherwise.

In studies of this sort, authors avoid reporting the marginal changes in probabilities. There might be some disadvantages to this procedure as ordered probit coefficients might be ambiguous to interpret in the absence of such calculation. In this study, we report the marginal change results for one of the composite variables (External Location × Skill quotient) in Table 2.4a, as it is likely to get particularly ambiguous to sign the coefficient estimates in this case.

2.5.2 Regression Results

Tables 2.4b,c, and d display the regression results with all three location measures (i.e., external location variable, distance, and regional location, respectively) for the three types of groupings described previously. External location, distance, regional location, skill quotient, alliance research activity, and alliance size variables are all statistically significant at the 5% level. In addition, the composite variable obtained by multiplying each location variable with the skill quotient (e.g., external location × skill quotient) is also significant for all location measures at the 5% level. These results are consistent across all alliance groupings. Further, we construct a simple diagnostic test (Table 2.5), which is a linear regression of the skill quotient variable on the external location variable. The result strengthens the main empirical results and the theoretical arguments in this essay. Specifically, the regression results indicate a strong correlation between the increased skills to conduct mutual alliance activities and the location of the partner.

All variables except the composite variable have the expected signs across all location measures and alliance groupings. The composite variable takes a negative sign, which is the opposite of what we expect. To check the accuracy of this result, we calculate the marginal changes in estimated probabilities of each alliance category for Alliance Grouping Type (1). In ordered models, it is expected that we will have inaccuracy in the signs of different probability measures. Marginal change results help disentangle the direction of change, particularly for composite variables.

Table 2.4a Results from ordered probit regressions, alliance grouping type (1)

Dependent variable: Partnership type	Model 1: With external location dummy	Model 2: distance	Model 3: With regional location dummy
External location dummy	0.88* (0.34)		
Regional location			0.34* (0.13)
Distance		0.62* (0.16)	
Skill quotient	2.92* (0.84)	2.78* (0.71)	2.84* (0.82)
External location × skill quotient	–4.20* (0.94)		
Distance × skill quotient		–2.5* (0.47)	
Regional location × skill quotient			–1.59* (0.35)
External location × alliance research activity	0.13 (0.26)		
Distance × alliance research activity		0.10 (0.13)	
Regional location × alliance research activity			0.057 (0.098)
Alliance research activity variable	0.59* (0.21)	0.67* (0.18)	0.62* (0.20)
Trend	–0.03 (0.048)	–0.044 (0.049)	–0.032 (0.049)
Same SIC variable	–0.38* (0.19)	0.35* (0.19)	–0.41* (0.19)
Alliance Size variable	–1.06* (0.13)	–1.07* (0.13)	–1.03* (0.13)
Log-likelihood value	–327	–391.9	–324.9

Notes: 1. Alliance Grouping Type (1) indicates grouping inter-firm agreements as unilateral, bilateral, and joint venture agreements, where each agreement type is as described in the text.
2. External location dummy takes the value (0) if all partners of a Silicon Valley firm are located in Silicon Valley and (1) otherwise. The regional location dummy takes the value (3) if the partner is located outside the Western United States, (2) if it is located in a Western state, (1) if it is located in California outside Silicon Valley, and (0) if it is located in Silicon Valley. The distance variable measures the mileage between the partners. See Section 2.4 for more details on the variables.
3. Numbers in parentheses are standard errors. An asterisk indicates the significance of a variable at the 5% level.
4. The time period is 1991–1998, and the number of variables is 480.

As summarized in Table2.4a, it appears that firms choose agreements that are close to market end when all partners are located in the Valley and as the proportion of skills increase. This result is as previously hypothesized.

All results in Tables 2.4b,c, and d are robust at least for Model 1. Specifically, the results remain unchanged when some variables, for example, external location × alliance research activity, are excluded from the model. Alternatively, the robustness

Table 2.4b Marginal Change Results for Model 1, Alliance Grouping Type (1)

	Probability ($Y = 1$)	Marginal change	Probability ($Y = 2$)	Marginal change	Probability ($Y = 3$)	Marginal change
Skill quotient external location = 0 after 3% increase	0.203 0.211	0.008	0.784 0.773	–0.011	0.0163 0.018	0.002
Skill quotient external location = 1 after 3% increase	0.43 0.211	–0.01	0.44 0.45	0.01	0.13 0.134	0.004

Notes:1. Marginal change results are obtained for Model 1 with external location dummy and for Alliance Grouping (1). The inter-firm agreements are grouped as unilateral, bilateral, and joint venture.
2. The external location dummy takes the value (0) if all partners of a Silicon Valley firm are located in Silicon Valley and (1) otherwise
3. The marginal change calculations are obtained by using the standard normal cumulative distribution function where the probabilities are evaluated at mean values of the variables.
4. Row 1 represents the case where the external location variable is fixed at zero (all partners are located in Silicon Valley) and the skill quotient variable is increased by 3%. Row 2 represents the situation where the external location variable is fixed at one (at least one partner is located outside Silicon Valley) and the skill quotient variable is increased by 3%.
5. The time period is 1991–1998.

of the regression results is also verified by, for example, replacing the research activity variable with the alliance activity variable and comparing the coefficient estimates.

Once all three models with separate location measures and different alliance groupings are compared, the one with the distance measure seems to outperform the other two. This is because the*t* ratio values for each coefficient in this model are larger than the ones in the other two models. The log-likelihood values for Alliance Grouping Type (1) with the distance measure seem to outperform other models and alliance-type classifications. This fails to provide strong support for the separating effect of a cluster as a location from mere proximity.

In addition, we also report ordered probit results where joint venture, cross-technology transfer, and licensing agreements are the dependent variables. Results in Table 2.6 provide support for the incentive-building effects of joint ventures and cross-technology transfer agreements in comparison to licensing agreements. Although this result appears a bit short of being significant at the 5% level(*t* ratio = 0.93), the coefficient of the distance variable for licensing agreements is negative in comparison to joint venture and cross-technology transfer agreements. This result supports the monitoring story in this chapter.

2.6 Discussion and Concluding Remarks

Inter-firm alliances are widely observed among high-tech firms. These contract forms help high-tech firms access complementary assets, decrease uncertainty, and enter and exit different industries. Despite the wide acknowledgment of their promi-

Table 2.4c Results from ordered probit regressions, alliance grouping type (2), Model 1

Dependent variable: Partnership type	Model 1: With external location dummy
External location dummy	0.68*
	(0.33)
Skill quotient	2.97*
	(0.82)
External location × skill quotient	–3.77*
	(0.94)
Alliance research activity variable	0.34
	(0.20)
Trend	–0.03(0.047)
Same SIC variable	–0.38*(0.18)
Alliance size variable	–0.95*(0.13)
Log-likelihood value	–355.56
Number of observations	480

Notes:1. The results are obtained for Model 1 with external location variable and Alliance Grouping (2) where inter-firm agreements are grouped as unilateral, bilateral, and equity alliances.
2. External location dummy takes the value of (0) if all partners of a Silicon Valley firm are located in Silicon Valley and (1) otherwise
3. Numbers in parentheses are standard errors. Asterisks indicate the significance of a variable at the 5% level.
4. The time period is 1991–1998.
5. The equation includes the composite variable external location alliance research activity, which is not reported in this table. It is insignificant at the 5 and 10% levels. The cutoff parameter estimates are also not reported in this table.

nence in high-tech industries a critical issue related to inter-firm agreements (i.e., the efficient transfer of tacit skills among the partnering firms) is not studied in a geographical context.

Transferring tacit skills among firms has been highlighted separately in different disciplines of research such as research on multinationals and on organization theory. However, why and how geographical proximity might help such transfer is not cultivated in a structured way in the firm strategy and industrial organization literatures. In this study, our objective has been to describe this issue in a structured manner by explaining the underlying problems in the transfer of tacit skills and colocation might help ease these problems and thereby alter the efficiency of different contract forms. This is particularly important in advising high-technology firms by showing them the efficiency benefits of geographical proximity in forming alliances. The study also separately observes the efficiency effects of colocation within a networked cluster in comparison to being located in close proximity.

The results from this study show that firms located in Silicon Valley follow a systematic strategy in organizing the exchange of individual skills based on the location of their partners. Specifically, the results support the hypotheses that relate the intensity of tacit skills to different contract forms. It appears that the more tacit are the skills for conducting the mutual alliance activities, the more likely it is that a typical Valley firm would engage in integration-type agreements. The incentive-altering effects of geographical proximity among the partnering firms are also confirmed in this study. The results show that the farther away a firm's partner

Table 2.4d Results from ordered probit regressions, alliance grouping type (3), Model (1)

Dependent variable:Partnership type	Model 1:With external location dummy
External location dummy	0.52(0.45)
Skill quotient	2.93*(1.05)
External location × skill quotient	–2.95*(1.19)
Alliance research activity variable	–0.48(0.37)
Trend	–0.012(0.062)
Same SIC variable	0.61*(0.22)
Alliance size variable	–1.3*(0.19)
Log-likelihood value	–134.99
Number of observations	480

Notes:1. The results are obtained for Model 1 with the external location variable and for Alliance Grouping (3) where inter-firm agreements are grouped as joint venture agreements and the rest of the alliances as separate.
2. The external location dummy takes the value (0) if all partners of a Silicon Valley firm are located in Silicon Valley and (1) otherwise
3. Numbers in parentheses are standard errors. Asterisks indicate the significance of a variable at the 5% level. The external location variable falls a bit short of being significant *t* ratio = 1.15. This variable is significant when, instead, distance is used as an independent variable with *t* ratio = 2.12
4. The time period is 1991–1998.
5. The equation includes the composite variable external locationxalliance research activity, which is not reported in this table. It is insignificant at the 5 and 10 % levels. The cutoff parameter estimates are not reported in this table.

Table 2.5 Ordinary least squares: A simple diagnostic test

	Coefficient	Standard error	*t* Ratio
Dependent variable:			
Skill quotient			
External location	0.31**	0.03	22.8

Notes: 1. The regression is estimated to diagnose the correlation between the two key variables in the main regression (Table 2.3a): skill quotient and external location.
2. Double asterisks indicate significance at the 5% level.

is located, the more likely it will be to choose an integration-type alliance. In addition, the results show that when all partners are located in Silicon Valley, the increased level of tacit skills in a partnership increases the likelihood of market-end agreements vis-à-vis the integration-type agreements. These results are robust to the alternative specifications and groupings of the dependent variable.

The results have both practical and theoretical implications. On the practical side, they lay out the importance of partner location for the successful execution of inter-firm alliances. This is especially true when x-post skill transfer is a prominent factor in these agreements. The results also call for policy implications that are relevant to particular local policymakers. Local governments and associations can encourage the inception and growth of geographical clusters by promoting the network relations among the local firms. For example, Lechner and Dowling (1999) mention that the Munich/Martinsried Biotechnology Region suffers from a nonintegrative strategy. Apparently, firms in the Martinsried area, which is a suburban town in the Munich area, constitute the networked part of the biotech cluster. Firms in

Table 2.6 Ordered probit regressions for the joint venture, licensing and cross-licensing agreements with distance

Dependent variable: partnership type	Model 1: Joint venture agreements	Model 2: Licensing agreements	Model 3: Cross-licensings agreements
Distance	0.49* (0.24)	–0.14 (0.15)	0.39* (0.17)
Skill quotient	2.95* (1.003)	1.47* (0.69)	1.18 (0.77)
Distance × skill quotient	–1.62*(0.67)	0.14(0.45)	–1.58*(0.51)
Distance × alliance research activity	–0.57*(0.21)	0.17(0.13)	0.20(0.15)
Alliance research activity	0.37(0.25)	–1.16*(0.19)	0.98*(0.20)
Trend	–0.055(0.064)	0.16*(0.051)	1.07(0.052)
Same SIC variable	0.73(0.22)	0.36*(0.17)	–0.58*(0.20)
Alliance size variable	–1.39*(0.25)	1.211*(0.18)	–0.32*(0.15)
Log-likelihood value	–131	–253	–237

Notes:1. Numbers in parentheses are standard errors. Asterisks indicate the significance of a variable at the 5% level. The distance variable is a bit short of being significant at the 5% level in Model 2 (t ratio = 0.93). The skill quotient variable is a bit short of being significant at the 5% level in Model 3 (t ratio = 1.53).
2. The distance variable measures the mileage between the partners.
3. The time period is 1991–1998. The cutoff parameter estimates are not reported.

the Munich region, on the other hand, do not find themselves part of this network, despite their 10- to 15-km distance to Martinsried, which the authors argue is detrimental for the growth of the region.

On the theoretical side, our results show that geography and tacit skills are significant for having a more complete analysis of inter-firm contracts and that our introduction of a novel measure of tacit skills is a contribution to pushing the envelope in this field. Furthering this agenda within a game theoretical framework, for example, is likely to shed more light on the interaction of these variables with inter-firm contracting (e.g., see Aydogan and Lyon 2004).

Some improvements can be made on the study. For example, the tacitness measure that we employ in this study can be improved by conducting a survey to find the actual number of scientists, engineers, and technicians employed in each agreement. Another improvement would require one to detail the network relations among the Valley firms by finding out the frequency, density, and form (formal vs. informal) of these relations.

The results, obviously, call for future work to test the theoretical arguments stated in this chapter. This research area is fruitful and deserves the attention of both theoretical[2] and empirical economists. In Chapter 3, we construct a theoretical framework for technology trading coalitions (in this case, research consortia),

[2] For example, in subsequent work by Aydogan and Lyon (2004) we construct a theoretical work.

where the equilibrium results from a repeated game are determined by the distance between trading partners are and the degree of complementarity between the technologies.

References

Adams JD, Jaffe AB (1996) Bounding the effects of R&D: An investigation using matched establishment-firm data. Rand J Econ 27(4):700–721

Almeida P, Kogut B (1999) Localization of knowledge and the mobility of engineers in regional networks. Manag Sci 45(7):905–917

Anand BN, Khanna T (2000) The structure of licensing contracts J Ind Econ 48:103–135

Arora A (1996) Contracting for tacit knowledge: The provision of technical services in technology licensing contracts. J Dev Econ 50(2):233–256

Aydogan N, Lyon TP (2004) Spatial proximity and complementarities in the trading of tacit knowledge. Int J Ind Org 22(8):1115–1135

Di Nardo J, Johnston J (1997) Econometric methods, 4th edn. McGraw-Hill, New York

Doz YL, Shuen A (1995) From intent to outcome: The evolution and governance of inter-firm partnerships. Working Paper 95/19/SM, INSEAD, Fontainebleau, France

Fehr E, Simon G, Kirchsteiger G (1997) Reciprocity as a contract enforcement device: Experimental evidence. Econometrica 65(4):833–860

Greene WH (1993) Econometric methods, 2nd edn. Prentice-Hall, Englewood Cliffs, N.J.

Grossman S J, Hart OD (1986) The costs and benefits of ownership: A theory of vertical and lateral integration. J Polit Econ 94(4):691–719

Hagedoorn J (2001) Inter-firm R&D partnership: An overview of major trends and patterns since 1960. National Science Foundation, Division of Science Resources Studies, Strategic Research Partnerships: Proceedings from an NSF Workshop, Arlington, VA

Hart O, John J (1990) Property rights and the nature of the firm. J Polit Econ 98:1119–1158

Intercooled STATA 5.0, Econometrics software. Stata Corp. College Station, Tex

Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. Q J Econ 108:577–598

Kranton RE, Minehart DF (2000) Networks versus vertical integration. Rand J Econ 31:570–601

Lechner C, Dowling M (1999) The evolution of industrial districts and regional networks: The case of the biotechnology region Munich/Martinsried. J Manag Gov 3:309–338

Lerner. J, (1998) Venture capitalists and the decision to go public. J Financ Econ 35:293–316

LexisNexis Academic universe database

Marshall A (1920) Industry and trade. MacMillan, London

Morasch K (1995) Moral hazard and optimal contract form for R&D cooperation. J Econ Behav Org 28:63–78

Mowery DC, Oxley JE, Silverman BS (1997) Strategic alliances and interfirm knowledge transfer. Strateg Manag J 17:77–91

Murphy WJ (1991) R&D Cooperation among marketplace competitors. Library of Congress, Quorum Books, New York

Oxley JE (1997) Appropriability hazards and governance in strategic alliances: A transaction costs approach. J Law Econ Org 13:3877–3409

O*NET Database. http://www.onetcenter.org/whatsnew.html

Pisano GP (1991) The governance of innovation: Vertical integration and collaborative arrangements in the biotechnology industry. Res Policy 20:237–249

Polanyi M (1958) Personal knowledge: Towards a post-critical philosophy (Gifford Lectures, University of Aberdeen (1951–1952). Routledge & Kegan Paul, London and University of Chicago Press, Chicago

Porter M. (1990) The Competitive Advantage of Nations. Basic Books, New York

Porter M. Scott Stern (2001) MIT Sloan Management Review (Summer 2001) 42(4):28–36
Saxenian AL (1994) Regional advantage: Culture and competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, Mass.
SDC-Platinum Joint ventures and Strategic alliances data base. Thompson’s Financial Co., 1991–1998
SEC 10-K Filings and Forms, Edgar database. http://www.sec.gov/edgar.shtml
Small KA, Song S (1992) “Wasteful” commuting: A resolution. J Polit Econ 100(4): 888–898
State of California, Employment Development Division (1995) Santa Clara County industry occupation matrix
Storper M (1995) Regional technology coalitions: An essential dimension of national technology policy. Res Policy 24(6):895–911
Vuong HQ (1989) Likelihood ratio tests for mode selection and non-nested hypothesis. Econometrica 57:307–333
Wada T (1999) Cumulative innovation, appropriability and cross-licensing: An empirical study of patent citations and US-Japan licenses in electronics. Unpublished manuscript
Winter (1987) In: Teece D (ed) The competitive challenge. Ballinger, Cambridge, Mass
Williamson OE (1984) Perspectives on the modern corporation Q Rev Econ Bus 24(4):64–71
Williamson OE (1991) Comparative economic organization: The analysis of discrete structural alternatives. Admin Sci Q 36(2):269–296
Zip Find 2.0 software, Bridger Systems, Bozeman, Mont.

Chapter 3
Reciprocity, Proximity and Performance of Research Consortia

Neslihan Aydogan and Thomas P Lyon

3.1 Introduction

The Japanese model of publicly funded consortia was adopted inthe United States beginning in the 1980s following the relaxation of antitrust restrictions on joint research and development (R&D). A great example of such an effort is Sematech (Semiconductor Manufacturing Technology), a publicly subsidized consortium of 14 semiconductor firms. As we will explain, the economics of consortia has been tackled from several angles including knowledge transfer between the consortium and its members (e.g., see Sakakibara 1997; Link et al. 1996; Irwin and Klenow 1996). However, there is no theoretical work that models the transfer of tacit, that is, difficult to codify, knowledge between the member firms and the consortium within a geographical context. This is what we add to the consortium literature in this study. The relatively simple mathematicalsetup provides us with intuition to understand the mechanics behind knowledge transfer among the competing parties, which is the case of the firms located in Silicon Valley.

In the existing literature on consortia, several authors point out the importance of knowledge spillovers between the participant firms. For example, Branstetter and Sakakibara (2002) examine the impact of a large numbers of research consortia, which are sponsored by the American and Japanese governments, on the productivity of the participating firms. The authors find that such productivity is positively correlated with the level of potential spillovers among the consortium participants. In addition, Irwin and Klenow (1996) report that several executives of the participating firms of the Sematech Consortium describe such spillovers as occurring via people-to-people interaction and sending personnelto Austin, where the consortium

---

A version of this paper has been published with the following citation: Aydogan N, Lyon TP (2004) Spatial proximity and complementarities in the trading of tacit knowledge. Int J Ind Org. 22:1115–1135.

N. Aydogan
Department of Economics, Çankaya University

N. Aydogan, Y.P. Chen, *Social Capital and Business Development in High-Technology Clusters*, DOI: 10.1007/978-0-387-71911-5_3,

is located. Hence, in both cases the authors focus on the transfer of tacit knowledge, for which face-to-face meetings are required. In this study, we argue that geographical proximity plays a critical role due to two issues. First, exchange of tacit knowledge requires the transferring firm to train the transferee, for example, by demonstrating the knowledge. Hence, these two parties need to colocate for the transfer to take place. In addition, tacit knowledge is difficult to specify *ex ante*, and it is difficult to verify ex-post whether the promised knowledge is delivered or not. Hence, incentive alignment between the parties becomes an issue. We argue that proximity facilitates the alignment of these incentives as it enables parties to monitor each other's actions. Further, we argue that reciprocal learning among the consortium firms also improves the incentive alignment between the exchanging parties as it restores incentives to cooperate. If learning is one sided, we argue, the incentive to cooperate would be much weaker. This is a simple application of the reciprocity phenomena in the trust literature, where parties are observed to cheat less in agreements when they receive reciprocity (e.g., see Martin et al. 2004).

In particular, in this chapter following the Sematech example, we model a multi-member consortium where member firm employees travel to the consortium and engage in meetings to exchange knowledge. As we just explained, we aim to find out how the geographical proximity between the firms and the consortium as well as reciprocal learning among member firms affect the efficiency of knowledge exchange. In doing this, we do not model the research process, but rather concentrate solely on the transfer of knowledge. To the best of our knowledge although knowledge transmission is studied by some others (Cooper 2001; Berliant et al. 2000; Goyal and Moraga-Gonzalez 2001), the two most relevant features to such exchange —geographical proximity and reciprocal learning—are not modeled within a consortium framework.

Our findings show that when reciprocal learning is significant, individual firms are more likely to contribute to the consortium over time. In addition, knowledge exchange is observed to be more sustainable over larger distances between the firms and the consortium if it is more effective in reducing firm costs.

In the next section, we introduce an infinitely repeated technology trading game, where member firms exchange knowledge by traveling to the consortium. Firms have a choice to passively receive knowledge in the meetings or reciprocate; both of which decrease the costs of production. Subsequently, we model a symmetric Cournot quantity competition in the product market. In the third section we conclude.

3.2 Model

We envisage an N firm symmetrical consortium where firms are located around a circle with diameter d. Each firm is represented by a single employee who receives knowledge every period, and each period he or she has to decide whether to travel to the consortium and if done, whether to truthfully disclose knowledge in bilateral meetings with the other consortium firms. We assume that if, at the end of each period, any firm cheats, the firm, represented by the employee, is expelled from the consortium, and after that, the consortium stays intact. This equilibrium concept is described as *stacked reversion* by Curtis and Eswaran (1997), who claim that it is a better representative of cooperation in coalitions in comparison to Nash reversion.

We assume that traveling to a consortium is costly but that once the employee is at the consortium, bilateral meetings are costless. Firms are assumed to be located evenly around a circle, and they are assumed to be connected to the consortium via a spoke with length $d/2$. In order for each firm to be at another firms'facilities, each has to travel along the spoke to the consortium, incurring a travel cost in the amount of $d/2$, where the unit costs are normalized to a dollar. This is structured to assume away any asymmetries among the member firms, as such is not the focus of this chapter.

Let q_i be the output chosen by firm i and the total industry output be $Q = \sum_{i=1}^{N} q_i$. The industry demand is given by $P(Q) = a - bQ$. The cost function of each firm i is specified as the following:

$$c_i(x) = \alpha - \beta \sum_{i \neq j} x_{ji} - \gamma \sum_{i \neq j} x_{ij} x_{ji} \tag{3.1}$$

where x_{ji} is firm j's disclosure of its knowledge, which takes the value of (3.1) if firm j is truthful; β is the cost-reducing parameter for the unilateral knowledge transfer, and γ is the cost-reducing parameter for reciprocal learning.

If all firms truthfully transfer knowledge, given (3.1), they each have the following cost function:

$$c_i^{\text{cooperate}}(x) = \alpha - (\beta + \gamma)(N - 1) \tag{3.2}$$

And given (3.2) in the repeated game, the firm can either cooperate forever or cheat on the sharing agreement. If the firm chooses to cooperate, forever, it gets the following payoff at the equilibrium where δ is the discount factor:

$$\pi_i^{\text{cooperate}}(x) = \frac{\dfrac{[a - \alpha + (\beta + \gamma)(N - 1)]^2}{b(N + 1)^2} - \dfrac{d}{2}}{(1 - \delta)} \tag{3.3}$$

Alternatively, the firm can choose to cheat on the sharing agreement. This can occur in two alternative ways. If the reciprocal learning is sufficiently large, that is, $\gamma > \beta/(N - 1)$,[1] then the firm would cheat by not traveling to the consortium as, once traveled, it would choose to share its knowledge. If it chooses not to travel, its

[1] This is if profits from cooperation are larger than the profits from cheating in the stage game, that is, $\pi^{\text{cooperate}} > \pi^{\text{cheat}}$ implies $\gamma > \beta/(N - 1)$.

membership in the consortium would be ended or it would be ostracized from then on, making it no longer accepted as part of the consortium and knowledge exchange. In this case, the firm's payoff will be as follows:

$$\pi_i^{\text{nottravel}-\text{cheat}}(x) = \frac{\dfrac{[a-\alpha+(\beta+\gamma)(N-1)(N-2)]^2}{b(N+1)^2}}{(1-\delta)} \tag{3.4}$$

If, on the other, hand the reciprocal learning is sufficiently small, that is, if $\gamma \langle \beta/(N-1)$, then the firm can travel to the consortium and withhold its knowledge while receiving everybody else's knowledge for a period. After this first period, the firm would lose its membership in the consortium. Its profits in the repeated game can be represented as follows:

$$\pi_i^{\text{travel}-\text{cheat}}(x) = \frac{a-\alpha-(\beta+\gamma)(N-2)+\beta N(N-1)}{b(N+1)^2} - \frac{d}{2} + \delta\frac{\pi^{\text{nottravel}-\text{cheat}}}{1-\delta} \tag{3.5}$$

Our focus in this study is to figure out how distance plays out as a decisive equilibrium factor, that is, we would like to find the threshold distance below which cooperation is an equilibrium. The larger this distance, we would conclude, the more the firms would be able to cooperate over larger distances from the consortium. Hence, following (3.3) and (3.5), one could find the maximum distance, the distance threshold, at the equilibrium between the member firms and the consortium that would support knowledge exchange, given sufficiently small and large reciprocity.

Proposition[2] For traveling to the consortium, *knowledge sharing is an equilibrium in the repeated game if*

$$d < D_C = \begin{cases} \left[\pi_i^{\text{cooperate}}\frac{d-2}{(1-\delta)} - \pi_i^{\text{notravek}-\text{cheat}}(1-\delta)\right] if\ \gamma \geq \frac{\beta}{(N-1)} \\ \frac{2}{\delta}\left[\pi_i^{\text{cooperate}} - \delta(1-\delta)\pi_i^{\text{notravel}-\text{cheat}} - (1-\delta)\right. \\ \qquad \left.\left(\pi_i^{\text{travel}-\text{cheat}} + \frac{d}{2} - \delta\frac{\pi_i^{\text{notravel}-\text{cheat}}}{1-\delta}\right)\right] if\ \gamma < \frac{\beta}{(N-1)} \end{cases} \tag{3.6}$$

This proposition shows that as reciprocity gets larger, the distance between the member firms will grow and the consortium gets efficient for supporting cooperation. We also find that the distance below which cooperation is equilibrium is more sensitive to the reciprocal learning parameter in comparison to the unilateral learning parameter. We also find that knowledge exchange is sustainable across

[2] The proposition is obtained by subtracting (3.4) and (3.5) from (3.3) and pulling out the distance parameter d for each.

larger distances when it is more effective in reducing costs regardless of the exchange form and that the distance below which cooperation is sustainable is more sensitive to reciprocal learning.

3.3 Conclusion

Our main finding in this study supports the claim that increased reciprocity in learning supports knowledge exchange within a consortium over greater distances, giventhat travel costs matter in such exchange and cheating is possible. The findings are useful in motivating sharing plans that would involve and monitor increased reciprocity of learning within the consortium. Such a result certainly sheds light on the sustainability of research alliances, given the significance of knowledge transfer among member firm employees.

Chapters 4 and 5 delve into the black box of skilled labor agglomeration in terms of the issues that affect the retention of the highly skilled laborin regions focusing particularly on native-born versus naturalized American citizens. Further, in Chapter 5 we try to explain the spatial clustering of certain ethnicities in U.S regions focusing on native-born citizens versus H1-B visa holders. These two chapters nicely complement the first three chapters by focusing on the mechanics of agglomerations, detailing the factors that explain labor mobility, and taking into account the significance of highly skilled foreign labor in the United States.

References

Berliant M, Reed, R III, Wang P (2000) Knowledge exchange, matching and agglomeration. Working Paper, Washington University, St. Louis

Branstetter L, Sakakibara M (2002) Measuring the impact of ATP-funded research consortia on research productivity of participating firms. Working Paper NIST GCR 02-830, National Institute of Standards and Technology

Cooper D (2001) Innovation and reciprocal externalities: Information transmission via job mobility. J Econ Behav Org 45:403–425

Curtis B, Eswaran M (1997) Technology trading coalitions in super games. Rand J Econ 28(1): 135–149

Goyal S, Moraga-Gonzalez JL (2001) R&D networks. Rand J Econ 32:686–707

Irwin D, Klenow P (1996) Sematech: Purpose and performance. Proc Natl Acad Sci 93: 12739–12742

Link AN, Teece D, Finan WF (1996) Estimating the benefits from collaboration: The case of Sematech. Rev Ind Org 11(5):737–751

Martin B, Armin F, Ernst F (2004) Relational contracts and the nature of market interactions. Econometrica, Econometric Society 72(3):747–780, 05

Sakakibara M (1997) Heterogenity of firm capabilities and cooperative research and development: An empirical examination of motives. Strateg Manag J 18:143–164

Chapter 4
Citizenship, Social Capital, and Spatial Assimilation of Highly Skilled Labor and Location Choice

Yiu Por Chen

4.1 Introduction

In this chapter, we explain the differences in the location choices of science and engineering (S&E) PhDs who are native born versus those who are naturalized American citizens. Specifically, we compare these two groups as to their likelihood of staying in a specific state. A comparison of these two groups is interesting in that they might be very different in their respective cultures and hence in their behavior in selecting locations in which to work and live. Further, the mobility pattern of these two groups is likely to affect the supply of human capital in certain U.S. locations and hence have a strong impact on the country's educational and regional policies.

Naturalized American citizens can be described as first-generation U.S. immigrants, who are thereby not restricted by the conditions imposed by work visas such as H1-B. Hence, their patterns of assimilation in the United States have important economic implications. In contrast, H1-B visa holders are subject to important restrictions; for example, they are required to remain employed by their sponsor company for a certain period before they are allowed to apply for permanent residency. We focus on this latter group in the next chapter. Overall, in this chapter we connect immigrant assimilation to employee ethnicity and high-skilled labor mobility, a task that has not been previously tackled.

We argue that location choice and spatial assimilation are intertwined in ways that determine the ethnic groupings of employees across space. In this study, we aim to learn the factors that determine such groupings. Specifically, we aim to find out how, for example, one's social capital (here, we define *social capital* as one's social networks, which include, for example, family and friends along with colleagues) as wellas ethnicity, skill level, and family background affect location choice and a particular geographical grouping.

To achieve this task, we employ a data set of scientists and engineers with doctorate degrees, which we obtained from the 2001 Survey of Doctorate Recipients in the United States (SDR), a longitudinal panel survey, by the National Science

Y.P. Chen
School of Public Service, DePaul University, 25 E. Jackson Blvd, Suite 1250, Chicago, IL 60604, USA

N. Aydogan, Y.P. Chen, *Social Capital and Business Development in High-Technology Clusters*, DOI: 10.1007/978-0-387-71911-5_4,

Foundation (2001), of individuals who have received their doctorate degrees in science and engineering. In this study, given all the above-mentioned factors, we aim to find the difference in behavior between native-born and naturalized American citizens in their decision to remain in the same state in which they received their doctorate or to leave after graduation.

Our findings show that there are substantial differences in the spatial assimilation of these two groups. We show that spatial assimilation might change according to a variety of factors such as the individual's citizenship, family structure, ethnicity, social capital, field of study, type of industry, research and development (R&D) funding of his or her field of study, and whether he or she has a post-doc degree.

In Sect. 4.2, we review some of the recent work on spatial assimilation. In Sect. 4.3, we compare native-born and naturalized Americans by using the SDR database (NSF 2001). In Sect. 4.4, we display two sets of logistic regressions, where we compare the differences between the native-born and naturalized Americans by using a migration-decision model. In Sect. 4.6, we conclude and discuss the policy implications of brain retention in a region.

4.2 Literature Review: Highly Skilled Labor, Immigration, Assimilation, and Social Capital

The mobility of highly skilled labor has been a particularly important public policy issue in the United States. In general, this issue has a deeper focus in the United States than elsewhere because skilled immigrant labor is significant for the country's development. On a smaller scale, each state competes to attract highly skilled workers and retain them after their graduation. It appears that not much work has been done to explain the spatial distribution of highly skilled labor in the United States (except from Audretsch and Stephan 1996 and from Stephan et al. 2004). We now analyze some of the general theories on migrant assimilation.

Chiswick (1999) studies the human capital theory of human migration. The author views migration as an investment in human capital for local economies if the migrant selection is positive. He illustrates the way a place can "bank" the invaluable "brain" for local economic development. Specifically, he shows that there is a possibility of reducing and even stopping the so-called brain-drain problem.

The brain-drain problem occurs when individuals receive their education in one state and migrate to other states to work after graduation. Such behavior benefits other states at the expense of the state that invested in educating the individual. Thus, it is important, particularly to the state that provided the education, to understand the factors that affect the location choice of doctorate program graduates. For example, industrial policies such as R&D investment may help the states in which the highly skilled workers were educated retain their local human capital by discouraging their migration to other states.

Research and development facilities may also be conducive to the creativity of highly skilled workers and affect their location choice. Stephan et al. (2004) showed that doctoral education in science and engineering is critical to a university's role in fostering economic development. However, using data from the 1997 to 1999 Surveyof Earned Doctorates (SED), administered by the National Science Foundation to all doctoral recipients in the United States, these authors also show that the training locations of new PhDs who subsequently work in the industry appear to be different from what the university and R&D expenditure would suggest (Stephan et al. 2004). The authors found significant outflows of new PhDs from the Midwest and significant inflows of new PhDs to the Pacific and Northeast regions of the country. On the other hand, another study (Koo 2005), using the U.S. Patent and Trademark Office's 1995–1999 patent data, shows a high correlation between patents and industrial clustering at the county level. Hence, it appears that there is a positive relation between innovative activity and employee clusters. Therefore, the brain-drain problem may be mitigated by R&D and industrial polices, and it may actually be decreased by economic assimilation.

In terms of economic assimilation, Chiswick (1978), who used the 1970 Census to focus on immigrants in the United States, showed that foreign-born white male immigrants may take 10–15 years to "catch up" with their native-born counterparts. Chiswick (1977) also showed that if the first-generation immigrant does relatively well, this positive-assimilation effect may be "transferred" to the second generation and enable that second generation to earn more than their counterparts whose parents are native born. When comparing the income between the first- and second-generation immigrants, Borjas (2006), using the 1940 and 1970 Censuses and a 1995–2003 survey of population in the United States, demonstrates that second-generation immigrants earn 5–10% more than the first generation.

There might be many reasons why some immigrants need more time to assimilate into U.S. society. One barrier to assimilation for naturalized Americans is their family background. For instance, family background may affect one's ability to communicate effectively, one's ability to assimilate, and one's income. Focusing on the 15- to 44-year-old immigrant cohort with different arrival times and races, Alba and Nee (2003, p. 223) compared the percentage of first-generation household members with second-generation Americans. The authors found that the second generation (born in the United States of immigrant parents) speaks English much more often than the first cohort. When Alba and Nee compared the results across different ethnicities, they found that the second generation varies widely in terms of habitually speaking only English at home. It appears, for example, that this number is almost 90% for Filipinos and Koreans, but only 30% for Dominicans. This observation is important as Chiswick and Miller show in separate papers that language skills may affect earnings (Chiswick 1978; Chiswick and Miller 2002; Miller and Chiswick 1996).

In addition to language skills, a graduate's first job and his or her location choice might be highly correlated to informal and formal networks such as the effort of his or her advisors, formal school facilities for job hunting, and other remote employment sources such as Internet job sites and the school's networks. For example, Smeets et al. (2006, p. 169) show that the first job and its location are highly related to the informal network of the doctorate student's advisor, especially in the American economics profession. Also, the school's informal and formal networks via its advisors may translate into research activities in the industry. Audretsch and Stephan's (1996) work shows strong connections between university-based scientists and companies in biotechnology. This research shows that a school's social capital (formal and informal networks) has an effect on the location choice of highly skilled human capital. However, there are no studies that compare the ability of the graduate to benefit from the school's social capital and location choice across different citizenships.

Additional factors such as ethnic networks and language are usually considered important to immigrants' location choice. Recent work, such as that of Chiswick et al. (2002), develops a theoretical framework for the study of the tendency of immigrant groups to be geographically concentrated. In testing the model for Australia, the authors show that the extent of geographic concentration of immigrant groups is negatively related to age at migration, duration of residence in Australia, and the proportion of the countrymen who are fluent in English. The extent of geographic concentration appears to be affected by the availability of ethnic media and by the distance between the country of origin and the place of residence in Australia. The drawback of these kinds of models is that they do not explicitly take individual preferences into account, but rather describe the spatial displacement.

It comes into view that until now, little work has been done on spatial assimilation and the location choice of highly skilled labor. Later in this chapter, we compare the differences between native-born and naturalized Americans with science and engineering degrees, and we also discuss the factors affecting the different spatial distributions among these groups. We claim that it is important to study the connection between assimilation and socialization of different groups and the impact of these on the spatial distribution of science and engineering labor. Once we understand this, we can also understand the effects of these factors on the migration decision of doctorate graduates from a given locale, which can then help local governments tailor policies to retain locally trained human capital and hence achieve economies of agglomeration in certain clusters (Helsley and Strange 1990).

Hence, this chapter fills a gap in the high-technology labor-mobility literature by introducing a linkage between citizenship, a school's social capital, and spatial assimilation in the graduate's location choice. This work will also take into account the variables that embody cultural elements such as the aspiration to find job satisfaction, the individual's job preference, and ethnicity variables. In the next section we will identify the location pattern of the two groups of PhDs.

4.3 The Spatial Distribution of Science and Engineering PhD's

There are usually three major legal ways for a person to become a naturalized American citizen. First, some people immigrate to the United States as minors with their parents. Second, some apply to become naturalized American citizens after they arrive in the United States as adults. (For example, many graduate students apply for a green card once they find jobs in the United States, and then some later apply for citizenship.) Third, some have refugee status and are allowed to stay in the United

Table 4.1 Citizenship of American S&E PhDs, by year awarded (%)

Year S&E PhD awarded	Native-born	Naturalized	Permanent resident	Temporary resident	Row total
1990	66.67	18.75	14.58	0.00	240
1991	68.12	16.56	14.70	0.62	483
1992	65.00	15.60	18.60	0.80	500
1993	64.90	12.94	20.20	1.96	510
1994	61.85	12.48	23.89	1.78	561
1995	62.91	10.02	23.32	3.76	639
1996	59.30	9.38	23.85	7.47	629
1997	61.48	7.24	18.85	12.43	732
1998	59.04	8.01	15.76	17.18	774
1999	58.74	5.34	13.11	22.82	824
2000	61.50	5.43	11.37	21.71	387
Total	62.05	10.22	18.27	9.46	6,279

Source: National Science Foundation, Division of Science Resources Statistics (2001) Survey of doctorate recipients in the United States.

States legally. Our research focus is then to explain the factors relating to spatial distribution by comparing the naturalized Americans who hold PhDs in science and engineering with their native-born American counterparts.

Table 4.1 shows the distribution of recent graduates (from 1990 to 2000) in terms of residence status and whether they were born in that location or were born outside the United States and naturalized. We observe a decreasing proportion of naturalized Americans in the distribution. This pattern supports the argument that the naturalization process following the immigration of foreign students has taken a longer time in recent years than it did earlier.

Naturalized Americans may behave very differently from their native-born counterparts simply because of their different social and cultural backgrounds as well as the degree of their assimilation. In particular, Borjas (1984) shows that different cohorts of immigrants may earn differently in salary level. Speed of assimilation may affect immigrants' behaviors and thus their income growth. Chiswick et al. (2006) demonstrate that immigrants with higher skill levels and those with higher skills in economics are more likely to be associated with higher earning levels than other immigrants. In addition, their earnings appear to grow faster than do those of family-based and refugee immigrants. The analysis indicates that immigrant economic assimilation does occur, but that the speed of such assimilation depends on the background of the immigrant.

Table 4.2 shows the distribution of employees in major businesses by the ethnicity of those who are native-born and those who are naturalized American citizens. Each cell in the table represents the percentage of the ethnicity in each area of business relative to the total number of that ethnicity in the PhD graduate population of the 2001 SDR data set. We first compare cells within each group, and we then compare different groups with the same ethnicity.

When comparing native-born and naturalized Americans across different professional categories, we observe that the native-born Americans show similar characteristics and are distributed in similar patterns. When we compare the U.S.-born

Table 4.2 Distribution of major business employees and the ethnicity among Americans (%)

	Native born				
	White	Hispanic	Black	Asian	Native American
01 Agriculture, forestry, or fishing	1.4	1.0	1.4	0.4	1.9
02 Biotechnology	2.8	2.7	1.4	2.4	1.9
03 Construction or mining	0.3	0.0	0.3	0.4	1.9
04 Education	42.2	39.3	48.5	39.9	39.4
05 Finance, insurance, or real estate services	0.9	2.0	0.7	1.2	1.0
06 Health services	10.4	8.1	6.1	9.9	11.5
07 Information technology	4.5	3.0	4.1	7.1	1.0
08 All other services (e.g., social, legal, business)	2.7	3.4	3.8	2.8	1.0
09 Manufacturing	5.7	4.0	4.1	4.7	3.8
10 Public administration/ government	4.2	4.4	4.1	2.8	5.8
11 Research	19.3	26.5	21.5	24.9	23.1
12 Transportation services, utilities, or communications	0.9	1.0	0.3	0.4	1.0
13 Wholesale or retail trade	0.6	1.3	0.3	0.0	0.0
14 Other	4	3.4	3.4	3.2	6.7
Total percentage	100	100	100	100	100
Total number	2,949	298	293	253	104

	Naturalized			
	White	Hispanic	Black	Asian
01 Agriculture, forestry, or fishing	0	3.2	0	1
02 Biotechnology	0.6	0	0	6.3
03 Construction or mining	1.8	0	0	0
04 Education	35.9	38.7	42.9	21.9
05 Finance, insurance, or real estate services	4.8	3.2	3.6	3.1
06 Health services	6.6	3.2	10.7	8.4
07 Information technology	6.6	3.2	7.1	16.9
08 All other services (e.g., social, legal, business)	2.4	3.2	7.1	1.4
09 Manufacturing	6	12.9		11.1
10 Public administration/ government	4.8	0	7.1	4.3
11 Research	21.6	25.8	14.3	18.1
12 Transportation services, utilities, or communications	4.2	0	0	2.2
13 Wholesale or retail trade	0.6	0	3.6	0.5
14 Other	4.2	6.5	3.6	4.8
Total percentage	100	100	100	100
Total number	167	31	28	415

Source: National Science Foundation, Division of Science Resources Statistics (2001) Survey of doctorate recipients in the United States.

Asian Americans to naturalized Asian Americans, we find an interesting pattern. Specifically, we find a sharp contrast between the two groups; for example, while the percentage of native-born Asian Americans in the education field is around 40%, the percentage of naturalized Asian Americans in that field is only 21.9%. However, the pattern does not show to be so different between the comparisons of other ethnicities groups; the distribution of native-born African Americans in the education field is 48.5%, while the proportion of naturalized African Americans is 42.9%. The variations of occupation distribution among Americans, however, may also be due to differences in supply of and demand for graduates and the retention rate in each state.

Table 4.3 shows comparisons of retention rates of science and engineering PhDs of native-born and naturalized Americans, respectively . The diagonal cell of each table shows the retention rate, and these are quite similar in pattern between the two groups. Specifically, the average retention rates of the two groups are 26.3 and 27.4% for native-born and naturalized Americans, respectively.[1]

It appears that although the average retention rates of the two groups are similar, their distribution is not. When comparing the retention rate of the two groups, we find that the region with the highest retention rate is East South Central (where the retention rate of native-born Americans is 31.4% while that for naturalized Americans is 34.2%). However, the regions with the lowest retention rate are different between the two groups, with the lowest retention rate for native-born Americans being 21.3% in the East North Central region and for naturalized Americans being 19.1% in the West North Central region.

We also find that the highest-supply regions for new PhDs are East North Central (15.8%) and Middle Atlantic (16.7%) for native-born and naturalized Americans, respectively. In addition, it appears that the lowest-supply region is East South Central with 5.8 and 5.9% for native-born and naturalized Americans, respectively. We also observe that the largest graduate absorption comes from the Pacific region, with 14.2 and 19.3% for native-born and naturalized Americans, respectively, and the lowest graduate absorption is the East South Central region, with 5.8 and 4% for native-born and naturalized Americans, respectively.

One interesting observation concerns the Pacific region, where it appears that naturalized Americans have a 5.1% higher retention rate than do the native-born Americans. However, in the Mountain region, native-born Americans have a 5.1% higher retention rate than do the naturalized Americans.

When we compare the supply of and demand for the native-born Americans, we find that they tend to match for each region, while the East North Central region appears to have a slightly larger supply than demand (3.7%) and the South Atlantic region appears to have a slightly lower demand than supply (4%). However, for naturalized American citizens, the picture appears to be quite different, with the West North Central region appearing to have an oversupply (4.8%) and the Pacific region an undersupply (7%).

[1] The average retention rates calculated from Table 4.3.

Table 4.3 Comparison of retention rate for native-born and naturalized S&E PhDs by graduation region and citizenship

S&E PhD graduates' region	Native-born (%) Employer region: 01 New England	02 Middle Atlantic	03 East North Central	04 West North Central	05 South Atlantic	06 East South Central	07 West South Central	08 Mountain	09 Pacific	Graduate (column %)	Total
01 New England (row %)	26.3	15.3	9.9	3.5	17.7	1.6	4	6.7	15	9.6	373
02 Middle Atlantic (row %)	11.8	23	12.4	4.6	20.1	2.5	5	7	13.7	12.4	483
03 East North Central (row %)	8.6	12.7	21.3	8.3	15.7	6.3	6.2	9.3	11.7	15.8	616
04 West North Central (row %)	7.5	9	14.5	26.2	13.6	3.3	7.8	8.1	9.9	8.5	332
05 South Atlantic (row %)	8.1	12.1	9.8	5.8	27.4	6.2	11.1	7.5	11.9	13.6	530
06 East South Central (row %)	3.1	12.4	10.6	5.3	19.9	31.4	8	3.5	5.8	5.8	226
07 West South Central (row %)	3.3	8.2	11.3	6.4	15.9	6.9	27.8	8.7	11.3	10	389
08 Mountain (row %)	7.5	8.3	6.4	5.6	10.9	3.7	9.1	29.9	18.7	9.6	375
09 Pacific (row %)	10.1	13.1	9.3	6.1	15.4	2.4	6.3	13.1	24.1	14.7	572
Total employment % by region (row %)	9.8	13	12.1	7.6	17.6	5.8	9.2	10.6	14.2	100	
Total employment by region	382	506	473	297	686	227	358	412	555		3,896

Table 4.3 (continued)

	Naturalized (%)										
	Employer region										
S&E PhD graduates' region	01 New England	02 Middle Atlantic	03 East North Central	04 West North Central	05 South Atlantic	06 East South Central	07 West South Central	08 Mountain	09 Pacific	Graduate (column %)	Total
01 New England (row %)	25.7	11.4	15.7	5.7	15.7	1.4	5.7	1.4	17.1	10.9	70
02 Middle Atlantic (row %)	12.1	28	12.1	3.7	11.2	1.9	6.5	4.7	19.6	16.7	107
03 East North Central (row %)	4.9	15.9	24.4	4.9	14.6	2.4	9.8	4.9	18.3	12.8	82
04 West North Central (row %)	2.9	17.6	14.7	19.1	13.2	2.9	5.9	7.4	16.2	10.6	68
05 South Atlantic (row %)	6.3	22.1	9.5	2.1	29.5	3.2	6.3	2.1	18.9	14.8	95
06 East South Central (row %)	10.5	5.3	7.9	2.6	18.4	34.2	5.3	0	15.8	5.9	38
07 West South Central (row %)	7.8	7.8	4.7	7.8	20.3	0	32.8	4.7	14.1	10.0	64
08 Mountain (row %)	2.6	12.8	10.3	5.1	15.4	2.6	7.7	23.1	20.5	6.1	39
09 Pacific (row %)	13.9	12.7	10.1	2.5	10.1	2.5	10.1	7.6	30.4	12.3	79
Total employment % by region (row %)	10	16.5	12.6	5.8	16.5	4	9.8	5.5	19.3	100	
Total employment by region	64	106	81	37	106	26	63	35	124		642

Table 4.4 Retention of S&E graduates by ethnicity and by region

In-state % by ethnicity and region	01 New England	02 Middle Atlantic	03 East North Central	04 West North Central	05 South Atlantic	06 East South Central	07 West South Central	08 Mountain	09 Pacific	Total in-state number	Retention % by ethnicity	Total graduate by ethnicity
Native-born American												
White	15.4	17.4	18.8	19.7	10.9	19.8	27.9	20.7	17.7	528	17.9	2,949
Hispanic	33.3	16.3	9.7	12.5	8.5	0	29.3	20.6	21.3	53	17.79	298
Black	18.8	5.7	33.3	16.7	15.5	44.4	22.9	23.1	34.5	68	23.21	293
Asian	28.1	15	15	12.5	11.4	0	7.1	14.3	36.7	56	22.13	253
Native American	25	11.1	33.3	33.3	16.7	44.4	19	12.5	9.1	22	21.15	104
Naturalized American												
White	23.8	8.7	33.3	25	30.8	25	29.4	30.8	29.4	44	26.35	167
Hispanic	0	0	0	0	40	100	16.7	0	0	7	22.58	31
Black	0	22.2	20	0	0	33.3	0	0	100	5	17.86	28
U.S. naturalized Asian	12.2	18.6	13.5	23.1	11.8	40	28.2	16.7	14	71	17.11	415
Total no. retained in region	446	612	554	334	792	253	421	447	679	854		
Region's retention %	17.5	16.2	19.1	19.8	12.1	26.1	26.4	20.1	20.9			
Region's % of total retention	9.8	13.5	12.2	7.4	17.5	5.6	9.3	9.9	15	18.8		4,538

Note: This table also shows those who are native-born and naturalized American citizen S&E PhDs in the 2001 NSF data set.Source: National Science Foundation, Division of Science Resources Statistics (2001) Survey of doctorate recipients in the United States.

Table 4.4 displays the regional retention rates by ethnicity. The regional distribution of ethnic retention rate among different ethnicities is close to equal at around 20%. However, the overall retention rates across regions are not as close to even. The second to last row shows the retention rate of a region as a percentage of the total number of science and engineering PhDs in the data set. For example, it appears that the South Atlantic region has a fairly low overall retention rate at 12.1%. Also, the Asian and Hispanic retention rates within the South Atlantic region are much lower than those in other regions, even though it is the largest producer of the total number of PhDs. One might argue that this region is experiencing a brain-drain problem, especially if it is unable to attract graduates from other regions. In addition, it appears that some regions deserve particular attention. For example, in the Pacific region, although the overall retention rate is not high (20.9%), the rates of some ethnic groups are much higher, for example, native-born African Americans (34.5%), native-born Asian Americans (36.7%), and naturalized white American citizens (29.4%). Also, the East South Central region has a high retention rate of native-born African American citizens (44.4%) and native Americans (44.4%). Overall, this table shows that the interactions between citizenship and ethnicity may matter in retention rate, as was argued in Sect. 4.2. For example, as shown for the Pacific region, the retention rate of naturalized Asian American citizens is particularly low at 14%, while the native-born Asian American citizen rate is quite high at 36.7%. This is puzzling, as the two groups have similar skill levels and both are of Asian origin. These findings focus our attention on the role of ethnicity and citizenship in spatial assimilation. This is what we tackle in Sect. 4.4.

4.4 Hypotheses and Data

4.4.1 Hypotheses

We form the following hypotheses on spatial assimilation based on the observations in the previous sections:

Hypothesis 1 Citizenship and ethnicity in general matter to spatial assimilation and to the spatial distribution of PhD holders.

Hypothesis 2 Professional specializations, employers' industry, and family structure are typical variables that play a distinguishing role for the in-state retention rate of naturalized and native-born Americans.

Hypothesis 2a The term *home-state effect* summarizes the possibility that the individuals' local experience of being native to a location and having a local undergraduate degree may also affect the in-state retention rate. It may be that these factors will affect not only the individual's personal adaptation to local working conditions but also his or her connection to the local social networks, which might be decisive in his or her location choice.

Hypothesis 2b There is a systematic difference between native-born and naturalized Americans regarding their location choice (state) when they look for a job . This is interesting, particularly when considering two groups with the same ethnicity.

Hypothesis 2c The utilization of school-related social networks may be very different between the two groups. Hence, it might be that the first job location would differ between native-born and naturalized Americans as a result of their spatial assimilation.

Hypothesis 2d A state's R&D spending and availability of post-doc positions may also affect the two groups differently if their spatial assimilation is different.

4.4.2 The Data

The data on doctoral scientists and engineers come from the 2001 Survey of Doctorate Recipients (SDR), a longitudinal panel survey of individuals who have received their doctorates in science or engineering (S&E).[2] Since the 1970s, this study has been conducted biennially for the National Science Foundation (NSF) and other federal sponsors.[3]

One of the advantages of the 2001 SDR is that it retains the changes in the questionnaire that were implemented in 1993.[4] A large set of core data items is conveyed from year to year to enable trend comparisons. Also in each survey year, different sets of module questions on special topics of interest are included.[5] In 2001, a special module on publication and patenting, which was introduced in 1995, was again used as a field for activities during the preceding 5-year period. Also in 2001, new questions were added on individual satisfaction and the importance of various job attributes. The multiple-race question, as mandated by the U.S. Office

---

[2] The information in this section is retrieved from National Science Foundation (2001, 2003).

[3] The sampling frame for the 2001 SDR was compiled from the Doctorate Records File (DRF) to include individuals who (a) had earned a doctoral degree from a U.S. college or university in an S&E field; (b) were U.S. citizens or, if non-U.S. citizens, had indicated that they planned to remain in the United States after their degree was awarded; and (c) were under 76 years of age.

The 2001 frame consisted of the 1999 SDR sample, supplemented with new S&E doctorate graduates who had earned their doctoral degrees since the 1999 survey and who met the conditions listed here. Those who were carried over from 1999 but had attained the age of 76 (or were deceased) were deleted from the frame. The survey had two additional eligibility criteria for the survey target population. The sampled member had to be a resident of the United States and not institutionalized as of the survey reference week (April 15, 2001; National Science Foundation, 2001).

[4] The U.S. Census Bureau conducted the survey for the NSF in 2001. Data collected in the SDR is part of the Scientists and Engineers Statistical Data System (SESTAT) surveys that are sponsored and maintained by the NSF. Additional data on education and demographic information in the SDR come from the Survey of Earned Doctorates (SED), an annual census of research doctorates earned in the United States that has been ongoing since 1920, which forms the Doctorate Records File (DRF). The overall unweighted response rate for the 2001 SDR was 82.2%.

[5] For example, the 1995 SDR questionnaire has a post-doc module, and the 1997 version has special modules on alternative work arrangement, job security concerns, and recent doctorates' initial career experiences.

of Management and Budget (OMB), was also added in 2001 for research evaluation and future sampling purposes.[6]

In 2001, the SDR sample size was 40,000.[7] To ensure that the sampling rate of the new cohort was at least 15% higher than that of the old cohort, 4,000 of the total sample were drawn from the new cohort group.[8] The remaining 36,000 sample cases were then divided so that the nearly new cohort would have a 10% higher sample allocation than the old cohort.

Our sample consists of a hierarchical random sampling of the 2001 SDR data. The selection criteria were chosen according to the respondents' observable characters such as location of school awarding highest degree, graduation year, first S&E PhD graduation year, location of school, degree field of major, region of employer, employer sector, citizenship, and gender. The resulting sample size is 6,279. To correct the survey design effect, we use probability weight to form a cohort in all the regression models that follow.

4.4.3 Variables and Definitions

We will compare the in-state employment choice by using a discrete choice model. The set of variables can be divided into the following categories (see Appendix 4.1 at the end of the chapter for variable labels and further details).

4.4.3.1 The Dependent Variable

The dependent variable that we use is "in state," which takes the value of "0" when the respondents' reply indicate that their employment state is different from that of their first S&E PhD degree-awarding state. Alternatively, it will take the value of "1" when both are the same state. This dependent variable simplifies the multiple-destination choice by reducing the location choice to a dummy variable. We also

---

[6] The 2001 SDR data collection consisted of two phases: a self-administered mail survey, followed by computer-assisted telephone interviewing (CATI) of a sample of the non-respondents to the mail survey. The mail survey consisted of an advance letter and then two mailings of a personalized questionnaire package, with a reminder postcard between the first and second questionnaire mailing.

[7] The total sample was selected from three groups: (a) old cohort cases with doctoral degrees earned prior to July 1, 1994; (b) nearly new cohort cases with doctoral degrees earned between July 1, 1994, and June 30, 1998; (c) new cohort cases with doctoral degrees earned between July 1, 1998, and June 20, 2000.

[8] The goals of the 2001 SDR sample design include the following: (a) reduce the variation in the sampling weights of the old and nearly new cohorts; (b) allocate the sample so that the variance of overall population estimates are minimized; (c) allocate the sample to assure that the sampling rate of the new cohort is at least 15% higher than that of the old cohort; (d) allocate the sample to assure that the sampling rate of the nearly new cohort is at least 10% higher than that of the old cohort; (e) adjust the sample location if any large stratum receives a disproportionate amount of sample.

control the regional variable by the employers' regions as well as the region of the first S&E PhD graduation.

4.4.3.2 Independent Variables

Personal character and family background are quite significant in an individual's decision to consider whether he or she will remain in the area in which he or she studied. The variables in this category are gender, ethnicity, marital status, number of children in different age intervals (younger than age 18), as well as spouse's working status.

The ethnicity variable involves Chinese and Indians as representatives of Asian graduates, comparing them to other ethnic groups in term of spatial distribution.

The professional field specialization variable involves "computer and information science" as a reference for the dummy variable for comparing the different effects on the likelihood of being in state. In particular, the following fields are considered:

1. Computer and information science
2. Mathematical sciences
3. Biological and agricultural sciences
4. Health sciences
5. Physical and related sciences
6. Social sciences
7. Psychology
8. Engineering

We also construct two discrete variables to reflect the **home-state effects**. The first variable is obtained from the following item: "Indicate if the first S&E PhD degree-granting state is the same as the respondents' birth state." This variable focuses on the likelihood of those who are native to the state to settle in the same state in which they were born and obtained their first PhD degree. The second discrete variable was created from another survey item: "Indicate whether the undergraduate state is the same as the first PhD-awarding state." This variable captures the effect of the state in which one has received his or her undergraduate degree on the individual's future location choice, that is, after one has received his or her PhD degree. The dummy variables take the value "0" if different states and "1" otherwise.

We employ the state's 1998 **Research and development** (R&D) funding to investigate the R&D effect on the likelihood of its retaining high-tech human capital, that is, inducing these individuals to stay in the original state. We use lagged R&D to enable us to observe the causality between R&D and labor mobility. It appears that educated individuals move with their own research funding. We also construct some interaction terms to see some interactive effects with the state's R&D funding. For example, we use states' 1998 R&D funding and interact it with variables on industrial sectors and ethnicity.

Another variable is **social capital for job searching and hunting**. As we discussed earlier, Americans who are native to a state might have stronger social networks than naturalized Americans in general, simply because they were born and grew up in the same state. Some of the local social networks, such as family connections, may be utilized in job hunting by native-born Americans, but this is not applicable to naturalized Americans. In general, one would expect the school's social network effect to be more important in the first job search for fresh S&E PhD graduates. To capture this difference in utilization of the school social capital for job location, we generate a variable from an SDR survey question: "Which resource is the most effective for finding your first career path job?" The hypothesis is that if both groups benefit to the same degree from utilization of the school's resources such as advisors and the school's career centers, given other related factors such as their major and characteristics of the job, then the school's social capital shows that job assimilation works when one is searching for a job. The underlying observation is that native-born Americans may be in a better position to mobilize social capital than are naturalized Americans. Alternatively, naturalized Americans may rely more on the school's social capital than do their native-born counterparts. If two groups of Americans systematically differ in their use of the school's social capital, we will observe a very different pattern in their priority of utilization of resources.

Specifically, survey respondents answered the following question: Which resource was the most responsible for finding your first career path job? The choices include these:

1 = Faculty or advisors
2 = Formal institutions: Professional recruiters such as "headhunters"; college or department placement offices; professional meetings
3 = Media: Electronic postings; newspapers; professional journals
4 = Informal channels through colleagues or friends
5 = Direct contacts to companies and others

We use a series of dummy variables, with faculty or advisors as reference (with value = 0), to compare with other factors (with value = 1 for each factor) to observe the different patterns of the two groups' reactions to questions.[9]

The post-doc effect on the location choice is the last variable we consider. The post-doc effect may be more important to naturalized Americans than to native-born Americans simply because the naturalized Americans may have fewer social networks for job hunting than the native-born Americans. They may then rely on a post-doc position if they cannot find a permanent job when they graduate. Thus, it may be that naturalized Americans are more responsive to post-doc positions than are native-born Americans.

[9] Please see the data descriptions in the Appendix at the end of the chapter for details on the dummy variables.

4.4.3.3 Control Variables

In addition to the mentioned independent variables, we introduce two sets of control variables:

Individual ability is likely to help the individual change his or her location freely. It is also the case that creative activities that reflect individual abilities may differ across fields. We document personal creative activities by variables such as books accepted for publication, papers presented in conferences, patents applied for, patents granted, and patents commercialized.

Employment considerations and job satisfaction may be conducive to the location choice. These set of variables are derived from the question "When thinking about a job, how important is each of the following factors to you?" Respondents are asked to rank the factors provided in the survey—opportunities for job advancement, benefit, challenge, location, respect, security, salary, and contribution to society—from low to high in order of importance in relation to job opportunities.

We document also the year in which the first PhD was awarded, which we call the "cohort effect" (in the 1991–2000 range), as control variable.[10]

4.5 Methodology and Results

The empirical strategy is straightforward. We will use a simple logit model to estimate the likelihood of a respondent's staying and working in the state where he or she earned his or her first science and engineering PhD, which we call the "in-state decision." The dependent variable, y, is "in state" and takes the value "0" when the state in which the respondent's employer is located is different from the state in which the respondent received his or her first S&E PhD; it takes the value "1" when both the employer's state and the state in which the S&E PhD was granted are the same. Certainly, the data has the limitation of not being able to capture the circular migration or the job relocation more than once after the PhD recipient has graduated. In other words, we cannot trace the locus of job change, such as the relocation from the graduate's original state to another state and then back to the original state. One way to get around this problem is to introduce the "cohort dummy variable," which captures the year and the state where respondents earned their first S&E PhD. This cohort dummy captures the likelihood of their staying when the year of graduation changes.

[10] Beenstock et al. (2005) create a longitudinal data set by matching immigrants in Israel's censuses for 1983 and 1995. They show both the Immigrant Assimilation Hypothesis (IAH) and the synthetic cohort methodology (SCM) as rejected. They suggest that SCM is subject to survivor bias, which increases the apparent degree of assimilation. Because of the increased return to destination-specific skills during this period and increased immigration, the assimilation curve changes shape in a way that makes it difficult to estimate even using panel data. However, if the period is not very long and the amount of immigration change has not fluctuated greatly, the cohort effect could pick up some of the assimilation effect, as argued in Beenstock et al. (2005).

The estimation equation is constructed as follows:

$$y^* = \beta_0 + X\beta + u \tag{4.1}$$

The in-state decision, y^* , is dependent on a vector of independent variables, X, where X is a set of explanatory variables and a *row* vector containing elements $(x_1, \ldots, x_k)$, and β is a *column* vector containing corresponding coefficients: $\beta' = (\beta_1, \beta_2, \ldots, \beta_k)$. The term u is the random error term with a normalized mean (0) and variance (1). We hypothesize that the dependent variable takes the following values:

$y = 1 \mid y^* > 0$
$y = 0$ otherwise

where the likelihood of observing y is expressed as the following, given the normality assumption:

$$\begin{aligned} P(y = 1|x) &= E(y^* > 0|x) \\ &= P[e > -(\beta_0 + X\beta)] \\ &= 1 - G[-(\beta_0 + X\beta)] \\ &= G(\beta_0 + X\beta) \end{aligned} \tag{4.2}$$

In the logit model we assume that e has a logistic distribution such that the probability density function of u is given by

$$g(u) = \frac{e^u}{(1 + e^u)^2} \tag{4.3}$$

where G is the cumulative distribution function of u:

$$G(u) = \frac{e^u}{1 + e^u} \tag{4.4}$$

Since the log of the odds that $y = 1$ is a linear function of the explanatory variables, the marginal effect in the logit model can be expressed as

$$\frac{\partial P(y = 1)}{\partial x_i} = G'(\beta_0 + X\beta)\beta_i = \left(\frac{e^{\beta_0 + X\beta}}{1 + e^{\beta_0 + X\beta}}\right)\beta_i \tag{4.5}$$

for $i = (1, n)$.

This model is a preliminary step in investigating the spatial assimilation of highly skilled naturalized Americans. In the current model, we assume that the graduates' choice of employment location depends on the industry in which they apply and accept offers first. The variable "job location" is critical among the set of variables on the "preference of job satisfaction" because it clearly identifies the personal preference for a job with in-state decision. This way of modeling the estimation, though congruent to the theoretical setup, could be improved by, for example, applying a multinomial logistic model that takes the choice of industry and in-state decision into account simultaneously. Also, selectivity models, such as the Heckman model, could be used to account for selectivity and wage. In addition, one could use the gravity model, which takes both the origin and destination variables into account more rigorously.

For estimation purposes, we first pool samples on native-born Americans and naturalized Americans in the first model and then separately estimate individual group in second and third models, respectively. In doing so, we assumed a common error structure in the first model; however, in the separate sample models, we do not assume the same error structure for these two groups. Another advantage of separate sample model for estimation purposes is that in order to identify the "assimilation" effect, we rely on the differences in magnitude, sign, and significance levels for comparing different groups' variables. The basic idea is that if there is no assimilation problem, the independent variable should show similar magnitude, sign, and significance levels, once other independent variables such as individual, family, school, and personal preference are controlled for.

Table 4.5 shows comparisons among full sample, the native-born, and naturalized Americans. The results show a very interesting pattern: In the pool sample model, the "citizenship" dummy shows a difference between the native-born and the naturalized Americans. In particular, the marginal effect of citizenship shows a 5% higher chance for naturalized Americans to stay in the state in which they receive their first S&E PhD. The last row of the two separate citizenship models in Table 4.5 also shows that the predicted probability of naturalized Americans staying in state is 0.16, which is larger than that of native-born Americans who are also native to the state in which they received their first S&E PhD (0.13). This strong, positive, and significant effect on the "naturalized" dummy variable might partially detect the "lagged effect" of the visa restriction on naturalized Americans. In other words, our model shows that in both the pooled and the separate sample models, naturalized Americans are more likely to stay in the state in which they received their first S&E PhD than native-born Americans. At this point, this is the natural question to ask: "What factors explain the difference between the two groups in their likelihood of staying in state?"

We first discuss the factors that might influence the differences between the native-born and naturalized Americans' decisions to stay in state. We then provide a policy recommendation in the last section.

First, we observe only one consistent pattern of **personal character and family effects** in relation to the in-state decision, and this is related to gender. Females, in general, are more likely to stay in state regardless of their citizenship status. The marriage effect shows that local singles are less likely to stay in state than married people; however, this observation does not apply to naturalized Americans. While ethnicity does not play any role for Americans native to the state, we found

Table 4.5 Comparison between native-born and naturalized Americans on location choice for the first S&E PhD

	FullSample		Native-born American		NaturalizedAmerican	
Variable	dy/dx	SE	dy/dx	SE	dy/dx	SE
Citizenship (naturalized = 1)	0.05**	(0.02)	–	–	–	–
Personal and family characteristics						
Gender (female = 1)	0.04***	(0.01)	0.03***	(0.01)	0.09***	(0.04)
Marital status (single = 1)	–0.04***	(0.01)	–0.04***	(0.01)	–0.03	(0.04)
Ethnicity (Asian = 1)	–0.01	(0.02)	0.02	(0.02)	–0.06*	(0.03)
Ethnicity (other minority = 1)	0.00	(0.01)	0.01	(0.02)	–0.06*	(0.03)
No. of children aged 12–17	0.02*	(0.01)	0.03**	(0.01)	–0.02	(0.03)
No. of children aged 6–11	0.00	(0.01)	0.01	(0.01)	–0.05**	(0.03)
No. of children aged < 6	–0.03*	(0.01)	–0.02*	(0.01)	–0.06**	(0.03)
Spouse work (part time = 1)	0.02	(0.02)	–0.01	(0.02)	0.24**	(0.10)
Spouse work (no work = 1)	–0.01	(0.02)	–0.02	(0.02)	0.07	(0.05)
Field specialization and home-state effect						
Field of major for first S&E PhD (major group) (reference = computer and information science)						
Mathematical sciences = 1	–0.05**	(0.02)	–0.04	(0.03)	–0.09***	(0.03)
Biological and agricultural sciences = 1	–0.10***	(0.02)	–0.06***	(0.02)	–0.16***	(0.03)
Health sciences = 1	–0.04	(0.02)	–0.01	(0.03)	–0.09***	(0.03)
Physical and related sciences = 1	–0.07***	(0.02)	–0.04	(0.03)	–0.11***	(0.03)
Social sciences = 1	–0.09***	(0.02)	–0.06***	(0.02)	–0.15***	(0.02)
Psychology = 1	–0.07***	(0.02)	–0.04	(0.03)	–0.11***	(0.02)
Engineering = 1	–0.05***	(0.02)	–0.02	(0.03)	–0.12***	(0.03)
(Birth state = PhD state) = 1	0.10***	(0.02)	0.09***	(0.02)	–	–
(Undergrad state = PhD state) = 1	0.11***	(0.02)	0.12***	(0.02)	0.05	(0.05)

Table 4.5 (continued)

	Industry and R&D					
(other industries as reference)						
(Biotechnology, information technology and research) = 1	–0.03**	(0.01)	–0.01	(0.02)	–0.09***	(0.03)
(Educational institution) = 1	–0.05***	(0.01)	–0.04***	(0.01)	–0.07**	(0.03)
State R&D expenditure (1998)	0.00**	(0.00)	0.00*	(0.00)	0.00	(0.00)
	Social capital for job hunting and post-doc effects					
(academic advisor as reference)						
Formal institutions = 1	–0.06*	(0.03)	–0.06**	(0.03)	–0.04	(0.09)
Media = 1	–0.09***	(0.02)	–0.09***	(0.02)	–0.12***	(0.03)
Informal channels = 1	0.01	(0.03)	0.01	(0.03)	0.00	(0.08)
Direct contacts = 1	–0.02	(0.03)	–0.02	(0.03)	–0.08	(0.05)
Post-doc = 1	0.06***	(0.02)	0.04**	(0.02)	0.08	(0.07)
˙cons	–1.09*	(0.67)	–1.08	(0.71)	–2.69	(2.65)
No. of observation	4,538.00		3,896.00		642.00	
Wald χ^2	409.82		373.08		140.52	
Prob > χ^2	0.00		0.00		0.00	
Log pseudo-likelihood	–1944.58		–1645.84		–244.54	
Pseudo R^2	0.11		0.12		0.23	
Predicted probability	0.15		0.13		0.16	

***, **, * Significantly different from zero at the 1, 5, and 10% level, respectively. Except continuous independent variables, all other dy/dx is for discrete change of dummy variable from 0 to 1.Control variables are year of graduation, employers' regions, graduation regions, personal ability, employment considerations, and job satisfaction, income per capita (2001), and population size (2002).All coefficients reported are marginal effect.Source: National Science Foundation, Division of Science Resources Statistics (2001) Survey of doctorate recipients in the United States.

naturalized Americans, both Asian and those of other minorities, less likely to stay in the same state than were white naturalized Americans. This may be due to the fact that local-born minorities have fewer problems assimilating than do naturalized minorities when compared to native-born white holders of S&E PhDs. The variable **work status of the spouse** also shows diverse results when we compare the native-born and naturalized Americans. In particular, we find a slight positive effect of the spouse's working at a part-time job vis-à-vis a full-time job for naturalized Americans, but this variable does not seem to matter to native-born Americans. This may be due to the high probability that naturalized Americans will have spouses who are also naturalized Americans or foreign citizenship which might prevent the spouse from working full time. Child bearing is also observed as having different effects across the two groups. Naturalized Americans are observed to have a negative attitude to staying in state when their children are under age 6 or the number of their children between ages 6 and 11 increases, while local-born Americans with children between ages 12 and 17 are more likely to move to another state.

When we compare across the **major fields of PhD's**, we find large differences between the groups. While most native-born Americans graduate with majors in biological/agricultural sciences or social sciences, they are less likely to stay in state than their counterparts who graduate with majors in computer and information science. Naturalized Americans are observed to be much more mobile in all the seven fields than who graduate with majors in computer and information science.[11] Also, the coefficients of the naturalized Americans' significant professional variables are larger (with a range of 9–16%) than those of native-born Americans. In other words, naturalized Americans are more likely to stay in state if their major is in computer and information science than if their major is in one of the other fields. On the other hand, there are serious mismatches in the supply of and demand for naturalized American PhDs in state-level economies because those in all fields other than computer and information science are less likely to stay in state after their graduation.

Two surprising results seem to emerge when we compare the **home-state effects** dummies between the two groups. While native-born Americans are observed to be affected by both of the two home-state effects variables, there was no effect at all on the naturalized Americans. For example, while the native-born Americans have a quite strong (12%) "in-state effect" for being educated in undergraduate and graduate studies in the same state, there is no such effect on the naturalized Americans. Putting these differences into perspective, one may reflect that naturalized Americans are not bound by the "in-state effect." In other words, naturalized Americans may be more likely to look for a job that best matches their interest across the nation than are native-born Americans.

[11] The seven majors that show a negative significant effect are mathematical sciences, biological and agricultural sciences, health sciences, physical and related sciences, social sciences, psychology, and engineering. That means these seven majors are less likely (a range from 9 to 15%) to stay in state than those from computer and information science.

States' R&D funding has a positive effect on native-born Americans but not on naturalized Americans, nor does the industry that native-born Americans enter after graduation have a positive effect to the in-state probability. However, naturalized Americans show a strong negative effect when they switch to research-oriented industries from other industries. This implies that large research-oriented industries across the United States rely on immigrant human capital.

For the **social capital** variables, we find that native-born Americans choose formal institutions such as department career centers, headhunters, professional meetings, and media as the most important source of their first job, which would then be less likely to be in state than those who choose advisor networks. However, naturalized Americans select (marginally significantly) the faculty advisor rather than the media. This observation is consistent with other variable explanations in the original model, which states that naturalized Americans lack the advantage of having the social capital of native-born Americans to stay within the state in which they received their PhD. However, the post-doc position may help both native-born and naturalized Americans stay in state, while naturalized Americans have a larger coefficient and a strong post-doc effect, when we look at the bottom row of the last four columns in Table 4.5. The interpretation may well be that while a school's social capital may not be very helpful for naturalized Americans who want to remain in the state in which they obtained their PhD, other variables such as adjustment of industrial policy may help retain their "brains" in state.

4.6 Conclusions and Discussion

In this chapter, we have examined and compared native-born Americans with naturalized Americans with regard to their decision to stay in the state in which they received their first S&E PhD. This project is particularly important since this highly skilled human capital is trained in the original state and the training is often paid for by that state. If the state is not able to retain these highly skilled workers it trained, it is likely to suffer from the loss of human capital, the brain-drain problem.

Our interest in comparing the native-born and naturalized Americans is because, by definition, the naturalized Americans are first-generation immigrants who usually have less social capital than their native-born counterparts. These naturalized Americans are of particular interest because, while they are not originally American, they stay in the country for a long-term purpose (usually because naturalization takes a long time and they already have family in the country). Their situation is quite different from that of foreign students, who have visa issues when dealing with their mobility and job searches.

With a hierarchical random sample of National Science Foundation 2001 SDR data and a logistic model, we calculated the likelihood of S&E PhD graduates' staying in the state in which they received their first S&E PhD (we call this the in-state decision). These individuals' decision to stay in the state in which they received the first PhD may be a function of a set of related variables such as personal character, creative abilities, family factors, field specialization, educational institution characteristics, employer's field and the state's R&D policy, social capital for the job search, and post-doc placement. We have shown that these naturalized Americans are very different from native-born Americans either in significance or in magnitude on almost every observable attribute that affects their decision to stay in the state.

With regard to the assimilation problem of these naturalized Americans, states designing educational and industrial policies, may want to take into account the factors that may affect native-born and naturalized Americans differently when they are making decision about whether to stay in state.

For native-born Americans, states may try to target unmarried males with biological/agricultural and social science degrees since male PhD graduates with these two majors are the most likely to "drain away." Increasing opportunities for undergraduate and graduate education, R&D funding, and post-doc positions for local residents may also help to "bank the brains" in the state as well.

As for naturalized Americans, they are, in general, 3 percentage points more likely to stay in state than are the local-born Americans. This higher percentage may be due to the "lagged visa" effect, which is due to the requirement for citizenship applicants to stay in a particular position for a long-enough period to be granted citizenship. Given this higher predicted in-state percentage point for naturalized Americans, there are also many assimilation issues that need to be addressed. In our analysis, naturalized Americans do not show any "home-state" effects in the way that native-born American citizens do, and this may be because they lack the social capital to remain in the state. Indeed, some ethnicity-assimilation programs may need to be considered since the likelihood that S&E PhD recipients who are naturalized Asian and other minorities will stay in state is less than that of their naturalized white counterparts. For example, increasing opportunities for their spouses to work locally could also be a viable strategy to reduce brain drain. One major challenge is the naturalized Americans' choice of major. Our study shows that computer and information science majors are the most likely of all majors to move out of state upon completion of their first PhD. This issue may signify the mismatch between production of PhDs with certain field majors/specialties and local industrial demands for naturalized Americans. Moreover, biotechnology, information technology, research, and education majors are more likely to find jobs out of state than are majors in other fields. This tendency alone is worth a state's reconsidering its industrial policy. Furthermore, states may consider strengthening naturalized Americans' local social capital as a strategy to keep them in state after graduation. For example, states may consider providing more networking opportunities for naturalized Americans to the development local social capital in their graduate schools, work places, and community they live.

In summary, states that need to adjust for the effect of brain drain might want to reconsider their industrial structure, their higher-education policies, and the fields in which their universities are producing PhDs. State governments may also want to work with universities and local businesses to build up a "ladder" of career paths to encourage new PhD graduates to start their careers in the same state in which they graduate from a post-doc and then achieve a permanent position. In terms of a school's social capital, states may want to invest more in local industrial networking to keep their PhDs in state because those who elect to use formal institutions and media for their job search are more likely to move out of state than are those who find their jobs through social networking.

In this chapter, we have compared the effects of native-born and naturalized Americans' likelihood of staying in the same state in which they receive their first S&E PhD. In the next chapter, we will move our analysis to a comparison of local-born Americans' likelihood of staying in state after receipt of their first S&E PhD to that of foreign graduates (those with H1-B visas). In particular, we connect the idea of cultural clustering and technical clustering to the likelihood of graduates' staying in the state in which they receive their first S&E PhD.

Appendix: Variable Descriptions

Variable Name	Variable Descriptions	Further Details
In-state		
Female	Gender	0 = Male (reference), 1 = Female
_Imarsta_2	Marital status	1 = Married (reference), 2 = Single (widowed, separated, divorced, never married)
_Iracebb_2	Ethnicity	1 = White (reference), 2 = Asian (Chinese and Indian), 3 = other ethnicity
_Iracebb_3		
Ch12_7	Children number at age 12–17	
Ch611	Children number at age 6–11	
Ch_6	Children number at age under 6	
_Ispowk_2	Spouse working (1 = yes, full time = reference)	2 = Yes, part time
_Ispowk_3		3 = No
_Isdrmemg_2	Field of major for first S&E PhD (major group)	2 Mathematical sciences
_Isdrmemg_3	(1 = Computer and information science)	3 Biological and agricultural sciences
_Isdrmemg_4		4 Health sciences
_Isdrmemg_5		5 Physical and related sciences
_Isdrmemg_6		6 Social sciences
_Isdrmemg_7		7 Psychology
_Isdrmemg_8		8 Engineering
_Isdryr_1991	Year of first S&E PhD	1991=1
_Isdryr_1992		1992=1
_Isdryr_1993		1993=1
_Isdryr_1994		1994=1
_Isdryr_1995		1995=1
_Isdryr_1996		1996=1
_Isdryr_1997		1997=1
_Isdryr_1998		1998=1

Variable Name	Variable Descriptions	Further Details
_Isdryr_1999		1999=1
_Isdryr_2000		2000=1
Birthse	Birth state equal to first S&E PhD state	0 = Not the same,
	(Note: naturalized American does not have time variable)	1 = Same state
Underse	Undergraduate state equal to first S&E PhD state	0 = Not the same,
		1 = Same state
Books	Number of books accepted for publication	
Papers	Number of papers authored or co-authored which have been presented at conferences	
Patentapp	Number of patent applications named as inventor	
Patentgan	Number of patent applications which have been granted	
Patentcom	Number of patent applications which have been granted and resulted in commercial products	
_Iind_1	Employer's industrial field (0 = all other industries)	1 = Biotechnology, information technology, research
_Iind_2		2 = Education institution
rd98	State R & D expenditure at 1998	
Inc_per_01	Income per capita of each state at 2001	
Pop02	Population size of each state 2002	

When thinking about a job, how important is each of the following factors to you: opportunities for advancement?

_Ifacadv_2	Importance of opportunities for advancement : (1 = very important = reference)	2 Somewhat important
_Ifacadv_3		3 Somewhat unimportant
_Ifacadv_4		4 Not important at all
_Ifacben_2	Importance of benefits:(1 = very important = reference)	2 Somewhat important
_Ifacben_3		3 Somewhat unimportant
_Ifacben_4		4 Not important at all
_Ifacchala_2	Importance of intellectual challenge : (1 = very important = reference)	2 Somewhat important
_Ifacchala_3		3 Somewhat unimportant AND Not important at all
_Ifacinda_2	Importance of degree of independence : (1 = very important = reference)	2 Somewhat important
_Ifacinda_3		3 Somewhat unimportant AND Not important at all

_Ifacloc_2	Importance of location: (1 = very important = reference)	2 Somewhat important
_Ifacloc_3		3 Somewhat unimportant
_Ifacloc_4		4 Not important at all
_Ifacrespa_2	Importance of level of responsibility: (1 = very important = reference)	2 Somewhat important
_Ifacrespa_3		3 Somewhat unimportant AND Not important at all
_Ifacsala_2	Importance of salary: (1 = very important = reference)	2 Somewhat important
_Ifacsala_3		3 Somewhat unimportant AND Not important at all
_Ifacseca_2	Importance of security: (1 = very important = reference)	2 Somewhat important
_Ifacseca_3		3 Somewhat unimportant AND Not important at all
_Ifacsoc_2	Importance of contribution to society: (1 = very important = reference)	2 Somewhat important
_Ifacsoc_3		3 Somewhat unimportant
_Ifacsoc_4		4 Not important at all

Which TWO resources were most responsible for finding your first career path job?

_Ipathpria_2		2 = Formal institutions: professional recruiters such as “head hunters”; and college or department placement office; professional meetings
_Ipathpria_3	(01 = Faculty or advisors = reference)	3 = Media: electronic postings; newspapers; professional journals
_Ipathpria_4		4 = Informal channels through colleagues or friends
_Ipathpria_5		5 = Direct contacts you initiated with company and others
_Ipathpria_6		6 = Logical skip: the question was asked since 1999 and for those job seekers
_Ipdix_2	Postdoctoral appointment indicator	No = 0 = reference; yes= 1

Source: National Science Foundation, Division of Science Resources Statistics (2001) Survey of doctorate recipients in the United States.

References

Alba RD, Nee V (2003) Remaking the American mainstream: Assimilation and contemporary immigration. Harvard University Press, Cambridge, Mass

Audretsch D, Stephan P (1996) Company – scientist locational links: The case of biotechnology. Am Econ Rev 86: 641–652

Beenstock M, Chiswick BR, Paltiel A (2005) Endogenous assimilation and immigrant adjustment in longitudinal data. IZA Discussion Papers 1840, Institute for the Study of Labor (IZA)

Borjas GJ (1984) The impact of assimilation on the earnings of immigrants: A reexamination of the evidence. National Bureau of Economic Research Working Paper No. 151

Borjas GJ (2006) Making it in America: Social mobility in the immigrant population. National Bureau of Economic Research Working Paper No. 12088

Chiswick BR (1977) Sons of immigrants: Are they at an earning disadvantage? Am Econ Rev 67: 376–380

Chiswick BR (1978) The effect of Americanization on the earnings of foreign-born men. J Polit Econ 86: 897–921

Chiswick BR (1999) Are Immigrants Favorably Self-Selected? Am Econ Rev 89: 181–185

Chiswick BR, Miller PW (2002) Immigrant earnings: Language skills, linguistic concentrations and the business cycle. J Popul Econ 15: 31–57

Chiswick BR, Lee YL, Miller PW (2002) The determinants of the geographic concentration among immigrants: Application to Australia. IZA Discussion Paper No. 462, Institute for the Study of Labor (IZA)

Chiswick BR, Lee YL, Miller PW (2006) Immigrants' Language Skills and Visa Category. Int Migr Rev 40: 419–450

Helsley R, Strange W (1990) Matching and agglomeration in a system of cities. Reg Sci Urb Econ 20: 189–212

Koo J (2005) Knowledge based industry clusters: Evidenced by geographic patterns of patents in manufacturing. Urb Stud 42: 1457–1505

Miller PW, Chiswick BR (1996) Ethnic networks and language proficiency among immigrants. J Popul Econ 9: 19–35

National Science Foundation, Division of Science Resources Statistics (2001) 2001 Survey of doctorate recipients in the United States

National Science Foundation, Division of Science Resources Statistics (2003) Characteristics of doctoral scientists and engineers in the United States: 2001. NSF 03-310

Smeets V, Warzynski F, Coupé T (2006) Does the academic labor market initially allocate new graduates efficiently? J Econ Perspect 20: 161–172

Stephan PE, Sumell AJ, Black GC et al (2004) Doctoral education and economic development: The flow of new PhD's to industry. Econ Devel Q 18: 151–167

Chapter 5
Ethnic and Technical Clustering: Native-Born Americans Versus Foreign S&E Graduates

Yiu Por Chen

5.1 Introduction

This chapter seeks to explain the differences in the spatial distribution of native-born American as opposed to foreign PhD holders in the fields of science and engineering (S&E), with a specific focus on graduates from China and India. In this chapter, unlike in Chapter 4, we highlight the effects of ethnic clustering and high-technology clustering on spatial assimilations. Further, we concentrate on H1-B visa holders as opposed to naturalized Americans. *Ethnic clustering* is defined as the inclination of immigrants to settle in places where their ethnicity is concentrated. *Technical clustering* refers to the tendency of workers with similar skills to move to places where the industries in which they work are concentrated. By connecting these two concepts, this study advances the debate on spatial displacement of foreign S&E PhDs and technology clustering to another level of analysis.

We ask: To what extent can the spatial concentration of, for example, Asians in California be explained by ethnic clustering and to what extent by technical clustering?[1] We focus on the effects of ethnic clustering, which could induce S&E PhDs from different nationalities to have different responsiveness to a region where there is a high concentration of their ethnicity. Our analysis shows that foreign graduates from China are more responsive to Asian communities. However, Indian graduates do not appear to show any particular pattern.

We disentangle the technical clustering effects as *spillover effects* and *matching effects*. The spillover effects of technical clustering can be described as the enticement of R&D funding to a particular industry and to high-tech human capital's graduates' field of specialization to make a specific location choice. The matching effects of technical clustering relate to the matching between graduates' field of

[1] See Collaborative Economics (2005) for the recent ethnicity distribution in Silicon Valley.

Y.P. Chen
School of Public Service, DePaul University, 25 E. Jackson Blvd, Suite 1250, Chicago, IL 60604, USA

N. Aydogan, Y.P. Chen, *Social Capital and Business Development in High-Technology Clusters*, DOI: 10.1007/978-0-387-71911-5_5,

specialization and the types of industries in a particular state. We show that both effects are quite diverse among different nationalities.

In addition, we observe that the role of school-related social capital (social networks) and the effects of personal creativity are quite dissimilar, particularly between the locals and the foreigners in terms of their agglomeration patterns. Hence, the policymakers are in a position to tailor different educational, industrial, and community strategies to retain foreign or native-born American brainpower based on their different preferences and their relation to social capital. This study contributes to the literature by providing a microperspective to technical agglomeration and by incorporating different citizenships, ethnicity, and human capital.

This chapter is organized as follows: The next section reviews the recent trend of H-1B petitions by citizenship. Section 5.3 is a literature review, where we analyze the literature on technical clustering and ethnic clustering. In Sect. 5.4, by using the 2001 Survey of Doctorate Recipients (SDR) sample data, we first show the spatial distribution of S&E graduates in terms of patent numbers and their citizenship. We then match the distribution of the institutions in different regions with the most recent doctorate graduates and their current job location. Deduced from the observations in the preceding section, we derive several hypotheses and seek to explain the reasons behind the different location choices with two sets of logistic regressions in Sect. 5.5. These models show the effects of technical clustering and ethnic clustering on the likelihood of individuals staying in a particular state (or relocating out of the state) where their first PhD degree was granted. Section 5.6 is a discussion of the issues and the policy implications of our findings on ways to retain highly skilled human capital in the United States.

5.2 H-1B Petitions: A Brief Review by Citizenship

In this section, by focusing on the H-1B visa, we investigate the distribution of visa petitions by foreign citizenship and the effect of immigration policies. It is important to understand the effect of work visa on the foreign graduates' likelihood of staying in a state before we attempt to explain the differences in the job location choice between graduates who are foreign and who are native-born Americans.

According to a National Science Foundation (NSF) study around 47% of the foreign doctorate degree recipients for the year 1990–1991 were still working in the United States under a temporary visa in 1995 (National Science Foundation 1998). Table 5.1 shows the approved H-1B petitions by country of birth and type of petition (number) for fiscal years 2002 and 2003.[2] India and China are the top two birth countries among beneficiaries' origins. For example, observing all beneficiaries in 2003, in the second column, we find that around 30% are of Indian origin and around 10% are of Chinese origin. It appears that most Chinese and Indian

[2] For detailed distribution of H-1B holder in the United States for fiscal years 2002 and 2003, please see U.S. Department of Homeland Security, Office of Immigration Statistics (2004).

Table 5.1 H-1B petitions approved by country of beneficiary's birth and type of petition: fiscal years 2002 and 2003 (no.)

	All beneficiaries		Initial employment		Continuing employment	
Country of birth	Fiscal year 2002	Fiscal year 2003	Fiscal year 2002	Fiscal year 2003	Fiscal year 2002	Fiscal year 2003
Total	197,537	217,340	103,584	105,314	93,953	112,026
Country of birth known	197,092	217,031	103,350	105,185	93,742	111,846
India	64,980	79,166	21,066	29,269	43,914	49,897
China, People's Republic of	18,841	20,063	11,832	11,144	7,009	8,919
Canada	11,760	11,160	7,893	6,201	3,867	4,959
Philippines	9,295	10,454	6,648	6,852	2,647	3,602
United Kingdom	7,171	7,599	4,192	3,871	2,979	3,728
Korea	5,941	6,614	3,886	3,893	2,055	2,721
Japan	4,937	5,716	2,970	3,287	1,967	2,429
Taiwan	4,025	4,076	2,366	2,109	1,659	1,967
Pakistan	3,810	3,549	1,955	1,501	1,855	2,048
Germany	3,291	3,408	1,955	1,788	1,336	1,620
Mexico	3,082	3,407	1,905	1,969	1,177	1,438
France	3,145	3,190	1,925	1,621	1,220	1,569
Colombia	3,320	3,125	2,362	1,771	958	1,354
Russia	2,864	2,905	1,523	1,265	1,341	1,640
Venezuela	2,398	2,677	1,610	1,798	788	879
Brazil	2,287	2,354	1,414	1,307	873	1,047
Turkey	2,004	2,305	1,319	1,311	685	994
Argentina	2,148	2,230	1,611	1,479	537	751
Australia	1,846	1,925	1,107	986	739	939
Israel	1,620	1,841	1,042	1,085	578	756
Other countries	38,327	39,267	22,769	20,678	15,558	18,589
Country of birth unknown	445	309	234	129	211	180

Source: U.S. Department of Homeland Security, Office of Immigration Statistics (2004, Table 4A).

graduates who eventually settle in the United States come to the country with an F-1 student visa and then convert to an H-1B visa after they are employed in the country. Graduates from China and India are of particular interest to our research, not only because of the density of their population in the United States but also because of their general tendency to concentrate in science and engineering, particularly information technology (IT). According to a report by the U.S. Immigration and Naturalization Service, 70% of H-1B visa holders in 2000 are employed in either computer- or engineering-related occupations (U.S. Immigration and Naturalization Service, 2002).

In reality, foreign graduates may have mobility constraints on their H-1B visa applications. For example, the duration and limitation of a student visa are likely to induce foreign graduates to rush into a job at an institution where they can apply for a work visa. This may be especially important to recent graduates in science and engineering since they are likely to be locked in at certain locations, and it is likely to take several years before they will be able to relocate. We are interested in learning how "lock-in" may create imperative impacts on foreign graduates' spatial distribution patterns.

5.3 Literature Review: Technology Clustering, Ethnic Clustering, and Social Capital

In this section, we discuss the foreign graduates' settlement pattern, the concepts of technical clustering and ethnic clustering, and institutional factors that contribute to the mobility of foreign S&E PhDs.

Recently, the study of highly skilled labor migration within the United States has been combined with studies on economic geography and urban economics. These studies are particularly focused on the location of the firm and the factors that contribute to the local economic agglomeration and its effects (e.g., Feldman and Audretsch 1999). Economic agglomeration is hypothesized to result in increasing returns to scale (as explained in the previous chapters) and growth of the local economies. It could be that increasing returns might attract labor mobility because of the better job opportunities and networking possibilities in local economies.

The study on localization economies was pioneered by Marshall (1890). Market agglomeration at a location can be defined as a "cluster," which is characterized by the spillover effects of specialized supply and demand of certain types of products and skills (Porter 2000). Indeed, the factors that affect location choice also depend on the agglomeration benefits (Koo 2005a,b). Recent studies in new economic geography suggest that the increasing return to scale has been an important factor in inducing industrial clusters and urban agglomeration (Krugman 1991, 1996). Furthermore, recent reviews of spillover effects and urban agglomeration suggest that it is important to study these closely related concepts together (Koo 2005b).

In this chapter, we focus on technical clustering, which has emerged from the localization of industrial economies concept and which describes attraction for the mobility of highly skilled labor. Florida (2002) shows a high (0.72) correlation between talented human capital and high-technology industries, using the 1990 U.S. decennial census Public Use Microdata sample. Perhaps two of the most important measures of technical clustering effects on labor mobility are the spillover effect of industries and of R&D, on the one hand, and the matching effect between government educational institutions and industrial structures, on the other.

Certainly, government policies, such as R&D policy, local institution development policy (e.g., development of universities and research laboratories), and land use policy, may also play key roles in industrial clustering (Chakravorty et al. 2003).[3] More importantly, the question linked closely to these observations is: How sustainable is industrial clustering? We argue that the mobility of fixed and human capital sustains an industrial cluster due to its spillover and matching effects. [4] The geographical proximity of industrial sectors and research institutions, such as the location of university and specialized industrial fields, may have both complementary and opposite effects on high-technology human capital mobility. Social capital, generated from university and ethnic groups by immigrants who located to the United States earlier than the recent foreign graduates, may be very important to those recent foreign graduates in job placements and spatial distribution.[5]

The concept of *ethnic groups* in America is very complex to define. An ethnic group can be defined as an organic combination of immigration, race, culture, geography, and social groups (Phinney 1996). Studies show that different ethnic groups have assimilated into the mainstream society at varying speeds. For example, Alba and Logan (1991) show that the white ethnic group assimilated in the American mainstream society much better and faster than other ethnic groups. Ethnicity may pass from generation to generation even though a person is locally born in the United States. Further, socioeconomic groups may matter a lot to the second generation's performance.[6]

The association between *ethnicity and location choice* has been shown in labor mobility models that focus on the background characteristics and resource outputs

---

[3] For example, Chakravorty et al. (2003) study the formation of industrial clusters in three metropolitan areas in India and show that the fundamental factors in industrial clustering may be related to land use planning

[4] Using the R&D data, Helsley and Strange (1990) show that firms may be colocated with the mobility of labor pools. Fujita and Ogawa (1980) also show that talented labor pools may create "knowledge spillover" and attract the relocation of firms.

[5] Please see Chapter 4 for the discussion of a school's social capital.

[6] Borjas (1994) also argues that the living environment (such as residential segregation in an ethnic community) and "ethnic capital" (measured by the average skill level of an ethnic group's parent generation) may affect the second generation's future accumulation of human capital. He then uses the 1970 U.S. Census data to validate his claim and shows that these factors may explain second-generation immigrants' slow convergence of skill level. Borjas says that "Ethnicity has an external effect, even among persons who grow up in the same neighborhood when children are exposed frequently to persons who share the same ethnic background" (Borjas 1994, page 365). However, Borjas did not directly address immigrants' location choice problems in his studies.

of the backgrounds of individuals and households and their location attainment models.[7] Gross and Schmitt (2003) define a concept called *cultural clustering*, which signifies the advantage of being in a (smaller) immigrant group, when the immigrant student looks for a job or makes connections with firms that employ larger immigrant groups. These advantages start diminishing when the immigrant group expands in a particular industry in an area. This is because it becomes more difficult to verify someone's ability as the group grows. Also, one may feel more comfortable working and/or living with people from the same country. However, the concept of cultural clustering effects on spatial assimilations is still controversial.[8] We define *ethnic clustering* as the concentration of certain ethnic groups that may (or may not) constitute an attraction to newcomers from a particular country. The attraction (or repulsion) may be due to a common language or location preference. The definition of *ethnic clustering* is ethnicity based, a broader definition than that of *cultural clustering*, which is country based. This is because many countries can be considered to have the same ethnicity. For example, Chinese and Indians are both Asian. We believe a more general definition of an immigrant group by ethnicity of graduates, such as *ethnic clustering* in this study, can contribute to the understanding of the ethnic stock effect on the spatial concentration of high-tech foreign graduates of U.S. institutions.

---

[7] An example of this is the resources in the neighborhoods where they live (see Alba and Logan 1991, 1992, 1993; Logan and Alba 1993). This model posits a hierarchical ordering of neighborhoods and evaluates the way individual and household characteristics translate into placement in a particular area. Thus the location attainment model is conceptualized in the same manner as the status attainment model (Blau and Duncan 1967); both models reveal the way individual- or household-level characteristics are converted into access to the larger societal groups, such as occupations or communities, which form not only the stratification system within the society, but also the spatial displacement of ethnicity. Chiswick et al. (2002) also show that geographic concentration of immigrant groups is negatively related to age at immigration, duration of residence in the country to which the person immigrated (in their study, Australia), and the proportion of the birthplace group that is fluent in English. However, the extent of geographic concentration is also affected positively by the availability of ethnic media and the distance between the country of origin and the place of residence in the receiving country (Australia). Thus, family background, job satisfaction, and occupation aspiration are factors affecting the choice of job may depend on some cultural factors that affect the choice to stay within a state (of the United States).

[8] Using the 1980 U.S. Census of population in standard metropolitan statistical areas (SMSAs) to track immigrants, aged 22–64, into the United States, Bartel and Koch (1991) found no systematic evidence of cultural clustering. In particular, they noted "the high mobility rates of Asian immigrants were unrelated to the percentages of Asians in various cities, while Europeans who moved actually experienced a decrease in the concentration of fellow countrymen." This finding, however, provided exactly the opposite evidence to the positive externality effect of cultural clustering to spatial concentration of immigrants. More recent work (Gross and Schmitt 2003, 2006) used the proportion of immigrants to the size of the immigrant population in each country in OECD countries and France, respectively, to measure the ethnic clustering effect. The work also argues that the ethnic clustering effect is more likely to affect low- than high-skilled immigrants.

5.4 Data, Key Variable Distribution, Hypotheses, and Method

5.4.1 Data

The major data set on doctoral scientists and engineers contained in this chapter comes from the 2001 Survey of Doctorate Recipients (SDR), a longitudinal panel survey administered by the National Science Foundation, of all individuals who received their doctorate degrees in science and engineering (S&E) in that year.[9] All the respondents were actively working during the week of the survey. Using a hierarchical random sampling of the data, we obtained a sample with 6,322 observations. Additionally, we employed another data set, with variables on R&D and ethnicity distributions, collected by the U.S. Census Bureau (2000). It is important to understand the distribution of the key variables prior to deriving our hypotheses on the effects of ethnic and technical clustering.

5.4.2 Spatial Distribution of S&E PhDs, Including Citizenship, Ethnicity, and Occupation in the United States, Using the SDR 2001 Sample Data

In this section, we walk readers through the information on spatial distribution of talented human capital (measured by patent applications and the number of granted patents), graduates and their retention rates, the top 10 employer states, and the role of the schools' social capital, by citizenship.

Where do the talented graduates concentrate? To show the geographical distribution of talented human capital, we use the proportion of patents applied for and patents granted in each region by citizenship in Fig. 5.1a and b. When we combine the findings of the two figures, we find two compelling observations: First, the figures show a similar pattern, that is, an uneven distribution of talented human capital among regions. Although the native-born Americans' distribution is similar among regions, both Chinese and Indians are concentrated in the Pacific and Middle Atlantic regions. A closer look at the data shows that several states, such as California in the West region, Indiana and Minnesota in the Midwest region, and New York and New Jersey in the Middle Atlantic region, are dominant over other states, in terms of both the share of patent applications and the share of patents granted.

Our second observation from Fig. 5.1a and b is the likely impact of (both working and permanent) visa status on the highly uneven spatial distribution of patents held by foreign citizens. First, to get a job and then a green card, these graduates have to work very hard to find employment, and subsequently, they have to prove their skills to their employers if they are to be able to apply for a green card. One way of doing this is to apply for research opportunities to show their talent by patent invention.[10]

---

[9] For details of the survey, please see National Science Foundation, Division of Science Resources Statistics (2001 and 2003).

[10] See Freeman et al. (2001) for the career path of foreign students in biosciences.

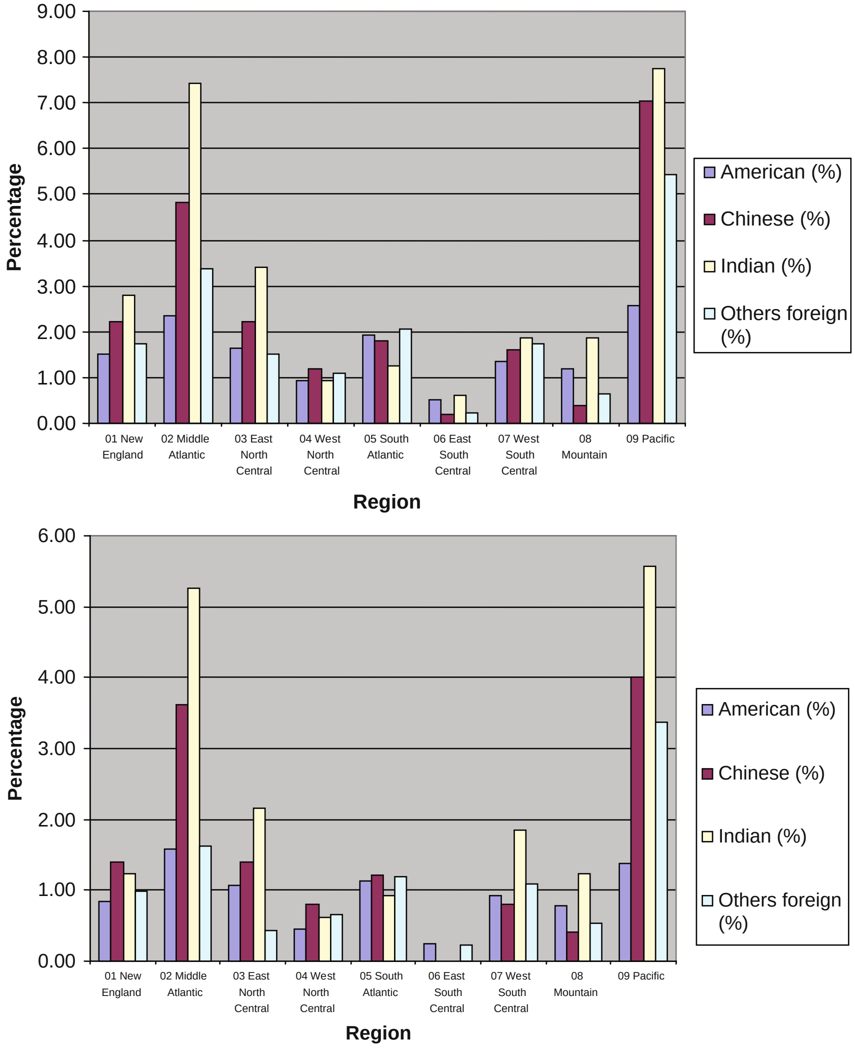

Figure 5.1 Share of patents for each citizenship group. (**a**) Patents applied for, by each citizenship group (the percentage is calculated from the number of patents applied by each citizenship at regions to the overall number of patents applied in the United States). (**b**) Patents granted, by each citizenship group (the percentage is calculated from the number of patents granted to each citizenship at regions to the overall number of patents granted in the United States). All the bars sum up to 100%.
Source: Hierarchical random sample of SDR 2001 data (NSF 2001).

Hence, they could use their patented inventions as proof of their ability for purposes of applying for a green card. The express tracks for green card applications are called "national interest" and "outstanding researcher." These two figures raise questions about the factors affecting the concentration of patents and retention of talented human capital. Tables 5.1 and 5.2 show the distribution of retention of S&E PhD visa holders by the region in which they received their degree and the region in which they were employed.

The distribution of S&E graduates' retention rate by region and visa status is shown in Table 5.2. In particular, this table shows the spatial distribution of graduates with different visa statuses. Native-born American, permanent visa holder, and temporary visa holder categories are shown in the three sections of the table. The diagonal cells of each table are the region's retention rates, indicating foreign S&E PhD recipients employed in the state in which they received their degree. The first section shows the native-born American graduates' retention rate. The native-born Americans' retention rates are fairly consistent across regions, though not high at around 20–30%. We, however, found that the retention rates of foreign graduates with permanent visas (see second section of the table) were spread more widely than those of their native-born American counterparts, ranging from 23.8% in the East South Central region to 55% in the Pacific region.

Interestingly, the temporary and permanent visa holders exhibit even higher variations than the native-born Americans. An even more remarkable fact is that the permanent visa holders' lowest and highest retention rate regions coincided with those of the temporary visa holders; the lowest retention rate region for permanent visa and temporary visa holders was the East South Central (with 23.8 and 13.6%, respectively), while the highest retention rate region was the Pacific region (with 55.2 and 44%, respectively). Based on the previous comparison of citizenships, we hypothesize that citizenship may be instrumental in the spatial displacement of graduates. One of the reasons behind this type of spatial distribution may be that graduates who are foreign citizens and who would like to stay in the United States may be bound by visa availability and the ability of employers to help them apply for a work visa. In Table 5.3, we analyze both permanent and temporary visa holders together by focusing on their citizenships in a subsequent analysis as both groups were not U.S. citizens at the time of the survey.

Table 5.3 shows a state's capacity to attract graduates from other states. Each cell in the table represents the percentage of employed S&E PhD recipients in the corresponding state in comparison to the total S&E PhD recipients of the same citizenship in the United States. California, New York, Texas, and Massachusetts seem to attract the largest proportion of out-of-state graduates; their absorption rate is large: as shown in second last column in the first section of the table, 25% of the total S&E PhD graduates in the United States go to work in those four states.

More interestingly, in the first row of the last column of the first section of Table 5.3, it is seen that 9.44% of all graduates in the sample settled in California; that state's out-of-state graduate absorption rate is more than double that of the next highest state in the rankings. California also attracts the largest proportion of foreign graduates of all the other states; more than 14% of Chinese along with more than 12% of Indians and other foreign graduates with different citizenships seem to settle in California. This finding is consistent with the general immigration pattern found by Borjas (2005). Using the 1970–2000 integrated Public Use Microdata samples of the U.S. Census, Borjas shows that while New York is still a hot spot of immigration

Table 5.2 Regional distribution of S&E PhD graduates by visa status

S&E PhD graduates' region	Native-born (%) Principal employer region after graduation (%)									
	New England	Middle Atlantic	East North Central	West North Central	South Atlantic	East South Central	West South Central	Mountain	Pacific	Total graduates by region
01 New England	26.3	15.3	9.9	3.5	17.7	1.6	4.0	6.7	15.0	373
02 Middle Atlantic (row %)	11.8	23.0	12.4	4.6	20.1	2.5	5.0	7.0	13.7	483
03 East North Central (row %)	8.6	12.7	21.3	8.3	15.7	6.3	6.2	9.3	11.7	616
04 West North Central (row %)	7.5	9.0	14.5	26.2	13.6	3.3	7.8	8.1	9.9	332
05 South Atlantic (row %)	8.1	12.1	9.8	5.8	27.4	6.2	11.1	7.5	11.9	530
06 East South Central (row %)	3.1	12.4	10.6	5.3	19.9	31.4	8.0	3.5	5.8	226
07 West South Central (row %)	3.3	8.2	11.3	6.4	15.9	6.9	27.8	8.7	11.3	389
08 Mountain (row %)	7.5	8.3	6.4	5.6	10.9	3.7	9.1	29.9	18.7	375
09 Pacific (row %)	10.1	13.1	9.3	6.1	15.4	2.4	6.3	13.1	24.1	572
Total employment % by region (row %)	9.8	13.0	12.1	7.6	17.6	5.8	9.2	10.6	14.2	
Total employment by region	382	506	473	297	686	227	358	412	555	3,896

Table 5.2 (continued)

S&E PhD graduates' region	New England	Middle Atlantic	East North Central	West North Central	South Atlantic	East South Central	West South Central	Mountain	Pacific	Total graduates by region
					Permanent visa (%)					
01 New England (row %)	37.4	22.4	6.5	0.9	11.2	0.0	0.9	1.9	18.7	107
02 Middle Atlantic (row %)	10.9	33.2	9.5	3.3	16.1	0.5	5.7	2.8	18.0	211
03 East North Central (row %)	9.0	17.5	24.7	6.3	14.3	2.2	5.8	4.0	16.1	223
04 West North Central (row %)	10.3	15.4	10.3	25.6	5.1	3.8	6.4	3.8	19.2	78
05 South Atlantic (row %)	9.4	14.1	7.6	2.4	38.2	2.4	10.0	2.4	13.5	170
06 East South Central (row %)	2.4	11.9	16.7	2.4	21.4	23.8	4.8	4.8	11.9	42
07 West South Central (row %)	3.6	10.7	10.7	4.5	10.7	3.6	35.7	3.6	17.0	112
08 Mountain (row %)	5.1	8.9	3.8	1.3	7.6	1.3	15.2	26.6	30.4	79
09 Pacific (row %)	3.2	15.2	8.0	0.8	8.8	1.6	4.8	2.4	55.2	125
Total employment % by region (row %)	10.5	18.5	11.8	4.7	16.1	2.6	9.4	4.7	21.7	
Total employment by region	120	212	135	54	185	30	108	54	249	1147

Table 5.2 (continued)

S&E PhD graduates' region	New England	Middle Atlantic	East North Central	West North Central	South Atlantic	East South Central	West South Central	Mountain	Pacific	Total graduates by region
				Temporary visa (%)						
01 New England (row %)	30.2	27.0	6.3	3.2	7.9	1.6	1.6	1.6	20.6	63
02 Middle Atlantic (row %)	14.4	33.7	9.6	3.8	10.6	1.9	6.7	1.9	17.3	104
03 East North Central (row %)	6.0	16.4	31.9	2.6	6.9	2.6	9.5	2.6	21.6	116
04 West North Central (row %)	8.5	14.9	10.6	31.9	4.3	2.1	10.6	2.1	14.9	47
05 South Atlantic (row %)	6.9	17.2	10.3	2.3	34.5	0.0	11.5	1.1	16.1	87
06 East South Central (row %)	4.5	9.1	9.1	9.1	31.8	13.6	9.1	4.5	9.1	22
07 West South Central (row %)	6.7	4.4	11.1	0.0	11.1	2.2	48.9	2.2	13.3	45
08 Mountain (row %)	0.0	11.4	8.6	11.4	2.9	5.7	5.7	40.0	14.3	35
09 Pacific (row %)	5.3	13.3	5.3	5.3	16.0	2.7	6.7	1.3	44.0	75
Total employment % by region (row %)	9.9	18.7	13.3	6.1	13.6	2.5	10.9	4.2	20.7	
Total employment by region	59	111	79	36	81	15	65	25	123	594

Note: Native-born retention is pretty consistent at around 20–30%, while permanent visa foreign distribution is more volatile than native-born retention's distribution with the in-state percentage ranging from 24.7 to 55%. The highest volatility is temporary visa holders, due in part to their visa nature that foreigners may go to work in the place where they received their visa and then change jobs when they obtain a permanent visa. The high variation ranges from only 13.6% in the East South Central to 44% in the Pacific region.

Source: National Science Foundation, Division of Science Resources Statistics 2001 Survey of doctorate recipients in the United States.

Table 5.3 Spatial distribution of graduates by employment state for each citizenship in the United States by in or out of state

Top 10 in state	Native-born American (%)	Native-born American Top 10 states	Chinese (%)	Chinese Top 10 states	Indian (%)	Indian Top 10 states	Other foreign (%)	Other foreign Top 10 states	In-state total (%)	In-state total Top 10 states
	The share of employees to total graduates in each group from in state									
1	2.8	California	2.6	California	3.4	California	5.3	California	3.1	California
2	1.9	Texas	2.4	Texas	3.4	Texas	2.6	Massachusetts	2.1	Texas
3	1.6	Massachusetts	2.2	New York	1.5	Massachusetts	2.5	Texas	1.7	Massachusetts
4	1.3	New York	1.6	Pennsylvania	1.5	New Jersey	2.1	New York	1.5	New York
5	0.8	Colorado	1.4	Illinois	1.2	New York	1.5	Illinois	1	American Samoa
6	0.7	Illinois	1.4	Ohio	0.9	Ohio	1.2	Pennsylvania	1	U.S. Virgin Islands
7	0.6	Minnesota	1.2	North Carolina	0.6	Arizona	1	New Jersey	0.8	Illinois
8	0.6	Ohio	1	Colorado	0.6	Maryland	1	Ohio	0.8	Pennsylvania
9	0.6	Pennsylvania	0.8	Alabama	0.6	Michigan	0.9	Arizona	0.7	Colorado
10	0.5	Alabama	0.8	Arizona	0.6	Minnesota	0.9	Maryland	0.7	Ohio
%In state	18.74		22.89		19.50		27.66		20.4	

Table 5.3 (continued)

Top 10 out of state	The share of employees to total graduates in each group from out of state									
	Native-born American		Chinese		Indian		Other foreign		Out-of-state total	
	(%)	Top 10 states	(%)	Top 10 states	(%)	Top 10 states	(%)	Top 10 states	(%)	Top 10 states
1	8	California	14.7	California	12.7	California	12.8	California	9.4	California
2	4.8	New York	7.4	Massachusetts	7.4	New Jersey	6.1	New York	5.1	New York
3	4.7	Texas	6.4	Texas	7.1	Texas	5.5	Texas	5	Texas
4	4.1	Massachusetts	5	New Jersey	6.8	New York	5.2	Massachusetts	4.7	Massachusetts
5	3.8	Pennsylvania	4.6	New York	6.5	Massachusetts	4.6	New Jersey	3.7	Pennsylvania
6	3.6	Maryland	3.8	Pennsylvania	3.7	Pennsylvania	3.1	Illinois	3.4	New Jersey
7	2.8	Illinois	3.2	Illinois	3.4	Illinois	3	Pennsylvania	3.3	Maryland
8	2.7	New Jersey	3.2	Maryland	3.1	Washington	2.4	Maryland	2.9	Illinois
9	2.6	Colorado	2.4	Georgia	2.5	Maryland	2.3	Washington	2.3	Ohio
10	2.5	Ohio	2.4	Michigan	2.5	Michigan	2.1	Florida	n	Michigan
%Out of state	81.26		77.11		80.5		72.34		79.60	
Total by citizenship	4579		498		323		922		6322	

Note: The percentages in the cells are from the total number of employed in each citizenship in the whole United States. For example, if 13 Chinese graduates in our sample stay in California after graduation, we use the number of students in state/total number of Chinese graduated = (%) of total graduates. That is, 13/498 = 2.6%. This table shows which state has gained the most graduates from the point of view of the whole United States.

Source: National Science Foundation, Division of Science Resources Statistics 2001 Survey of doctorate recipients in the United States.

(with 15.7% of all foreign-born workers living in the New York metropolitan area), California, Texas, and Florida are centers that have drawn them in recent decades (see also Saxenian 1994, page 163). The immediate question is this: In which industries is the talented human capital concentrated?

Table 5.4 shows the distribution of graduates by citizenship and by industry. The two subsections analyze the data in two different subsets: in state and out of state. The distribution of employment among citizenships is found to be quite consistent when comparing the in-state and out-of-state employment pattern. The largest employers are in the fields of education, information technology, manufacturing, and research; these four industries constitute over 70% of total employment, and it holds true across different citizenships. In fact, the distribution of employment is quite similar between in state and out of state, with the out-of-state employment around three times higher than the in-state employment. This consistent pattern of employment may lead to a hypothesis that industrial distribution matters for the graduates' mobility and that if the in-state industries offer a job opportunity, a local educational institution's graduate is more likely to accept, holding all other factors constant.

Furthermore, when we observe the table across different citizenships, we find that foreign graduates exhibit a similar pattern while native-born Americans' employment pattern is different in two aspects. Comparing across citizenships, native-born American graduates are more likely to seek employment in health services and public administration/government, while foreign graduates are more likely to seek employment in information technology and manufacturing. This tendency is reasonable because certain professions, such as public administration/government, require American citizenship. We observe that Chinese and Indian graduates are more concentrated in biotechnology and information technology. Both the Chinese and Indian employment ratios in the biotechnology field are about twice as high and in the information technology field three times as high as those of the native-born American citizens in these industries. These differences in the distribution of different professions may suggest another hypothesis, namely, that citizenship may matter in occupation selection and thus in the opportunity to have an in-state occupation.

Another factor that may affect the spatial distribution of human capital is the school's social capital for job hunting. Table 5.5 shows the distribution, by citizenship, of job-hunting practice and its relationship with a school's social capital. When we look across the columns of Table 5.5, we see that foreign graduates chose "Faculty or advisor" as their most important source of finding their first career path job, while native-born Americans chose "Informal channels through colleagues or friends" as their most important source of finding their first career path job. This sharp contrast shows in part how graduates rely on resources for their job hunting. It also shows that job-hunting resources are highly related to the graduates' access to social capital. Another interesting observation is that for Indian graduates, the percentage choosing "Direct contacts with company" is three times higher than that of native-born American citizens and Chinese. For a more rigorous analysis of the relationship of these factors to the spatial concentration of highly skilled human capital, we derive these hypotheses:

Table 5.4 Distribution of employment from in state and out of state

Industry	Native-born Americans	Chinese	Indians	Other foreign	In-state total
	In-state % of total graduates in the United States				
01 Agriculture, forestry, or fishing	0.13	0.00	0.00	0.00	0.09
02 Biotechnology	0.33	1.41	0.93	0.65	0.49
03 Construction or mining	0.15	0.20	0.31	0.00	0.14
04 Education	6.60	5.42	4.64	11.93	7.18
05 Finance, insurance, or real estate services	0.26	0.60	0.00	0.98	0.38
06 Health services	2.42	1.61	0.93	1.19	2.10
07 Information technology	0.94	4.02	5.26	2.93	1.69
08 All other services (e.g., social, legal, business)	0.63	0.40	0.00	0.65	0.59
09 Manufacturing	0.96	2.61	2.17	1.84	1.28
10 Public administration/government	1.07	0.40	0.31	0.65	0.92
11 Research	4.00	4.62	3.72	5.42	4.24
12 Transportation services, utilities, or communications	0.22	0.20	0.31	0.33	0.24
13 Wholesale or retail trade	0.11	0.00	0.31	0.22	0.13
14 Other	0.92	1.41	0.62	0.87	0.93
%In state	18.74	22.89	19.50	27.66	20.40
	Out of state % of total graduates in the United States				
01 Agriculture, forestry, or fishing	1.07	1.61	0.00	0.22	0.93
02 Biotechnology	2.53	6.63	5.57	2.82	3.05
03 Construction or mining	0.20	0.60	1.24	0.33	0.30
04 Education	33.92	15.86	19.50	28.85	31.02
05 Finance, insurance, or real estate services	1.11	3.41	2.17	1.63	1.42
06 Health services	7.14	5.42	3.41	3.69	6.31
07 Information technology	4.67	15.66	14.86	7.81	6.52
08 All other services (e.g., social, legal, business)	2.03	0.40	0.62	0.33	1.58
09 Manufacturing	4.94	9.44	7.12	6.40	5.62
10 Public administration/government	3.10	0.20	0.31	0.43	2.34
11 Research	16.14	13.65	19.20	14.43	15.85
12 Transportation services, utilities, or communications	0.85	2.01	0.93	1.41	1.03
13 Wholesale or retail trade	0.46	0.40	0.31	0.22	0.41
14 Other	3.10	1.81	5.26	3.80	3.21
%Out of state	81.26	77.11	80.50	72.34	79.60
Total of graduates in the sample by citizenship in 2001	4579	498	323	922	6322

Source: National Science Foundation, Division of Science Resources Statistics 2001 Survey of doctorate recipients in the United States.

Table 5.5 Which resource is the most responsible for finding your first career path job? (%)

Answer	Native-born American	Chinese	Indian	Other foreign	Total
01 Faculty or advisors	24.2	27.1	31.1	30.6	25.9
02 Professional recruiters	1.6	4.5	1.6	0.4	1.7
03 College or department placement office	3.8	6.8	3.3	2.6	3.9
04 Professional meetings	5.8	4.5	8.2	5.2	5.7
05 Electronic postings	15.8	16.5	13.1	18.5	16.2
06 Newspapers	1.5	3.8	0.0	1.3	1.6
07 Professional journals	8.9	7.5	8.2	5.6	8.2
08 Informal channels through colleagues or friends	25.9	19.5	13.1	21.6	24.0
09 Direct contacts with company	7.8	8.3	21.3	14.2	9.5
10 Other	4.7	1.5	0.0	0.0	3.4
	100.0	100.0	100.0	100.0	100.0
Total no.	962	133	61	232	1388

Note: Answers are valid only from 1998 on respondents from the hierarchical random sampling of the SDR 2001 data (NSF 2001).
Source: National Science Foundation, Division of Science Resources Statistics 2001 Survey of doctorate recipients in the United States.

5.4.3 Hypotheses

Drawn from the preceding discussions, the spatial assimilation argument can be deduced to the following hypotheses:

Hypothesis 1: Citizenship affects spatial assimilation, and the spatial distribution of PhD graduates.

Hypothesis 2: Ethnic clustering, represented by the proportional differences of ethnicity in a state in comparison to the national average, is likely to be very different across different citizenships. In particular, different citizenships may respond to ethnic clustering effect quite differently.

Hypothesis 3: The spillover effects of technical clustering in relation to the amount of industry and R&D in the state may affect different citizenship groups differently.

Hypothesis 4: The matching effects of technical clustering in relation to graduates' field of specialization and the industry in which they are employed, as well as the amount of R&D funding in the state, may affect different citizenship groups differently.[11]

Hypothesis 5: Professional specializations also play an important role in the in-state retention rate; however, the effect will differ in magnitude if citizenship is introduced as a variable here.

Hypothesis 6: The way in which school social capital affects labor mobility is likely to differ between the native-born American and foreign graduates. School

[11] We will use the pooled data regression to test this hypothesis due only to the limitations in accessing data.

social capital may be very important for the first job location; however, graduates of different citizenships may perceive and utilize school social capital very differently.

5.4.4 Method: Estimation Strategy and Variables

5.4.4.1 Estimation Strategy

The method in this chapter is simple: first, we construct a logistic discrete model, and then, we run and discuss the two sets of logistic regression results. This model is a preliminary step to investigation of the spatial assimilation of highly skilled native-born American and foreign graduates (H-1B visa holders).

In these models, we assume that graduates' employment location choice depends on the industry in which they have applied for employment and accepted an offer. We also assume that wage is competitive all over the United States and thus that wage does not play a role in job location selection. In this regard, other than personal and family factors, we are interested in the way technical clustering and ethnic clustering affect the location choice.[12] Currently, we focus on the ex post outcome where the migration decision has been made and the migrant continues to stay in the state. In other words, we focus our interest on the effects of the employers' states on the retention rate of graduates in state.

The estimation equation is constructed as follows:[13]

$$y^* = \beta_0 + X\beta + u \tag{5.1}$$

The in-state decision, y^*, is the likelihood of a person staying in the same state after graduating from PhD. y^* is dependent on a vector of independent variables, $\boldsymbol{X}$, where $\boldsymbol{X}$ is a set of explanatory variables and a *row* vector containing elements ($x_1, \ldots, x_k$), and β is a *column* vector containing corresponding coefficients: $\beta' = (\beta_1, \beta_2, \ldots, \beta_k)$. The explanatory variables include the following: personal characteristic and creative ability factors, family factors, field specialization and the characteristics of the educational institution, the employer's field and the state's R&D, personal aspirations on the job, social capital for a job search and post doc placement, technical clustering, ethnic clustering, and other control variables. The term u is the random error term with a normalized mean (0) and variance (1).

We first estimate a logistic model on each group, native-born Americans, Chinese citizens, Indian citizens, and other foreign citizens, separately to show the contrast of the different effects of spatial concentration. The advantage of doing this is that if the technical clustering's spillover effect and ethnicity clustering variables do not

---

[12] We acknowledge that a gravity model with both pull and push efforts on both origins and destinations may be even better at capturing the forces which determine migration. This is the next step for our research agenda.

[13] The description here is a shorter version of similar equations in Chapter 4.

matter across different citizenships, this independent variable should show similar magnitude, sign, and significance level after other independent variables, such as individual, family, school, and personal preferences, are controlled for. Next, we pool the data set and run an extended regression with the full set of technical clustering variables, which includes the spillover and matching effects, and with variable interactions on the foreign dummy variable. For the pooled model, we expect to see a significant effect on the "citizenship dummy" if assimilation has an effect on the location decision.

5.4.4.2 The Dependent Variable

The dependent variable "in state" takes the value of "0" when the respondents indicate that their state of employment is different from the one where their first S&E PhD was awarded and "1" when they indicate that the two states are the same. This dependent variable simplifies the multiple destination choice by reducing the location choice to a dummy variable. We also control the regional variable by the region of employers as well as the first S&E PhD-granting region.

5.4.4.3 Independent Variables

We include all the independent variables we employed in Chapter 4 along with the following:

For the estimation of *technical clustering* effects, we extend Stephan et al.'s (2004) approach to dividing the data from the state's 1998 R&D funding into two sets of interaction terms in the following regression models.[14]

The R&D spillover effect on labor mobility consists of three subset of interaction terms. The first subset of interaction terms of R&D and field of specialization shows the attractiveness of R&D to a particular type of specialization. The second subset of variables that measure the technical clustering effects is the R&D and industrial sector's spillover effect on labor mobility because of a particular type of industry and its agglomeration effect. For simplicity, we collapse the original industrial major field to three categories and use this new industrial field with R&D. The first industrial category is biotechnology, information technology, and research work; the second category is education; and all other industries are included in the reference category. The third subset of interaction terms of R&D and citizenship signifies each citizen's inclination, across different citizenships, to stay in state.

[14] However, there are problems with data availability and measurement in R&D, as suggested by Stephan et al. (2004), namely that complete sets of R&D data are hardly available at the state level and that R&D funding may be highly concentrated in some institutions in the locality but not at the MSA or the state level. This may suggest why some measure of R&D expenditure may not be able to explain the agglomeration effects. In addition to the available investment (R&D) capital, another closely related issue is the way in which talented human capital can be attracted to localities and participate in the spillover effects and technological agglomerations. In this regard, the effect of R&D is still a subject for further study in S&E labor migration literature, at least in the context of the United States.

The other set of variables that measure the technical clustering's *matching effects* includes the interaction terms of field specialization and employers' industry. First, we use this additional set of variables because these interaction terms can signify matching effect, that is, the effects of matching between major S&E fields to employers' industries and to the likelihood of S&E PhD graduates staying in state. Second, technical clustering should be the effect between matching skills and industry and should not be affected by citizenship.

For *ethnic clustering*, we use the demographic data from the U.S. Census Bureau (2000). We use the variables "concentration of ethnicity in each state" by calculating the difference of Asian, black, and Hispanic populations from the national average to capture the ethnicity clustering in the employers' state. This is a step ahead of the usual percentage of national migrant stock (see Gross and Schmitt 2003 and 2006).

The formation of the variable is as follows: ethnicity clustering of each ethnicity at state i = (% of an ethnicity e in state i – national average % of ethnicity e) / (national average % of ethnicity e).

The advantage of this specification is that it can provide a more rigorous index measure on ethnicity, thus reducing the effect of the size of ethnicity in each state. It also provides more precise variations on each ethnicity since all the means of these ethnicities in the United States are not the same.

For citizenship variables, we use Chinese, Indian, and other foreign students to compare with the native-born Americans (controlling their ethnic groups) in terms of spatial distribution. We will use some interaction terms with the citizenship dummy variables to signify the difference among groups under comparison for the extended model.

5.4.4.4 Other Control Variables

School characteristics such as professional field specialization and school-related variables certainly are important factors in the choice of mobility. We also use the two "home-state effect" dummy variables to capture the effect of the degree-granting school's geography and the location choice. The other question is the social capital available for job searching.

We continue to use variables such as family characteristics, cohort effect, personal character and family background, gender, ethnicity, marital status, number of children in different age groups (up to age 18), as well as spouse's working status. In addition to individual characteristics, individual ability variables, including personal creative activities such as books accepted for publication, papers presented at conferences, patents applied for, patents granted, and patents commercialized, are used.

5.5 Results

In Table 5.6, which estimates foreign graduates separately, we use only the state R&D variables and their interaction terms with variables in industrial sectors to signify one set of *basic* technical spillover effects of technical clustering, due to the limitation of observations. In Table 5.7, the full sample model, we can then introduce a full set of technical clustering's spillover and matching variables since we pool the individual citizenship sample in one model.

Table 5.6 shows that the effects of independent variables are quite different among citizenships.

The first set of regression results, on *ethnic clustering*, shows a striking pattern that has not been discussed previously. We have two interesting observations: First, the model with the native-born Americans shows a strong negative effect at locations where Hispanics are concentrated. This suggests that the states with high Hispanic concentrations are likely to attract graduates from out of state. These Hispanic-concentrated states include Texas, New Mexico, California, and Arizona, with 32, 42.1, 32.4, and 25.3%, respectively. We also find that the coefficient in the Chinese model shows an expected significant (positive) sign on the Asian dummy variable, after we control the regional dummy variables for the employers' and degree-granting regions. From our data, we find that states with the highest Asian concentration are California, Washington, New York, and New Jersey, with 10.9, 5.5, 5.5, and 5.7%, respectively. These are popular destinations for Chinese migrants. One may suspect that the cumulative causation argument may be at work in these regions (see details of the argument in Massey et al. 1993). The second observation is that Indians are not at all responsive to these variables; the other foreign model shows a combined effect of native-born Americans and Chinese, where the variable for Hispanics is negative and the variable for Asians is positive. This may be due to the construction of "Other foreign" to include graduates from other Asian countries and other European countries, who may respond similarly to Chinese and Americans.

The second set of the regression results in Table 5.6 shows that the diverse effects of *field of specialization* increase the likelihood of staying in state among different citizenships. For native-born Americans, the biological and agricultural sciences and the social sciences have a strongly negative effect when compared to the reference, which is computer and information science, meaning that native-born Americans in these fields are less likely to stay than those in the other groups. The effect of field of specialization in the Chinese model, however, does not show any difference among different variables. While only the coefficient of the health sciences has a negative effect, the model of other foreign PhDs shows only psychology majors as being likely to stay in state when compared to the reference (computer and information science majors). What other variables explain the difference in likelihood of staying in state among different citizenships? We will now further our investigation into the effects of technical clustering.

To test the hypothesis of spillover effects on *basic technical clustering*, we use the industry dummies, interact them with R&D funding, and show the results in Table 5.6. Terms that interact with R&D are biotechnology, information technology, and research work. The second interaction term is education, while all other industries are in the reference category. Our intuition is that, by definition, technical clustering should have similar effects on all graduates with different citizenships. However, we show some interesting differences among citizenships.

Table 5.6 Basic logistic regression by citizenship on ethnic clustering and technical clustering

	Native-born American		Chinese		Indian		Other foreign	
Variable	*dy/dx*	*SE*	*dy/dx*	*SE*	*dy/dx*	*SE*	*dy/dx*	*SE*
Ethnicity clustering								
Hispanic density	–0.05****	(0.01)	0.04	(0.08)	–0.08	(0.06)	–0.11**	(0.05)
Black density	–0.01	(0.01)	–0.02	(0.06)	–0.09	(0.07)	0.01	(0.04)
Asian density	0.01	(0.01)	0.08***	(0.03)	0.02	(0.02)	0.04***	(0.02)
Field specialization effect								
Major field for first S&E PhD (major group) (reference = computer and information science)								
Mathematical sciences = 1	–0.04	(0.03)	0.21	(0.22)	0.03	(0.19)	0.05	(0.10)
Biological and agricultural sciences = 1	–0.07***	(0.02)	0.19	(0.16)	–0.06	(0.08)	0.04	(0.08)
Health sciences = 1	–0.02	(0.03)	0.17	(0.25)	–0.17***	(0.03)	0.14	(0.12)
Physical and related sciences = 1	–0.04*	(0.02)	0.12	(0.16)	–0.06	(0.07)	0.06	(0.08)
Social sciences = 1	–0.06***	(0.02)	0.24	(0.22)	–0.04	(0.09)	0.02	(0.08)
Psychology = 1	–0.04*	(0.03)	0.23	(0.36)			0.24*	(0.13)
Engineering = 1	–0.02	(0.03)	0.13	(0.16)	–0.02	(0.08)	0.04	(0.07)
Spillover effects of technical clustering								
State R&D expenditure in 1998	0.00	(0.00)	–0.00**	(0.00)	0.00	(0.00)	0.00	(0.00)
(Other industries as reference)								
(Biotechnology, information technology) =1	–0.02	(0.02)	0.10	(0.07)	0.00	(0.08)	–0.09	(0.06)
(Educational institution) = 1	–0.07***	(0.02)	–0.01	(0.07)	–0.10	(0.07)	–0.05	(0.06)
R&D × biotechnology, information technology	0.00	(0.00)	–0.00**	(0.00)	0.00	(0.00)	0.00**	(0.00)
R&D × educational institution	0.00**	(0.00)	0.00	(0.00)	0.00**	(0.00)	0.00**	(0.00)
Social capital for job hunting and post doc effects (academic advisor as reference)								
Formal institutions = 1	–0.06*	(0.03)	–0.07	(0.07)	–	–	–0.21***	(0.04)
Media = 1	–0.08***	(0.02)	–0.06	(0.07)	0.01	(0.17)	–0.18***	(0.04)
Informal channels = 1	0.01	(0.03)	–0.11**	(0.05)	–	–	–0.03	(0.07)
Direct contacts = 1	–0.02	(0.03)	0.10	(0.15)	0.04	(0.17)	–0.08	(0.07)

Table 5.6 (continued)

Variable	Native-born American		Chinese		Indian		Other foreign	
	dy/dx	*SE*	*dy/dx*	*SE*	*dy/dx*	*SE*	*dy/dx*	*SE*
Constant	−1.21	−(0.78)	4.38	−(0.03)	0.15	−(3.64)	−1.01	−(1.88)
No. of observations	3896		496		286		920	
Wald χ^2	258.50		96.44		67.05		109,50	
Prob > χ^2	0.00		0.00		0.15		0.00	
Pseudo R^2	0.12		0.17		0.23		0.11	
Predicted probability	0.15		0.17		0.14		0.25	

Note: All regressions are weighed by the cohort in the sample.

Robust standard errors are within parentheses in the column next to the marginal effects of logistic regressions.

One industrial dummy and two schools' social capital dummies for Indian graduates are not shown due to inadequacy of the observations on control variables: income per capita in employer's state, population in 2002, S&E PhD graduate's region, and employer's region. Family structure, spouse working, book published, paper presented, patent applied for, patent granted, patent commercialized, post doc and home-state effect (for native-born Americans' and other foreigners' regressions only).

***,**,*Significantly different from zero at the 1, 5, and 10% levels, respectively.

Table 5.7 Extended regression for ethnic clustering and technical clustering for full sample

	Model 1		Model 2	
Variable	*dy/dx*	*SE*	*dy/dx*	*SE*
Citizenship				
(Reference = native-born American)				
Chinese	0.01***	(0.03)	0.21***	(0.06)
Indian	0.08***	(0.03)	0.08	(0.07)
Other foreign	0.16***	(0.02)	0.17***	(0.04)
Ethnic clustering				
(Reference = White)				
Hispanic density	−0.06***	(0.01)	−0.06***	(0.01)
X Chinese			0.02	(0.02)
X Indian			−0.01	(0.03)
X Other foreign			−0.01	(0.02)
Black density	−0.01	(0.01)	−0.01	(0.01)
X Chinese			0.00	(0.02)
X Indian			0.01	(0.03)
X Other foreign			−0.01	(0.02)
Asian density	0.02***	(0.01)	0.01	(0.01)
X Chinese			0.04**	(0.02)
X Indian			−0.01	(0.03)
X Other foreign			0.02*	(0.01)
Field of specialization				
(Reference = computer and information science)				
Mathematical sciences = 1	−0.02	(0.03)	−0.11**	(0.05)
Biological and agricultural sciences = 1	−0.04*	(0.02)	−0.07	(0.05)
Health sciences = 1	0.00	(0.03)	−0.07	(0.09)
Physical and related sciences = 1	−0.02	(0.02)	−0.06	(0.06)
Social sciences = 1	−0.04*	(0.02)	−0.04	(0.06)
Psychology = 1	−0.01	(0.03)	−0.03	(0.06)
Engineering = 1	−0.01	(0.02)	−0.05	(0.06)
Technical clustering				
Employer industries and their matching effects with major fields (other industries as reference)				
Biotechnology, information technology, and research) = 1	−0.01	(0.02)	0.00	(0.07)
X Mathematical sciences			0.17	(0.16)
X Biological and agricultural sciences			0.01	(0.07)
X Health sciences			−0.10***	(0.04)
X Psychology and related sciences			−0.02	(0.07)
X Social sciences			−0.01	(0.08)
X Psychology			−0.02	(0.08)
X Engineering			0.06	(0.08)
(Educational institution) = 1	−0.06***	(0.02)	−0.03	(0.07)
X Mathematical sciences			0.08	(0.13)
X Biological and agricultural sciences			0.00	(0.07)
X Health sciences			−0.11***	(0.04)
X Psychology and related sciences			0.02	(0.08)
X Social sciences			−0.07	(0.05)
X Psychology			−0.08	(0.05)
X Engineering			0.04	(0.08)

Table 5.7 (continued)

	Model 1		Model 2	
Variable	*dy/dx*	*SE*	*dy/dx*	*SE*
R&D and its spillover effects with major fields and citizenships (reference = computer and information science)				
State R&D, expenditure in 1998	0.00	(0.00)	0.00	(0.00)
X Mathematical sciences = 1			0.00*	(0.00)
X Biological and agricultural sciences = 1			0.00	(0.00)
X Health sciences = 1			0.00	(0.00)
X Physical and related sciences = 1			0.00*	(0.00)
X Social sciences = 1			0.00**	(0.00)
X Psychology = 1			0.00**	(0.00)
X Engineering = 1			0.00	(0.00)
R&D's spillover effects with employer industries (other industries as reference)				
X (Biotechnology, information technology, and research) = 1	0.00	(0.00)	0.00*	(0.00)
X (Educational institution)	0.00***	(0.00)	0.00***	(0.00)
R&D's spillover effects with citizenship (native American as reference)				
X Chinese			– 0.00***	(0.00)
X Indian			0.00	(0.00)
X Other foreign			– 0.00*	(0.00)
Cons				
No. of obs	5637		5637	
Wald χ^2	458.30		524.15	
Prob > χ^2	0.00		0.00	
Pseudo R^2	0.09		0.11	
Predicted probability	0.18		0.17	

Notes:
Model 1 is the baseline pool data mode. Model 2 includes interaction terms from S&E major fields and citizenship interactions.
The letter *X* before a variable means that it is the interaction term of the variable with the lead variable in each section.
All regressions are corrected for survey design effects and have been weighed by the cohort in the sample.
Robust standard errors are enclosed in parentheses in the column next to the marginal effects of logistic regressions.
The expression *dy*/*dx* indicates discrete change of the dummy variable from 0 to 1.
The technical clusterings' "spillover effects" represent the effects to labor mobility of R&D and/or industrial attraction with an employee's major field and citizenship.
The term "matching effects" represents the effects to labor mobility of industrial match with an employee's major field.
Control variables: gender, number of children, spouse work, year graduate cohort, home-state effects, personal creativities, income per capita, population size in 2002, school social capital, employer's region, S&E graduation region, post doc index.
At the extended model, we also control foreign and its interaction with post doc and industries.

Native-born Americans show a strong negative effect (a 7% reduction in in-state likelihood) in response to the education profession as compared to the reference (other industries), interaction between education and R&D funding, which shows a fairly strong positive effect. This means that the strong R&D capacity of an educational institution may encourage more native-born Americans to stay; however, a strong educational institution effect alone would generally lead them out of the state where they received their first S&E PhD. While Indians and other foreign graduates show a similarly small (close to zero) positive response to the interaction terms between education and R&D funding, other foreign graduates also show a small positive effect to interaction terms between the biotechnology, information technology, and research work category and R&D funding. One of the possible explanations is that Indians and some other foreign graduates generally speak English at home before they come to the United States and that they therefore assimilate more quickly than other Asians, for example, Chinese.

However, Chinese are likely to move away from the state in which they received their first S&E PhD when the R&D funding is higher and its interaction term with the biotechnology, information technology, and research work category is higher, while all other foreign graduates' models show very different, diverse patterns. The interpretation may be that Chinese graduates all over the United States might, in fact, be seeking opportunities for gaining work visas, and they might be good at comparing research industries to other industries for R&D funding. If this interpretation is likely to be the case, we would observe its negative significance on the interaction term of the Chinese dummy and R&D in the extended model, using the pooled data, when compared with the native-born Americans (which is indeed the case and is shown in the extended model that follows).

The current analysis of the idea of technical clustering in Table 5.6 suggests that there are tremendous differences among people of different citizenships when responding to technical clustering. These results cast doubts on the traditional idea of technical clustering, in which labor is considered to respond homogeneously to the generation of agglomeration effects by local industries. We, however, further the inquiry into the spillover effects of technical clustering by introducing an additional interaction term, graduates' field of specialization and employers' industries, in the full sample model, which features homogeneous matching between fields and employers.

In Chapter 4, to signify the effect of *school social capital* on the likelihood of graduates' staying in state, we use "advisors" as the reference for comparison of the different effects of other job-hunting options that graduates may choose. This is because academic advisors usually have more connections (social capital) for local jobs than for those out of state. When comparing the effect of school social capital, we found that native-born Americans who selected media rather than their advisor as their source of job hunting showed a strong negative likelihood (8% less likely) of staying in state. Chinese who elected informal channels were less likely (we observed an 11% reduction) to stay in state, signifying that their social networks are not in that state. Similarly, other foreign graduates who selected formal institutions and media as their most important job hunting sources were also less likely (with 21 and 18% reductions, respectively) to stay in state. This confirms our hypothesis that graduates of different citizenships may choose different media types for job hunting, depending on their background and social networks. This also leads us to further investigate the possibility that citizenship and ethnicity may play an important role in the extended model using pooled data of the native-born Americans and foreigners.

Table 5.7, using the full sample, focuses on the extended effects of technical clustering and ethnic clustering on different citizenships regarding their likelihood of staying in state. Model 1 shows the basic model without foreign interactions, while Model 2 includes interaction terms of foreign citizenship and graduates' field of specialization as well as interaction terms with employers' industries and with R&D funding for testing the analysis of the "technical clustering's matching effect." Model 1, the basic pooled data regression, shows strong significant (at the 1% level) positive effects on all the foreign dummies for native-born Americans. That means, in general, Chinese, Indian, and other foreign graduates are more likely (10, 8, and 16% higher, respectively) to stay in state than are native-born American citizens.

However, the coefficients of the citizenship variables are different when their significance levels and magnitudes are compared in Models 1 and 2, where the extended pool sample includes the foreign citizenship interaction terms with other independent variables. The first observation in Model 2 is that the coefficients of Indian graduates are not as significant as they are in Model 1, but that the coefficients of Chinese and other foreign citizenships' dummies have increased their influence on the in-state likelihood. This suggests that the introduction of foreign interaction terms has absorbed the effects of Indian citizenship on the in-state decision, but not on Chinese and/or the other foreign graduates' citizenships. In other words, the inclusion of interaction terms may help explain the dynamics of citizenships and the in-state decision. In fact, throughout Model 2, only Chinese graduates' interaction terms show a significant difference from the native-born Americans'. For example, Chinese are less likely to be from in state (with a small coefficient) if the R&D is high. In this regard, homogeneous movement due to skill agglomeration, underlined by technical clustering, may be challenged by different ethnicity preferences, that is, ethnic clustering.

When investigating the variables representing ethnic clustering, we find perhaps one of the most important results in this pooled data model. Model 1 shows a strong negative effect on Hispanic concentration and a strong positive effect on Asian concentration to the likelihood to stay in state in general. However, when we add the foreign interaction terms, the significance of the Asian dummy is reduced because Chinese foreign graduates are also inclined to go to the Asian community, while Indian and other foreign citizens do not show such effects. The general finding is that we extend the idea of ethnic clustering in a sense that we not only investigate situations where the foreign immigrants are attracted by the locals but also preferences of immigrants *not* to locate where other ethnicities are concentrated.

In fact, the variables that represent the spillover effects of technical clustering do not show a systematic pattern between the two models, except the interaction terms between R&D and the education industry, which show a consistent positive effect. However, there are some signs of technical clustering that are worthy of notice. The additional set of interaction terms between industries and major fields shows that education industries may attract human capital from out of state in Model 1; however, Model 2 shows that the effect is basically driven from the effect of the interaction of health science with education industries. A similar effect can be found in the interaction of health science with the biotechnology, information technology, and research work variable.

It appears that the spillover effect of industries on the likelihood that new S&E PhDs will stay in state is complex. When one compares Models 1 and 2, the matching effect of R&D and major fields, the interaction terms between R&D and major fields show some interesting patterns. The R&D interaction with social science and with psychology records a 5% significance level, while R&D interactions with mathematical sciences and with physical science record a 10% significance level. Although the (per unit) marginal change is small, if we consider the cumulative marginal effects (from minimum to maximum value) of these R&D interaction variables, their impact on the likelihood that the new S&E PhDs will stay in state cannot be ignored. For example, the cumulative marginal effects of the R&D interaction with social science, psychology, mathematical sciences, and physical science would reach an increment of 21, 31, 18, and 16.8%, respectively, compared to the reference (computer and information science technology).

5.6 Conclusion and Policy Implications: United States

What factors determine the differences of spatial distribution among native-born Americans, Chinese, Indian, and other foreign highly skilled workers (human capital) after they graduate? In particular, what are the factors conducive to the decision of newly minted S&E PhDs to remain in the state in which they were trained? The answer is not simple.

The concept of ethnic clustering is that people like to go to areas where there is a high concentration of their ethnicity. It is a more general measure than that of cultural clustering which uses the country of origin as its base. The match among main foreign S&E PhD graduates' citizenships may help to reveal how these foreign PhD recipients decide on job location.

From our model, we find a very diverse orientation on the effects of ethnic clustering. In general, states with high concentrations of Hispanics tend to attract PhDs from out of state; these states include Texas, New Mexico, California, and Arizona, with Hispanic concentrations of 32, 42.1, 32.4, and 25.3%, respectively. These states also have high concentrations of research industries. Another finding is that Chinese have a clear preference for Chinese-populated regions, which shows an expected significant and positive sign on the Asian dummy, controlling over the regional dummies for employers and graduation regions. From our data, states with the highest Asian concentrations are California Washington, New York, and New Jersey, with 10.9, 5.5, 5.5, and 5.7%, respectively. These results suggest two very interesting extensions to the concept of ethnic clustering: The first is that ethnic concentration may be due to the cumulative causation effects of additional migrants to a location, increasing the likelihood of still more migrants from the same origin moving to the same location in the future. Second, the effects of ethnic clustering may work to keep the graduates trained in a state and/or attract graduates of the same ethnicity from outside the state. This result may extend future research on the effects of culture on location choice.

There is also some truth in the technical clustering argument regarding the effects of graduates' location choices. For example, the nature of industries, such as the interactions of their research orientation and R&D budgets in the employers' state, may have product different spillover effects on the likelihood of graduates of different citizenships to stay in the state after graduation. What kind of spillover effects would attract graduates from out of state? States with high R&D funding in the information technology and research industries are particularly likely to attract Chinese from out of state, given that R&D in general may attract Chinese from out of state. What kind of spillover effects would encourage graduates to stay after graduation? From our models, it appears that states with high R&D funding in biotechnology would likely keep Indian graduates who had received their first PhD there. Also, education industries with high R&D funding may keep native-born Americans, other foreign citizens, and Indians (although the significance level of this finding is 10%) in the state in which they graduated, but this effect is not very significant in the case of Chinese graduates. Hence, there is a spillover effect from R&D and industrial policies from other states that determines whether the state is able to attract graduates. Thus, investment in R&D and the presence of research-oriented industries such as biotechnology and computer-related industries may attract more foreign talent. This research may suggest some facts about the strategies states need to pursue to keep S&E graduates in state, according to the states' comparative advantage.

Also generated from technical clustering are matching effects, the matching between graduates' field of specialization and employers' industrial and R&D funding. Our results show that the matching effects variable, indexed by the graduates' field of specialization that interacts with employers' industries, does not have a distinctly significant pattern. However, other matching effects variables, indexed by the interaction of graduates' field of specialization with R&D, did show some interesting positive effects on the likelihood of graduates staying in state after graduation. The marginal changes from minimum to maximum value of the R&D interaction with social sciences, psychology, mathematical sciences, and physical science will have a 21, 31, 18, and 16.8%, respectively, increment over the reference (computer and information technology).

These findings suggest that industrial policies and R&D policies may not have unilateral effects on all the S&E PhDs. Ethnic clustering also seems to play an important role in S&E graduates' location choices, depending on the match of citizenship and ethnicity. In summary, this chapter shows that technical clustering and ethnic clustering may be far more complicated than past studies have shown, when different citizenships are compared. In this regard, our work extends the previous regional competitive advantage and the network-based entrepreneurship and industrial agglomeration story in Silicon Valley's startups and Route 128's big companies, described in Saxenian (1994), by incorporating the dynamic effects of technical and ethnic clustering.

Nonetheless, further research may be helpful for sorting out the dynamic effects of technical clustering and ethnic clustering in these models. Currently our model focuses on the way location of specific industries and level of R&D funding may affect the likelihood of different foreign graduates staying in the state in which they were educated. In a dynamic context, one may also need to consider the mobility of firms/industries to a location where there is a rich labor pool. For example, one leading manufacturer of personal computers, with a historical location in North Dakota, moved to San Diego for a better pool of human capital (Porter 2000, page 267).

In addition, we may want to extend the analysis to the employers' labor preferences. It is the labor demand side of the story that may matter to the spatial pattern of human capital because employers' hiring practices may be based on their personal or cultural preference. An extreme example would be labor market discrimination driven by employers' preference; for example, instead of hiring the most talented job candidate in the field, an employer may respond differently to different citizenships and ethnicities. Bertrand and Mullainathan (2004) conducted an Internet-based experiment on the effect of race on the likelihood of employers responding to job applicants. For their "applicants," they constructed identical backgrounds but different names that could be identified as either black or white; they found that white-sounding names had a 50% higher chance of receiving a callback from an employer than did black-sounding names.

Our findings suggest several policy implications. States may want to develop their educational and industrial strategies to take into account the specific location preferences of different citizenships. Given that, we also would like to highlight the importance of matching the following factors: local competitive advantage and industrial structure; educational institutions and industrial policies; educational outcomes and R&D policies. These policy matches, indeed, are not new to our readers. However, when taking into account the diverse responsiveness from citizenship and ethnicity to policy matching, we will see how powerfully this matrix may transform to viable entrepreneurships that contribute to development of local economies, such as the phenomenon of Silicon Valley, with extended network, capital, and invaluable labor resources.

References

Alba RD, Logan JR (1991) Variation on two themes: Racial and ethnic patterns in attainment of suburban residence. Demography 28:431–453

Alba RD, Logan JR (1992) Analyzing locational attainments: Constructing individual-level regression models using aggregate data. Sociol Meth Res 20:367–397

Alba RD, Logan JR (1993) Minority proximity to whites in suburbs: An individual-level analysis of segregation. Am J Sociol 98:1388–1427

Bartel AP, Koch. MJ (1991) Internal migration of U.S. immigrants. In: Bowd JM, Freeman RB (eds) Immigration, trade, and the labor market. University of Chicago Press, Chicago
Bertrand M, Mullainathan S (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Am Econ Rev 94:991–1013
Blau P, Duncan OD (1967) The American occupational structure. Free Press, New York
Borjas G (1994) Ethnicity, Neighborhoods, and Human-Capital Externalities. Am Econ Rev 85:365–390.
Borjas G (2005) Immigration trends in the New York metropolitan area. Fed Res Bank NY Econ Pol Rev 2005:91–101
Chakravorty S, Koo J, Lall SV (2003) Metropolitan industrial clusters: Patterns and processes. Paper presented at the 2003 UNU/WIDER Tokyo conference on Spatial Inequality in Asia
Chiswick BR, Lee YL, Miller PW (2002) The determinants of the geographic concentration among immigrants: Application to Australia. IZA Discussion Paper No. 462
Collaborative Economics (2005) 2005 Index of Silicon Valley. Joint Venture: Silicon Valley Network
Feldman MP, Audretsch DB (1999) Innovation in cities: Science-based diversity, specialization and localized competition. Eur Econ Rev 43:409–429
Florida R (2002) The economic geography of talent. Ann Assoc Am Geog 92:743–755
Freeman R, Weinstein E, Marincola E et al (2001) Competition and careers in biosciences. Science 294:2293–2294
Fujita M, Ogawa H (1980) Equilibrium land use patterns in a non-monocentric city. J Reg Sci 20:455–475
Gross DM, Schmitt N (2003) The role of cultural clustering in attracting new immigrants. J Reg Sci 43:295–318
Gross DM, Schmitt N (2006) Why do low- and high-skill workers migrate? Flow evidence from France. Working Paper No. 1797, CESifo
Helsley R, Strange W (1990) Matching and agglomeration in a system of cities. Reg Sci Urb Econ 20:189–212
Koo J (2005a). Knowledge-based industry clusters: Evidenced by geographic patterns of patents in manufacturing. Urb Stud 42:1457–1505
Koo, J (2005b). Technology spillovers, agglomeration, and regional economic development. J Plan Lit 20:99–105
Krugman P (1991) Increasing returns and economic geography. J Polit Econ 99:483–499
Krugman P (1996) Development, geography and economic theory. MIT Press, Cambridge, Mass
Logan JR, Alba RD (1993) Locational returns to human capital: Minority access to suburban community resources. Demography 30:243–268
Marshall A (1890) Principles of economics. MacMillan, London
Massey DS, Arango J, Hugo G et al (1993) Theories of international migration: A review and appraisal. Pop Dev Rev 19:431–466
National Science Foundation (1998) International mobility of scientists and engineers to the United States – Brian drain or brain circulation? NSF 98-316, June
National Science Foundation, Division of Science Resources Statistics (2001) 2001 Survey of doctorate recipients
National Science Foundation, Division of Science Resources Statistics (2003) Characteristics of doctoral scientists and engineers in the United States: 2001. NSF 03-310
Phinney JS (1996) When we talk about American ethnic groups, what do we mean? Am Psychol 51:918–927
Porter ME (2000) Locations, clusters, and company strategy. In: Clark GL, Feldman MP, Gertler MS (eds) The Oxford handbook of economic geography. Oxford University Press, Oxford, UK
Saxenian A (1994) Regional advantage: Culture and competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, Mass.
Stephan PE, Sumell AJ, Black GC et al (2004) Doctoral education and economic development: The flow of new PhD's to industry. Econ Devel Q 18:151–167

U.S. Census Bureau (2000) Census 2000 summary file 1, matrix P8
U.S. Department of Homeland Security, Office of Immigration Statistics (2004) Characteristics of specialty occupation workers (H-1B): Fiscal year 2003
U.S. Immigration and Naturalization Service (2002) Report on characteristics of specialty occupation workers (H1-B): Fiscal year 2000

Conclusion

Neslihan Aydogan

In this book, we probe into agglomerations from economics and social perspectives. It is no secret that clusters of economic activity are invaluable for the growth of regions in particular and countries in general. Our analyses show the characteristics of these formations of economic and social density across space.

To this end we angle ourselves so that we distinguish between high and low-tech economic activities as these two have very different implications on innovation and economic growth. High-technology industries employ knowledge workers intensely and involve frequently in R&D activities producing process and product innovations. Because of these two characteristics, firms in such industries have a different social and economic existence. One such reality of existence is portrayed in Chapter 1. In particular, we find that an environment with diverse economic activity helps high-tech firms to thrive as opposed to the concentration of a single type of economic activity.

As such is the case, we change tracks into the social and indulge the reader into the social world of contracts that take place among high-technology firms. The intriguing feature of such contracts is the very incentive and ability to cheat on a partner as knowledge tends to be sticky, i.e., difficult to transfer. The beauty of geographical proximity and, one step further, a networked cluster is to curb these incentives via their social and hence economic implications. Put simply a network and proximity prevents cheating to be an economically profitable or socially attractive endeavor. This is precisely what we observe in Chapter 2. We find that proximity and networked clusters allow economic agents to be able to engage in cheaper-to-administer contracts.

Next we further investigate the role of reciprocity in transferring knowledge across geographical distances. The premise is that as economic and social beings we value reciprocity as it is a motivating factor in all types of exchange. In the context of exchanging difficult-to-transfer knowledge we find that if parties to the exchange have this social structure, that is if learning is mutual, then partners in a coalition of exchange do not need to economize on distance. The bottom line is that regardless of how far we are if we complement each other and learning mutually takes place, cheating is unlikely to be a problem. Hence reciprocity is pivotal in altering the role that proximity plays in exchange. Therefore, Chapter 3 in great harmony with the previous two chapters nullifies the statement that economic exchange does not have a social component. In fact it does so to the extent that formal and informal contracts live in the context of the social and places individuals in the physical world allowing geographical proximity to create the social sphere and making social exchange inadvertently a part of exchange.

Considering the role exchange plays across space allures us into the world of knowledge workers who are the actors of such exchange. In a country like the United States, with many immigrant workers, we have to be greatly concerned in educating and attaining knowledge workers. Hence, a region while it invests in human capital has to also invest in retaining that capital. Human capital is likely to be sticky if it is nurtured via the social ties in a region. Having stated this, the challenge in an immigrant economy is the difficulty to form such atmosphere for its foreign workers so that they stay. Chapter 4 compares naturalized Americans with native-born Americans as this very particular characteristic, i.e., social networks, is difficult for the latter group to cascade given the relatively short time span they spend in a region and their background. The chapter also focuses on the other characteristics of a region that could motivate the knowledge workers to stay after graduation, such as the family factors and field of specialization. We find significant differences between the naturalized Americans and natives in these regards. Hence we find that keeping a diversity of knowledge workers in a locale calls for targeted education and industrial policies.

Chapter 5 delves into understanding the accumulation of knowledge workers by their ethnicity. We compare the native-born Americans and two large immigrant groups in the United States, Indians and Chinese. In this chapter the main social component of the analysis is ethnicity, taking into account largely the social networks and ethnic clustering and the reasons behind such occurrence. The more economics-related component of the analysis is what we call the technical clustering which addresses the clustering of workers with similar skills. We find that both of these play a role in the clustering activity of knowledge workers with diverse backgrounds.

Overall we tap into a relatively virgin area: we dare to question the social and economic components of clustering from a variety of angles and we believe our findings are likely to push the envelope, in particular in relation to the dynamics of social components of exchange and industrial clusters.

Index

Note: The letters t and f in the index refers to tables and figures respectively
For example 32t refers to table in page 32