2005

Three essays on migration: Its interaction with regional wage differentials and occupational choice and its application in different discrete choice models

Christiadi
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Christiadi, "Three essays on migration: Its interaction with regional wage differentials and occupational choice and its application in different discrete choice models" (2005). *Graduate Theses, Dissertations, and Problem Reports*. 2323.

<https://researchrepository.wvu.edu/etd/2323>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

**Three Essays on Migration:
Its Interaction with Regional Wage Differentials and Occupational
Choice and Its Application in Different Discrete Choice Models**

Christiadi

**Dissertation submitted to the
College of Business and Economics
at West Virginia University
in partial fulfillment of the requirements for the degree of**

Doctor of Philosophy

in

Economics

**Brian J. Cushing, Ph.D., Chair
Sudeshna C. Bandyopadhyay, Ph.D.
Stratford M. Douglas, Ph.D.
Ge Lin, Ph.D.
Peter V. Schaeffer, Ph.D.**

Department of Economics

Morgantown, West Virginia

2005

**Keywords:
Migration, Wage Differentials, Occupation, Discrete Choice, IIA, Nested Logit,
Mixed Logit**

ABSTRACT

Three Essays on Migration: Its Interaction with Regional Wage Differentials and Occupational Choice and Its Application in Different Discrete Choice Models

Christiadi

This dissertation consists of three essays that examine two important roles of migration, promoting regional wage convergence and increasing economic efficiency. The first essay is a survey of the migration and migration-related literatures, examining why migration and regional wage differentials can co-exist over time. The direct reason is that the classical assumptions leading to regional wage convergence are rarely fully applicable in the real world. This essay reveals various real-world conditions violating those assumptions and explains how they can either create or magnify regional wage differentials. The essay also reveals how the difficulty in distinguishing wage differentials from true wage disequilibrium has posed serious measurement problems. The second essay examines how individuals' occupational choice affects their destination choice in a simultaneous equation framework. The results show that individuals are more likely to choose destinations where they can maximize the returns to their occupational skills. This essay also develops the application of a two-step maximum likelihood method to examine two simultaneous choices. Considering that individuals frequently face simultaneous choices, this methodology fills an increasingly important need. The third essay investigates whether violating the Independence from Irrelevant Alternatives (IIA) assumption, as standard discrete choice models do, poses serious estimation problems. It compares the results from nested logit and mixed logit models, which relax IIA, with the standard conditional logit model, which assumes IIA. The essay shows that their results, while statistically different, are qualitatively similar. Their parameter estimates are of the same sign, of comparable statistical significance, and of comparable magnitudes. Given the substantial computational cost of the more complex nested and mixed logit models, the finding that a well-specified, but computationally much simpler, conditional logit model may suffice is encouraging. The results also suggest that the growing body of conditional logit migration models may not be inherently flawed.

To my Mom, whose love and prayer guide me all the way.
To Isna, Syihan, and Fahmi, who make every minute of my life worthwhile.

ACKNOWLEDGEMENTS

I would like to thank Dr. Brian Cushing, to whom I am forever indebted for his endless support, care, and encouragement. He is my advisor, my mentor, as well as the best friend I could ever have, who is always there for me and makes things better.

I would also like to thank all the committee members for their overall support. I thank Dr. Peter Schaeffer for his willingness to work closely with me on the first essay, passing on his expertise. Thanks to Dr. Stratford Douglas for providing me with thoughtful comments and all the assurance I need to deal with econometrics-related challenges. I am thankful to Dr. Ge Lin for helping me start working on this dissertation. I thank Dr. Sudeshna Bandyopadhyay for her thoughtful advice and comments.

I extend my gratitude to Dr. George Hammond for his caring and complete advice, his pep talk, as well as his intervention to keep the dissertation moving ahead. Thanks to my other colleagues, Dr. Tom Witt, Dr. Mehmet Tosun, Randy Childs, as well as Dr. Sean Snaith for their support and willingness to accommodate my “dissertation time” during my work at the Bureau of Business and Economic Research and Business Forecasting Center. Finally, thanks to my wife, Isna, for her companionship, prayers, and sacrifices she has made in making sure that I would finish this dissertation.

The standard disclaimers apply.

CONTENTS

Abstract	ii
Dedication	iii
Acknowledgement	iv
Contents	v
List of Tables and Figures	vii
Chapter 1. Introduction	1
1.1. Introduction	1
1.2. Dissertation Outline	2
Chapter 2. Why Migration Does Not Necessarily Equate Regional Wage Differentials: A Survey	6
2.1. Introduction	6
2.2. Regional Wage Differentials in the United States	7
2.3. Classical Theory on the Wage-Equilibrating Role of Migration	10
2.4. Increasing Returns to Scale	11
2.5. Self-Selection	13
2.6. Migration Costs	14
2.7. Productivity Differentials	15
2.8. Compensating Differentials	16
2.9. Wage Rigidity	17
2.10. Other Factors	21
2.11. Wage Differentials versus Wage Disequilibrium	22
2.12. Migration and Regional Wage Differentials	24
2.13. Migration and Regional Wage Disequilibrium	26
2.14. Future Directions	27
2.15. References	30
Chapter 3. The Interaction between Individual's Destination Choice and Occupational Choice, A Simultaneous Equations Approach	37
3.1. Introduction	37
3.2. Literature Review	38
3.2.1. Aggregate Model and Individual Model of Migration	38
3.2.2. Migration and Occupational Change	39
3.2.3. Modeling Simultaneous Discrete Choice Models	40
3.3. The Basic Models	41
3.4. Data	45
3.5. Empirical Results	46

3.5.1.	Multinomial Logit Model of Occupational Choice	46
3.5.2.	Conditional Logit Model of State Destination Choice	46
3.6.	Notes on the Independence from Irrelevant Alternatives (IIA) Assumption	48
3.7.	Concluding Remarks	51
	Appendix 3.1. Adjusting the Variance	58
	Appendix 3.2. Adjusting the Variance: The Real Form of the Conditional <i>Logit</i> Model Equation	62
	References	66
Chapter 4.	Comparing Conditional Logit, Nested Logit, and Mixed Logit State Destination Choice Models	69
4.1.	Introduction	69
4.2.	Nested Logit and Mixed Logit Models	71
4.3.	Discrete Choice Models	73
4.4.	Model Specifications	84
4.5.	Data	86
4.6.	Empirical Results	86
4.7.	Conclusions	88
	Appendix 4.1 Explanatory Variables	96
	Appendix 4.2 Characteristics of the Individuals in the Sample	96
	References	97
Chapter 5.	Conclusions	99

LIST OF TABLES AND FIGURES

List of Tables

Table 3.1	Distribution of Managerial and Professional Occupations across Industrial Sectors by State, 1990	52
Table 3.2	Distribution of Sales Occupations across Industrial Sectors by State, 1990	53
Table 3.3	Distribution of Operator Occupations across Industrial Sectors by State, 1990	54
Table 3.4	Distribution of Laborer Occupations across Industrial Sectors by State, 1990	55
Table 3.5	Multinomial Logit Model of Occupational Choice Selected Variables (N=6359)	56
Table 3.6	Conditional Logit Model of State Destination Choice (N=6359)	57
Table 3.7	Small-Hsiao IIA Test Results	57
Table 4.1	The Relationship between Probability Distribution Imposed on the Random Components and Type of Discrete Choice Model Applied	91
Table 4.2	Regression Results: Comparison among Conditional Logit, Nested Logit, and Mixed Logit Models	92
Table 4.3	Likelihood Ratio Tests	93

List of Figures

Figure 2.1	Standard Deviation of Wages per Job across States in the U.S. by Industry (SIC)	8
Figure 2.2	How Migration Equilibrates Wage Differentials	11
Figure 4.1	Nesting Specification of the First Nested Logit Model	94
Figure 4.2	Nesting Specification of the Second Nested Logit Model	95

Chapter 1

Introduction

1.1. Introduction

Population migration is the primary determinant of regional population change, dwarfing the small net natural change. Between 1995 and 2000, over 22 million individuals moved across states, representing over 8.5 percent of the U.S. population 5 years old and above. A change of this magnitude has a large impact. For example, migration affects regions' age composition, businesses' location decisions, wage rates, state and local tax revenues, public service allocations, and industrial clustering. Considering the significant impact of migration on the economy of a region, it is important to better understand the migration process. This dissertation furthers our understanding of migration by looking into two of its most important roles: (1) promoting regional wage convergence and (2) increasing economic efficiency through the sorting of workers and, indirectly, businesses as they search for the best location.

Neoclassical economic theory states that workers would migrate from low to higher wage regions and that such mobility should result in interregional wage convergence. Facts, however, have not adequately supported this claim. Despite large migration flows across U.S. states, previous studies found that state wage differentials have not disappeared. Considering the sound logic of the neoclassical argument, this suggests that other factors are at work. The first part of this dissertation provides a comprehensive survey of the migration and migration-related literatures to analyze how various factors affect the role of migration in promoting regional wage convergence. Previous studies have not analyzed this issue thoroughly enough, providing piecemeal analysis instead.

The second part of the dissertation looks into the role of migration in allowing workers and businesses to search for the best fit. In this process of searching for the best fit, individual's occupation plays an important role. Firms look for workers whose occupational skills best fit their operation and individuals look to find firms that can maximize the returns to their occupational skills. This study examines how individuals'

choice of occupation affects their choice of destination. Unlike previous studies, the research recognizes that some people decide their occupational choice and destination choice simultaneously. The analysis applies a two-stage maximum likelihood estimation, which has not been applied in migration studies. Since individuals frequently face simultaneous choices, this methodology would also be useful for other individual-level studies.

The above study uses a conditional logit model to examine the destination choice among U.S. states. This model, however, violates the Independence from Irrelevant Alternatives (IIA) property it imposes, which can cause the model to incorrectly predict the probability of a destination being chosen. The third part of this dissertation applies two other models: a nested logit, which partially relaxes IIA, and a mixed logit, which fully relaxes IIA. These two models are much more complex, compared with a conditional logit model. Estimating them incurs significant time and computational costs. This study then examines how much the outcomes of these two models differ from those of the standard conditional logit model. Based on the comparison, this study assesses whether relaxing the IIA assumption warrants application of the more complex nested logit or mixed logit models. The remainder of this introductory chapter provides a more detailed outline of the three essays.

1.2. Dissertation Outline

The next three chapters present three essays on the interaction between migration, regional wage differentials and individual's occupational choice. Chapter 5 makes concluding remarks.

Chapter 2 reports the results of a survey of the migration and migration-related literatures. In this essay I examine why migration and regional wage differentials can co-exist over time while classical economic theory states that migration should eliminate wage differentials. The direct reason for the persistent differentials is that the classical assumptions imposed are rarely fully applicable in the real world. Migration promotes regional wage convergence only when seven classical assumptions are met: (1) perfect

competition existing in all markets; (2) wage flexibility; (3) labor and job homogeneity; (4) no legal barriers in moving across regions; (5) insignificant costs of migration; (6) perfect information flows on wages across regions; and (7) production functions exhibiting constant returns to scale.

I then discuss various real world conditions that violate these classical assumptions and explain how each affects the interaction between migration and regional wage convergence. Those conditions include increasing returns to scale (agglomeration), self-selection in migration, significant migration costs, productivity differentials, compensating differentials, wage rigidity, and other factors such as aggregation bias, sex or racial discrimination, and slow market adjustments. I find that each of these conditions can contribute to the persistent wage differentials in one of two ways: (1) it widens existing differentials or (2) it creates the wage differentials. The first two conditions can widen wage differentials. If destination regions benefit from increasing returns to scale operation or agglomeration, or if migration selects individuals of higher quality, then migration would further reinforce the economic activity in the destination, widening the wage gap between origins and destinations. The rest of the conditions can create wage differentials. That is, everything else considered, regional wage differentials develop to the extent that they represent migration costs, differences in productivity factors, compensating differentials, differences in the extent of wage rigidity, and so on across regions.

That various components can make up the observed regional wage differentials raises a measurement issue. Researchers cannot identify exactly which components constitute the observed wage differentials. In order to empirically test the role of migration, researchers must identify if a true wage disequilibrium exists within the observed wage differentials and, if so, how large. This is necessary because migration's equilibrating role is expected to work on wage disequilibrium but not necessarily on wage differentials. If researchers fail to control for the other components, then test results might falsely conclude that migration fails to promote regional wage convergence. These observed wage differentials may reflect factors such as migration costs or differences in productivity

rather than true regional wage disequilibrium. This measurement issue is discussed more fully in the last part of Chapter 2.

In Chapter 3, I examine the interaction between individual's occupational choice and destination choice among U.S. states. More specifically, I represent both choices as a multinomial logit model of occupational choice and a conditional logit model of state destination choice. I consider that the two choices are made simultaneously and, therefore, apply a two-step maximum likelihood estimation. The unusual merger of these models in a simultaneous equations framework requires derivation of a unique variance-covariance matrix. Otherwise, statistical tests will be incorrect. I apply Murphy and Topel's method to compute the adjusted variance-covariance matrix.

As expected, I find that individuals are more likely to choose destinations where they can maximize the returns to their occupational skills. After adjusting the variance, I find that the statistical results do not vary qualitatively, compared with the unadjusted results. This, however, does not imply that the adjustment is unnecessary. In some other situations, the adjustment may significantly alter statistical conclusions.

Chapter 4 can be perceived as an extension of Chapter 3. In Chapter 4, I examine and compare three models of state destination choice: a conditional logit, a nested logit, and a mixed logit. The three models differ in terms of their treatment of the IIA assumption. The conditional logit model carries the IIA assumption, the nested logit model partially relaxes it, and the mixed logit model fully relaxes it. I find that estimating the two latter models, the nested logit and the mixed logit, take considerably more time than the standard conditional logit model.

The results of these three models, while statistically different, are qualitatively very similar. The parameter estimates of the three models are of the same sign, of comparable statistical significance, as well as of comparable magnitudes. Almost all variables maintain their statistical significance, in terms of not changing from being statistically significant to insignificant or otherwise. The results suggest that if researchers are more

concerned with average preferences, the standard, easy-to-use conditional logit model should provide a good approximation. On the other hand, if they want to use the results to forecast substitution patterns among alternatives, models that relax IIA would likely be a better choice, despite the significantly increased complexity and computational cost.

Chapter 2

Why Migration Does Not Necessarily Equate Regional Wage Differentials: A Survey

2.1. Introduction

The neoclassical economic theory of migration implies that workers move from low to higher wage regions and that such mobility should therefore result in movement toward interregional wage convergence. Despite large interstate migration flows in the United States during the past century, regional wage differentials have not disappeared. Easterlin (1961) found that regional wage differentials have persisted for long periods of time, while Borts (1960) concluded that they have not been responsive to migration flows. Recent research, Card and DiNardo (2002) and Card et al. (2003), reveal that wage inequality in the U.S. increased sharply during the 1980s and remained relatively stable in the 1990s. The fact that wage convergence has not occurred, combined with the sound logic of the neoclassical argument, suggests that other factors are at work.

Several migration literature surveys have discussed the interaction between migration and regional wage differentials (e.g. Greenwood (1975b), Goetz (1999), and Armstrong and Taylor (2000)). However, they do not specifically focus on this issue and the discussions are not comprehensive, providing only pieces of the analysis. This study fills the gap by providing a more comprehensive survey of migration and migration-related literature. It compiles a more complete list of those influential factors and discusses at length how each affects the migration's role in promoting regional wage convergence. More specifically, this study looks at factors such as increasing returns to scale and agglomeration, self-selection in migration, significant migration costs, productivity differentials, compensating differentials, wage rigidity, and other factors such as aggregation bias, sex or racial discrimination, and slow market adjustments. In addition, this study enhances the discussion by considering the fact that migration is about individuals' decisions rather than a region's aggregate decision. Consequently, because individuals' utility is by nature subjective, variations in the individuals' utility also determine migrants' destination choice.

In summary, by looking closely at those factors, this study lets us learn how the measures of observed regional wage differentials differ from that of true regional wage disequilibrium, how various factors and differences in individuals' utility come into play to distinguish between the two measures, and how migration interact with them. In the end, this discussion leads us to see the possibilities where migration may or may not necessarily promote regional wage convergence as depicted by the classical theory.

The next section of this essay describes regional wage differentials in the United States from 1969 to 2000, followed by a review of the classical economic theory of migration and wage rate convergence. Then follow several sections discussing various real world deviations from the classical assumptions and how each affects the migration's role in promoting wage convergence. After a discussion on what can be learned from preceding sections, the final section offers conclusions and suggestions for future research.

2.2. Regional Wage Differentials in the United States

Figure 2.1 shows the standard deviation of wages per job across U.S. states from 1969 to 2000. Since wages differ by industry and industrial mix varies by state, the figures compare wages across states controlling for type of industry. The standard deviation measures how average wages across states differ from the national average wages. In a perfect world, where the law of one wage applies to each of the labor markets in each sector, the standard deviations would equal zero. Nonzero standard deviations reflect regional wage differentials across states.

The figure shows that wages in the mining sector have the largest and most volatile deviations until 1998. Wage differentials in the finance-insurance and manufacturing sectors are also large and increasing over time. Compared with other sectors, wage differentials in the retail trade, state and local-government, and the federal civilian sectors are much more stable over time.

Figure 2.1

Standard Deviation of Wages per Job across States in the U.S. by Industry (SIC)

Source: U.S. Bureau of Economic Analysis and U.S. Bureau of the Census.

If persistent standard deviations in wage rates mainly represent regional cost of living differences, then we would expect wage deviations in each industrial sector to follow the deviations in the cost of living, both in terms of magnitude and trend. Unfortunately, a standard measure of time-series cost-of-living differences is not available. While not a perfect proxy, I use the deviations in wages for the state-local government or for the federal-civilian sector to measure of cost-of-living deviations. I then compare wage deviations in each industrial sector with those in the two government sectors.

Figure 2.1 shows that wage deviations in four out of seven industrial sectors (mining, manufacturing, finance-insurance, and wholesale trade) are in most cases larger than those in the two government sectors. In addition, contrary to the stable deviations in the two government sectors, their deviations tend to increase over time. This suggests that,

even after controlling for cost of living differences, wage differentials across states persist and tend to increase over time. Thus, within these sectors, factors other than cost of living account for regional wage differentials. In contrast, wage deviations in the three other sectors (transportation-utility, services, and retail) fall below or within the range of deviations in the two government sectors, suggesting that those deviations may indeed represent differences in the cost of living. At least for four industry sectors, these results contradict the findings of the convergence literature. For instance, Barro (1991) and Barro and Sala-i-Martin (1991) show that economies, including states, tend to converge. Their argument is based on finding a negative and significant β -convergence coefficient, suggesting that poor economies grow faster than, and catch up with, rich economies.¹

The two types of studies, however, may be different because they use different variables to measure regional economic differentials. In this study regional economic differences are measured as wage differentials while in typical convergence studies they are measured as personal income differentials. The two measures are different for two reasons. First, personal income includes not only wages and salary but also proprietor's income, other labor income, income on assets, and income from transfer payments such as Medicaid, Medicare, Social Security, unemployment compensation. Transfer income itself typically is negatively correlated with wages and salary. Secondly, personal income is an aggregate measure of a region's economy, which, unlike wages or salary, cannot take into account differences in the industrial mix across regions. Hoover and Giarratani (1999) suggest that the personal income measure is more appropriate when comparing rich versus poor regions, while an earnings measure is more appropriate when comparing how well regions pay workers – the focus of this study.

Quah (1993) casts some doubt on the conclusions of Barro (1991) and Barro and Sala-i-Martin (1991). Applying Galton's fallacy shows that finding a negative β -convergence does not necessarily imply that cross-sectional differences diminish over time. He further

¹ A typical convergence equation takes the form of: $\ln(Y_t - Y_0) = \alpha + \beta \ln Y_0 + \gamma \ln X$, where Y_t represents economic output at time t and Y_0 at time 0 (the initial point). $\ln(Y_t - Y_0)$ then measures economic growth. X represents other factors that affect economic growth. If convergence were true then poor countries, those with lower Y_0 , would likely grow faster. This implies a negative β .

argues that the world economies are “tending – in the long run – towards either the very rich or the very poor, with the middle-income classes vanishing (p.441).”

Thus, the debate regarding the validity of regional wage differentials remains unsettled (Dickie and Gerking (1989)). Different studies may use different measures to examine regional wage differentials. Regardless of the measures used, the literature agrees that, at least on the surface, real wage differentials exist in the United States. The issue is not whether differentials exist, but how we explain the meaning of those wage-differentials, which is this paper’s purpose.

2.3. Classical Theory on the Wage-Equilibrating Role of Migration

Classical microeconomic theory predicts that labor (capital) will tend to move from places with low returns to those with higher returns. Such mobility will shift the supply curves of labor (capital) in the destination to the right, which in turn will pull wages (rates of returns to capital) down. On the other hand, because of that mobility the supply curves of labor (capital) in the origin will shift to the left, which in turn will push wages (rates of returns to capital) up. The mobility occurs until wages or the returns to capital are equalized across regions (see Figure 2.2).

However, the above scenario works if and only if the economy satisfies several classical assumptions². Those assumptions consist of: (1) perfect competition existing in all markets; (2) wage flexibility; (3) labor and job homogeneity; (4) no legal barriers in moving across regions; (5) insignificant costs of migration; (6) perfect information flows on wages across regions; and (7) production functions exhibiting constant returns to scale. These classical assumptions guarantee that markets clear, which implies that migration equalizes regional wage differentials.

The classical assumptions do not adequately represent reality in a world where markets do not always clear. Classical theory refers to long-run and not to short-run equilibrium conditions. In the short run, an economy constantly changes, with its short-run

²As described for example by Armstrong and Taylor (2000) and Kaufman (1994).

disequilibria influencing its long-run trends and rate of convergence. The complex interactions among economic agents that characterize a modern economic system mean that a shock occurring in one market can create ripple effects and generate new shocks elsewhere in the system, so that the economy continues to experience shocks almost constantly. It then becomes difficult to project the long-run equilibrium. Thus, maybe the only pragmatic lesson offered by neoclassical theory is the general direction of change.

Figure 2.2. How Migration Equilibrates Wage Differentials

Higher wage in Region A attracts migration from Region B, increasing the supply of labor in Region A, and pulling W_{A1} down. On the other hand, the supply of labor in Region B falls, which in turn pushes W_{B1} up. The process stops when wages in the two regions reach the equilibrium wage of W^* .

The next sections discuss six possible real-world conditions that violate classical assumptions and their roles in migration and regional wage convergence. These six conditions include increasing returns to scale, self-selection, migration costs, productivity differentials, compensating differentials, and wage rigidity.

2.4. Increasing Returns to Scale

At least two traditions deal with scale economies relative to migration and equilibrium, one represented by the gravity of spatial interaction (Lowry (1966) , Vanderkamp (1976)) and the other by the works of Myrdal (1957), Hirschman(1958) and, very recently,

Krugman (1991). Gravity models are based on the assumption that migration flow is positively correlated with “mass,” represented by population or other measures of economic opportunities. Empirical studies find a significant and positive role of a destination’s population or employment size in attracting migrants (Greenwood (1975b), Greenwood (1997), and Mueller (1985)). Because of agglomeration economies, migration can continually increase the number of people and businesses in the destination, until the positive agglomeration effects have been exhausted.

Regional economic theory and empirical results confirm that firms and workers benefit from increasing returns to scale and positive externalities generated by urban agglomerations. Agglomeration benefits make the region more productive (O’Sullivan (1995) and Krugman (1991)). Krugman (1991), citing Arthur (1990), and harking back to the much earlier work of Myrdal (1957), describes this as a circular causation: “manufactures production will tend to concentrate where there is a large market, but the market will be large where manufactures production is concentrated.” This agrees with Muth (1971), revealing a simultaneous relationship between migration and a destination’s economic performance. Several studies found that growing regions attract migrants and that strong population growth from migration fuels economic growth (e.g. Greenwood and Hunt (1984), Greenwood (1975a), Greenwood (1981), and Crown (1991)). The opposite is also evident, where deteriorating regions typically suffer sustained population losses (e.g. Gramlich et al. (1992), Goetz and Debertin (1996), and Bedau (1998)).

How might cumulative causation affect convergence? If increasing returns to scale and agglomeration govern the economic interactions between origin and destination, then migration from the low-wage to the high-wage region may not equalize wage differentials, but magnify them instead. The flow of migrants to the destination region shifts the demand curve there to the right. Because of scale and agglomeration economies, the new equilibrium at larger employment does not necessarily result in a lower wage rate, so that the incentive to leave the origin region may not diminish until such scale effects decline. Isserman (2005) shows that the presence of large cities in a region often distinguishes high-income from low-income regions and Hammond (2004)

found evidence of income divergence between metropolitan and nonmetropolitan regions in the United States during the last thirty years.

2.5. Self-Selection

The human capital model of migration (Sjaastad (1962)) implies that migrants are not a random sample of the whole population, but systematically self-select based on age, education and training, and other personal traits. The migration literature presents strong empirical evidence in support of self-selection (Greenwood (1975b) and Greenwood (1997)).

Inmigration of individuals with human capital above the average of the destination regions is referred to as positive self-selection (e.g., Plane and Rogerson (1994)). Blau and Duncan (1976) argue that domestic migration in the United States is positively selective. Even after controlling for productivity factors, they found that migrants still have higher productivity than nonmigrants. This finding suggests that migrants may be more strongly motivated to succeed at work and make a greater effort. As Allen and Eaton (2005) claim, migrants are less risk averse than those who stay.

Migration allows people to relocate to regions where the returns to their skills are higher as they find a better match between job requirements and their skill endowments (Borjas (1992)). The larger the benefits realized from this match, the higher the rewards the region can offer. Over time this fuels a continuous cycle whereby successful regions are able to offer higher rewards to attract skilled workers, thereby further enhancing their competitive advantage in the labor market.

In summary, labor migration tends to select people with superior skills. Destination regions benefit from positive self-selection, while origin regions suffer a loss of human capital. The literature refers to this phenomenon as the brain drain (e.g., Stark et al. (1997) and Schaeffer (forthcoming)), as well as Watanabe (1969) and Bellante (1978)). Similar to the effect of increasing returns to scale, positive self-selection prevents

migration from equalizing regional wage differentials, making them wider instead (Alonso (1971)).

2.6. Migration Costs

Following Sjaastad (1962), we distinguish between two types of mobility costs: monetary and psychic. The migration literature indicates that the monetary cost is small relative to the expected gains from migrating (Greenwood, 1975b). Thus, it is much more likely that psychic costs, not monetary costs, determine the migration decision. Psychic costs include the loss of location-specific human capital and moving away from family and friends. The longer individuals have lived in one place, the greater their attachments and location-specific knowledge are likely to be, therefore the greater are their psychic costs of moving. We can relate psychic costs to empirical evidence including the deterrent effect of distance, the decreasing likelihood of migration as age increases (Schwartz (1973) and Schwartz (1976)), and the formation of pockets of migrants of the same race or origins in destination regions (O'Sullivan (1995)).

Despite the small monetary costs associated with distance, empirical studies find a strong and significant deterring effect of distance. The effect remains strong even after controlling for adjacency as well as some measures of closeness to friends or relatives. Other unmeasured psychic costs account for this strong deterrent effect of distance. The more distant the destination, the greater the psychic costs that the migrant must bear.

The human capital model is compatible with empirical observations that the rate of migration decreases with age (Sjaastad (1962)). The model also allows incorporation of psychic costs. If older age correlates with length of time an individual has lived in one place, then older people have greater attachments and location-specific knowledge and, therefore, greater psychic costs of moving (Schwartz (1973) and Schwartz (1976)).

Psychic costs account not only for what people leave behind but also the uncertainties of a new place. Ability to make contact with friends of the same background or culture in the destination will definitely ease the psychic costs of dealing with unfamiliar

circumstances. Studies have suggested this as the reason we find pockets of migrants of the same race or ethnicity in big cities (O'Sullivan (1995)).

Overall, the literature agrees that migration costs have a significant deterrent effect on migration. How would this affect regional wage convergence? In this case, regional wage differentials can exist to the extent that they represent the magnitude of migration costs. If migration costs exceed the wage benefit of moving, then wage differentials will persist. They do not represent market disequilibrium, and thus do not necessarily induce migration.

2.7. Productivity Differentials

Different economies take different paths to become what they are today. Their natural endowments, productivity factors, and development strategies determine how they progress (Krugman and Obstfeld (1994) and Dixit and Norman (1980)). Depending on these factors, economies move forward to attain their own steady state level (Barro (1991) and Barro and Sala-I Martin (1991)). In the end, each economy finds its own economic structure. Typically, this manifests as regions having different industrial mix and occupational mix.

More importantly, differences in economic structure give rise to differences not only in terms of composition of the workers (by occupational skills, age, educational attainment etc.) but also in terms of workers' productivity. Hanna (1959) and Sjaastad (1962) showed that low-income states are dominated by low paying occupations and high-income states by high-paying occupations. Wojan (2000) also revealed that compared with high-income urban areas, low-income rural areas have greater concentration of low-paying operator occupations. In addition, the same occupations in different regions can yield different pay because they perform at different levels of productivity. For instance, an operator in a large manufacturing company in the rich region *A* may have higher-level responsibility than an operator in a small company in the poor region *B*, even though they have the exact same educational attainment, years of experience, and other productivity characteristics.

In summary, because of differences in economic structure across regions, workers with seemingly similar jobs in different regions can have job-related productivity differences, which in turn give rise to regional wage differentials. Thus, the observed regional wage differentials may actually represent the magnitude of job-related productivity differentials across regions, not real wage disequilibrium. This means that they will not necessarily induce migration, at least in the short run, and accordingly may persist.

2.8. Compensating Differentials

The concept of equalizing or compensating wage differentials in general refers to “observed wage differentials required to equalize the total monetary and nonmonetary advantages or disadvantages among work activities and among workers themselves” (Rosen (1986)). People are willing to take lower wage jobs when they are compensated with better job attributes, and to take unattractive jobs, such as mining jobs located in a remote village, when they are compensated with higher wages. The concept of compensating wage differentials emphasizes that workers consider not only wages but also job attributes. Typical job attributes include working condition, differences in location amenities such as climate, crime rate, and congestion; time flexibility; the composition of pay packages such as benefits, paid-annual leave, pensions; and cost of living (Rosen (1986)).

Job attribute differences can explain why wage differentials exist between rural and urban regions. Jobs in rural labor markets typically are associated with stronger ties between firms and workers than jobs in the urban labor market. In other words, jobs in rural labor markets provide more stability, thus receiving a lower wage than the same jobs in an urban labor market (Schaeffer and Gebremedhin (2004)).

The migration literature typically represents job attributes in terms of local amenities that come with the job. They showed empirically that destination amenity variables are statistically significant, implying that local amenities affect migrants’ destination choice

(e.g. Clark et al. (1996), Cushing (1987), Knapp and Graves (1989), and Schachter and Althaus (1989)).

Though theoretically sound, applying the concept of compensating wage differentials can be complicated in practice. Mortensen (2003) claims that while occupational risk differentials truly are compensated, the same might not be true for amenities. He conceptually shows that assuming both wages and amenities are normal goods, higher wages can be positively associated with better amenities when both employers and workers seek to maximize profits and utility. This does not agree with the concept of compensating differentials where wages should be negatively correlated with amenities.

Another complication is that job attributes can be hard to quantify. Job attributes can include factors such as job prestige, work hour flexibility, the presence of favorite recreational facilities, the presence of entertainment facilities, and proximity to friends and relatives. Each individual's utility function determines the value of these influences, thus making them hard to measure. For those reasons, empirically, regional wage differentials may partially, but not fully reflect compensating differentials.

Regardless of measurement issues, compensating differentials may explain persistent regional wage differentials. Two seemingly identical jobs can have different pay because one job may come with better job security, better opportunities for career advancement, better stability, as well as more preferable amenities. In this case, regional wage differentials do not represent wage disequilibrium, thus do not necessarily induce migration.

2.9. Wage Rigidity

Wage rigidity implies that wages do not adjust freely to changes in labor supply and labor demand. In the presence of wage rigidity, changes in supply or demand initiated by migration may not lead to the expected shifts in wage rates. Migration would not necessarily decrease regional wage differentials. Razin and Sadka (1995) found that with wage rigidity migration generates larger unemployment than it does without it. Among

various factors causing wage rigidity, four are well known, including efficiency wage theory, job contracts, labor unionization, and minimum wage legislation.

Efficiency Wage – Firms have the flexibility to set their own wage structure in order to maximize profits. One way is by applying efficiency wages whereby firms pay wages higher than the market wage. At least two reasons underlie efficiency wages (for more complete explanations see Stiglitz (1986) or Campbell III and Kamlani (1997)). First, efficiency wages can be considered as a premium wage designed to boost workers' morale while discouraging shirking, for which workers would lose the premium. Second, efficiency wages reduce firms' monitoring and turnover costs. This could be significant, especially for large firms and firms at risk to lose workers who have gained intensive firm-specific skills obtained through on-the-job training. Larger firms tend to bear higher monitoring and turnover costs, thus benefit more by applying efficiency wage policies (Katz (1986), Dunn (1984), and Brown and Medoff (1989)).

Firms applying efficiency wage policies may not lower wages even in slack labor markets because doing so could cut their long-term profits. As a result, efficiency wage policies do not let wages adjust freely to changes in supply and demand and, thus, can give rise to persistent wage differentials.

Job Contracts – A job contract refers to a condition where firms and workers agree to set their wage constant for a certain period of time. One type of job contract is an implicit contract where a firm offers risk-averse workers a steady wage, and in return pays a wage lower than it would otherwise pay (Stiglitz (1986)). Hashimoto and Yu (1980) added that even when workers are risk neutral, firms and workers can gain from making a long-term job contract.³ A long-term contract between firms and workers sets wages for a certain period of time and allows wages to be negotiated on a staggered basis (Fischer (1977) and Taylor (1979)). Firms that rely more on skilled workers can benefit more from a

³ Hashimoto and Yu (1980) describe some possible forms of long-term labor contracts, such as the regular long-term contract, quasi-long-term contract, tenure, and indenture contract. Firms and workers can agree to apply any type of contract that they see fit.

long-term contract because it gives them a period during which they can pay the same wage while workers keep acquiring more skills through on-the-job training.

In summary, whether it is an implicit contract or long-term contract, a job contract fixes wages for a certain period of time thus preventing wages from adjusting freely to the market's supply and demand forces. As a result, job contracts can also give rise to wage differentials, particularly where use of job contracts is widespread.

Labor unionization – Job contracts and efficiency wages represent a bargaining position that workers have as individuals. Workers can increase their bargaining position by forming a labor union. The union bargains on behalf of its members over the terms and conditions of employment and may strike the employers to achieve its objectives. The process to bargain with employers is called collective bargaining.

Labor unions can exercise different methods of collective bargaining, for instance through strike threat, restriction of supply, and increased labor demand.⁴ Through collective bargaining, unions and firms will agree to certain work arrangements that allow wages not to follow market wages. Whichever method the unions apply, literature agrees that in most cases unionized workers receive higher wages than nonunionized workers (Kaufman (1994), and Filer et al. (1996)).

Kalacheck and Raines (1980) claim that studies might have overestimated the effect of labor unionization on wages. They show that unionized jobs mostly consist of blue-collar jobs with a poor job environment. As a result, wage differentials between unionized and nonunionized jobs may represent not only the effect of unionization but also the effect of occupational compensating differentials. Regardless, labor unions provide ways for firms and workers to set their own wages, which would otherwise freely adjust to market supply and demand. This in turn can contribute to wage differentials.

⁴ For more complete explanation, see Kaufman (1994).

Minimum Wage Legislation – The minimum wage was initially introduced in 1939, and typically applies to nonsupervisory workers. While the share of workers with minimum wage coverage has increased over time, the minimum wage has not grown as fast as regular wages. Filer et al. (1996) show that the ratio of minimum wage to regular wage has fallen over time.

Most studies have found that a minimum wage increase reduces employment of unskilled workers (Kaufman (1994) and Filer et al. (1996)), although a few found opposing results (for example Card (1992) and Card and Krueger (1994)). Regardless, a minimum wage prevents wages from adjusting freely to the labor market, which can contribute to wage differentials.

In summary, all the factors above – efficiency wages, job contracts, labor unionization, and minimum wage legislation – allow firms (and workers) to set wages beyond the direct control of competitive market forces. This can give rise to persistent interregional wage differentials.

The extent of wage rigidity in a local labor market can depend on factors such as industrial mix, occupational mix, business sizes, and types of products. For instance, the 2003 Employment and Earnings data from the U.S. Bureau of Labor Statistics show that the unionization rate is highest in the government, communication-utilities, and transportation sectors and lowest in the finance-insurance and agricultural sectors. The same data show that the unionization rate is highest in precision production workers, operators, and professional and managerial occupations. Curme et al. (1990) confirm that the unionization rate differs across states. Similarly, Filer et al. (1996) showed that the level of minimum wage differs across states. In line with that, Campbell III and Kamlani (1997) showed that wage rigidity gives rise to wage differentials within three different groups of occupations: white-skilled, nonwhite-skilled, and unskilled occupations. Finally, Krueger and Summers (1987), Helwege (1992), and Borjas and Ramey (2000) claim that efficiency wages, unionization, and job contracts (rent sharing) give rise to interindustry wage differentials in the United States.

In summary, since industrial mix and occupational mix vary across regions, so does the extent of wage rigidity. This in turn implies that wage rigidity gives rise to wage differentials not only within a region but also across regions.

2.10. Other Factors

One may not be able to exhaust the list of possible factors that cause regional wage differentials. Besides factors mentioned above, the list can include aggregation bias, sex and racial discrimination, and slow market adjustments.

Aggregation bias occurs mainly because data sets do not allow studies to compare the exact same job across regions. Two main factors can cause aggregation bias: (1) the regions compared are too large; and (2) the job category is too broad. Studies typically are able to compare between job categories but not between specific job titles. The problem is that job categories are broad so that, across regions, the same category may have a very different mix of occupational skills. In other words, the observed regional wage differentials in aggregated data may not just represent pure wage disequilibrium, but also different occupational skills and productivity.

Studies typically compare wages by industry categorized by Standard Industrial Classification codes (SIC) or more currently by North America Industrial Code System (NAICS) codes. Technically, even the most detailed industry category, which would be based on 6 digit NAICS codes, will not completely eliminate the aggregation bias. Better data would break down those industries by occupational details. As of today, however, such data are rarely available, even at the state level. One possible way to avoid aggregation bias is by limiting the study to focus on a specific kind of job title, such as a study comparing wages of similar fast-food workers across regions. This, however, would typically require an expensive survey.

Sex and race discrimination can also cause persistent regional wage differentials. Though such practices are illegal, studies have found evidence of gender discrimination (Oaxaca (1973) as well as racial discrimination (Corcoran and Duncan (1979)).

Finally, regional wage differentials can still exist because the market has yet to respond to neutralize them. In this case, regional wage differentials would represent the transitional wage disequilibrium, which should be eliminated by migration. The process of wage equalization, however, takes time. The literature notes that spatial market adjustment can be slow due to imperfect information and transportation costs (Greenwood (1970), Bellante and Jackson (1983) and studies on the spatial mismatch hypothesis, such as Ihlanfeldt and Sjoquist (2000) and Martin (2004)). The speed of adjustment can also differ by age and race with young and blacks adjusting most rapidly (Dunlevy and Bellante (1983)).

I have discussed various factors that affect the interaction between migration and regional wage convergence. In the following sections I discuss what we can learn from the previous discussion. The goal is to answer the question of whether or not migration really promotes regional wage convergence. First, I will clarify two terms: regional wage differentials and regional wage disequilibrium. These seem interchangeable, but, as discussed in the previous sections, they are not. I then discuss how migration interacts with each of these.

2.11. Wage Differentials versus Wage Disequilibrium

Throughout the previous sections, I often mentioned regional wage differentials and wage disequilibrium as two interchangeable terms. They are, however, different concepts. The process illustrated in Figure 2.2 requires that the observed regional wage differentials indeed represent wage disequilibrium. Accordingly, such disequilibrium will induce migration. Assuming that jobs and the quality of workers are the same in both regions, any wage differences should represent wage inequalities. With zero mobility costs and perfect information, such inequalities are transitional and should be equalized by migration.

While conceptually acceptable, empirically this is hard to verify. The main reason is that the observed regional wage differentials do not necessarily represent wage disequilibrium. Not only do regional wage differentials comprise of transitional wage disequilibrium, they also comprise of many other factors including regional differences in local amenities, firms' wage policies, industrial and occupational mix, costs of living, and other job attributes such as job security, occupational risks, and wage-salary supplements. Observed regional wage differentials do not show what each of these factors contributes to differentials.

Since many components can make up the observed regional wage differentials, it is technically possible to find regional wage differentials that exist even when there is no real wage disequilibrium. Similarly, it is also possible to find zero regional wage differentials when we have real regional wage disequilibrium. For those reasons, migration can take place with or without observed regional wage differentials.

Technically, in order to identify the extent of wage disequilibrium within regional wage differentials, researchers have to control for all contributing factors. Due to the variability of the components included in wage differentials, however, as well as the difficulty of accounting for some nonmeasurable or subjective components, measuring what each contributes to regional wage differentials is difficult.

Further complication arises when taking into account migration being an individual decision. People decide to move only when they see economic disequilibrium and that adjusting to such disequilibrium would yield benefits. People, however, maximize utility rather than wages. In this context, people's utility can include job security, job prestige, opportunity for future career advancement, place amenities, as well as proximity to relatives and favorite places. In other words, for migrants, economic disequilibrium means utility disequilibrium, which is not necessarily the same as wage disequilibrium. Moreover, since utility is subjective then so is utility disequilibrium. This in turn implies that whether or not the observed regional wage differentials represent wage

disequilibrium depends also on individuals' subjective measures. Since utility differs across individuals, two identical jobs with different pay can be perceived as disequilibrium by one but as equilibrium by others.

Likewise, an equilibrium condition can also change due to personal changes. Consider a perfect condition where there is no shock in the market and two jobs in different markets receive the same equilibrium market wage. This can turn into disequilibrium because someone gets married, graduates from college, develops new preferences, or obtains new information that opens up new opportunities. Even when job attributes and wages stay the same, one may decide to migrate to find a new job or location in order to get new attributes that are more suitable with the individual's revised utility function. Graves and Linneman (1979) and Mueser and Graves (1995) discuss this as a "consumption theory" or "equilibrium theory" of migration.

In summary, while researchers know for sure that people move because of disequilibrium, empirically they cannot verify whether people move due to observed differences in wages or to differences in the other job/location attributes. This confirms the previous notion that migration can occur with or without regional wage differentials.

2.12. Migration and Regional Wage Differentials

If migration can occur with or without regional wage differentials, then it suggests that migration is not necessarily responsive to regional wage differentials. This agrees with the majority of empirical migration studies that find higher wages do not significantly attract migrants (Greenwood (1975b)). I offer three possible reasons: (1) measurement problems; (2) insignificant regional wage differentials; and (3) migrants being not necessarily interested in higher wages.

It's a problem of measurement – The wage is possibly the most important element in the individual's utility function as well as in the firms' profit function. Yet, throughout this chapter I have argued as if wages do not matter much for migrants. The fact is, while we agree on the importance of wages, empirically we have a problem verifying that. That is,

the problem lies not on the idea of the role of regional wage differentials, but on the measurement of regional wage differentials. To examine the real impact of wages on migration, one has to control for various other variables that potentially cause regional wage differentials – the *ceteris paribus* assumption of regional analysis. Given the great variety as well as the subjective measurement of those other factors, maintaining the *ceteris paribus* assumption is difficult. On top of that, data limitations can always lead to aggregation bias. In summary, as long as one cannot distinguish between regional wage differentials from wage disequilibrium, we may not be able to accurately gauge how migration responds to regional wage differentials.

Regional wage differentials are insignificant - Econometrically, studies can try to control for all those other factors by including a variety of explanatory variables. Finding the wage variable to be insignificant after controlling for those variables may suggest that the remaining regional wage disequilibrium is inconsequential. This is understandable when we look at wage as a composite variable whose value has taken into account not only the value of marginal product but also the values of other job attributes. In the regression model, this “composite” wage variable will be highly correlated with other variables measuring job attributes, leaving a strong likelihood of statistical insignificance (Gujarati (1995), Greene (2000)). In other words, after taking into account “job attribute” variables, the residual regional variation in wages is small. This suggests that the overall economy may be close to equilibrium with respect to regional wage differentials.

Migrants are not necessarily looking for higher wages – The third possible reason for a statistically insignificant wage effect is that migrants may not necessarily focus on higher wages. Migrants may look more at expected wages rather than the nominal wages. Accordingly, they are looking for places with higher probability of getting jobs (which would eventually yield a higher expected wage). Fields (1976) shows how job turnover rather than the wage significantly attracts migrants. Studies have more frequently found employment growth variables, rather than wage variables as having significant effects on migration (Mueller (1985); Greenwood (1975b); and Greenwood (1997)).

Migrants also look more at lifetime benefits rather than point-in-time benefits (Sjaastad (1962)). Accordingly, migrants might accept lower wages today in the hope of getting much higher wages in the future. Schaeffer (1985) shows that the migration decision can be part of an individual's career strategy. Migrants are willing to forego higher wages today in exchange for acquiring better human capital, with which they gain much higher wages in the future. This agrees with Greenwood (1975b) who concludes that although migrants may initially suffer losses, they will eventually gain higher wage growth than nonmigrants.

Another consideration is that migration often is a family decision, rather than an individual decision. An individual migrant might not respond to higher wages, but, as a whole, the individual's family might (Mincer (1978) and Schaeffer (1987)). Finally, some types of migration, such as return migration (Da Vanzo (1983)) and retirement migration (Clark et al. (1996)), are not inherently based on "economic" factors, but rather on other interests. Together, all of these factors may destine some regional wage differentials to appear as an inconsequential factor in typical studies of migration.

2.13. Migration and Regional Wage Disequilibrium

Let the regional wage differentials represent the true regional wage disequilibrium. In this case, following the classical theory, migration will flow from low-wage to high-wage regions, increase the labor supply and thus put downward pressure on the wage in the destination region, and in turn reduce the wage gap between origin and destination. This is true if and only if all the classical assumptions are met. It is practically impossible to meet all the classical assumptions. To the extent that those assumptions are met, migration should reduce regional wage differentials.

In reality, production functions are more likely to exhibit increasing returns to scale and workers are anything but homogenous. As people move into the destination region, migration would promote agglomeration, which in turn reinforces economic growth in the destination region. In addition, migration is more likely to select more productive people, thus further reinforcing economic growth in the destination region as well. Under

these scenarios, migration would widen the wage gap, not reduce the gap between origin and destination.

On the flip side, agglomeration will eventually stimulate an increase in property values and other costs of living and in disamenities. Over time, agglomeration economies may also give way to agglomeration diseconomies – disamenities such as congestion and crime. In the long run, the net benefits of agglomeration will eventually be exhausted, leading people and business to start migrating out of the region, as other regions now become relatively more interesting (Knapp and Graves (1989)). This may explain phenomena such as the reversal of the rural-urban and South-North migration patterns in the United States during the latter part of the 20th century (Greenwood (1997), and Greenwood and Hunt (1984)).

In summary, the answer to the question of whether migration would reduce or increase regional wage differentials depends on which of the above opposing scenarios dominates. Like this chapter, the migration literature addresses this question conceptually, though in bits and pieces, as opposed to this chapter's comprehensive perspective. The literature has not adequately addressed this issue, empirically. As discussed in this chapter, part of the empirical problem may be the difficulty in measuring the true regional wage disequilibrium. For example, the debate on whether migration follows an equilibrium versus a disequilibrium process roots in the disagreement in interpreting observed regional wage differentials. Evans (1990) claims regional wage differentials represent more of wage disequilibrium while Knapp and Graves (1989) and Graves and Mueser (1993) claim otherwise. Another reason for the difficulty in resolving this issue empirically may be that migration connects origin and destination regions in much more complex ways. I will discuss this issue in the next and final section of this chapter.

2.14. Future Directions

Note two facts on migration. First, data on domestic migration in the United States show that large outmigration frequently coincides with large immigration. That is, a large migration flow from region A to region B often coincides with a large flow from B to A.

Second, migration flows between prosperous regions are more dominant than flows from depressed to prosperous regions (Plane (2004)). This suggests that migration works more as a back-and-forth flow rather than a one-way or a one-point-in-time flow. Accordingly, the impact of migration on the involved regions may also go back and forth, with the current inflow affecting the reverse flow. It takes time for this process to play itself out. For these reasons, analyzing migration's effect on regional wage differentials with a model based on one-way or one-time flows could yield incorrect results. A better analysis should apply dynamic analyses, allowing researchers to take into account the dynamic nature of migration.

An appropriate model used to analyze how migration affects regional wage differentials might be an aggregate model equipped with tools for dynamic analyses. An even better model would be an individual migration model (e.g. a conditional logit model), which allows researchers to analyze migration as an individual, rather than aggregate, decision. This model also allows researchers not to worry about the possibly complex simultaneous or dynamic impact of migration. Finally, it also allows researchers to directly control for individual characteristics, thus takes into account some variations associated with individual's subjectivity.

Whether it is an aggregate or an individual model, an analysis of migration and regional wage differentials will always be subject to measurement problems, including aggregation bias and the difficulty in representing regional wage disequilibrium. One possible way to get around this problem would be to limit the scope of the study, focusing more on a specific labor market. This would greatly help researchers to control for various other factors that make up regional wage differentials. Studies focusing on, for instance, regional wage differentials in the fast food industry or grocery stores could be more meaningful than studies on an aggregated industry such as the whole services sector.

Comparing across industries based on their closeness to a perfectly competitive market would also help to see how much deviations from the classical assumptions change the

expected impact of migration on equilibrating regional wage disequilibrium. For instance, one can compare migration's effect on wages in the retail sales sector, which may most closely represent a perfectly competitive market, with migration's effect on manufacturing wages, which rely more heavily on increasing returns to scale operation.

Broadening out even further, it may be informative to compare migration between seemingly similar regions with migration between contrasting regions. Such analysis could tell whether migration occurs more between similar rather than contrasting regions, and how the similarity of regions affects the degree of back-and-forth flows. One could also look at the degree and speed of wage convergence between similar versus contrasting regions and between pairs of regions with significant back-and-forth flows versus pairs of regions dominated by one-way flows. From these many options, the studies detailed in Chapters 3 and 4 rely on models of individual migration. The model used in Chapter 3 also relies on significant detail by industry.

References

- Alonso, W. 1971. "Equity and Its Relation to Efficiency in Urbanization." in J.F. Kain and J.R. Meyer (eds). *Essays in Regional Economics*. Cambridge, MA: Harvard University Press.
- Allen, Jeremiah M. and Curtis. B. Eaton. 2005. "Incomplete Information and Migration: The Grass Is Greener across the Higher Fence." *Journal of Regional Science* 45: 1-19.
- Armstrong, Harvey and Jim Taylor. 2000. *Regional Economics and Policy*. Malden, MA: Blackwell.
- Arthur, W. Brian. 1990. "Positive Feedbacks in the Economy." *Scientific American* 262: 92-99.
- Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics* 106: 407-443.
- Barro, Robert J. and Xavier Sala-i-Martin. 1991. "Convergence Across States and Regions." *Brookings Papers on Economic Activity* 1: 107-182.
- Bedau, Klaus-Dietrich. 1998. "Population Trends, Employment Levels, Economic Performance, and Income Evolution in East and West Germany since Unification." *Journal of Income Distribution* 8: 207-23.
- Bellante, Don. 1978. "Changes in the Migration of Human Capital and the Southern "Brain Drain": Evidence from the Census of 1960 and 1970." *Journal of Behavioral Economics* 7: 53-77.
- Blau, Peter and Otis D. Duncan. 1967. *American Occupational Structure*. New York: Wiley.
- Borjas, George. 1992. "Self-Selection and Internal Migration in the United States." *Journal of Urban Economics* 32: 159-185.
- Borjas, George and Valerie A. Ramey. 2000. "Market Response to Inter-industry Wage Differentials." *National Bureau of Economic Research Working Paper*, 7799.
- Borts, G. H. 1960. "The Equalization of Returns and Regional Economic Growth." *American Economic Review* 50: 319-347.
- Brown, Charles and James Medoff. 1989. "The Employer Size Wage Effect." *Journal of Political Economy* 97: 1027-1059.

- Campbell III, Carl M. and Kunal S. Kamalani. 1997. "The Reasons for Wage Rigidity: Evidence From a Survey of Firms." *The Quarterly Journal of Economics* 112: 759-789.
- Card, David. 1992. "Do Minimum Wages Reduce Employment? a Case Study of California, 1987-1989." *Industrial and Labor Relations Review* 46: 38-54.
- Card, David and Alan Krueger. 1994. "Minimum Wages and Unemployment: a Case Study of the Fast Food Industry in New Jersey and California." *The American Economic Review* 84: 772-793.
- Card, J. David and E. DiNardo. 2002. "Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *NBER Working Paper No. 8769*.
- Card, David, Thomas Lemieux, and Craig W. Riddell. 2003. "Unionization and Wage Inequality: A Comparative Study of the U.S, the U.K., and Canada." *National Bureau of Economic Research Working Papers: 9473*.
- Clark, David E., Thomas A. Knapp, Nancy E. White. 1996. "Personal and Location-Specific Characteristics and Elderly Interstate Migration." *Growth and Change* 27: 327-51.
- Corcoran, Mary and Greg J. Duncan. 1979. "Work History, Labor Force Attachment, and Earnings Differences between the Races and Sexes." *Journal of Human Resources* 14: 3-20.
- Crown, William. 1991. "Migration and Regional Economic Growth: Origin-Destination Model." *Economic Development Quarterly* 5: 45-49.
- Curme, Michael, B. Hirsch, and D. Macpherson. 1990. "Union Membership and Contract Coverage in the United States, 1983-1988." *Industrial and Labor Relations Review* 44.
- Cushing, Brian. 1987. "Location-Specific Amenities, Topography, and Population Migration." *Annals of Regional Science* 21: 74-85.
- Da Vanzo, Julie. 1983. "Repeat Migration in the United States: Who Moves Back and Who Moves on? Evidence from Micro Data." *Review of Economic Statistics* 65: 552-559.
- Dickie, Mark and Shelby Gerking. 1989. "Interregional Wage Differentials in the United States: A Survey." in van Dijk et al. (ed.) *Migration and Labor Market Adjustment*. Dordrecht, The Netherlands: Kluwer Academic.
- Dixit A. K. and V. Norman. 1980. *Theory of International Trade*. Cambridge, Britain: Cambridge University Press.

- Dunlevy, James A. and Don Bellante. 1983. "Net Migration, Endogenous Incomes and the Speed of Adjustment to the North-South Differential." *Review of Economics and Statistics* 65: 66-75.
- Dunn, Lucia F. 1984. "The Effects of the Firm Size on Wages, Fringe Benefits, and Worker Disutility." in H. Goldschmid (ed.) *The Impact of the Modern Corporation*. New York: Columbia University Press.
- Easterlin, R. 1961. "Regional Income Trends, 1840-1950." in S. Harris (ed.) *American Economic History*. New York: McGraw-Hill.
- Evans, Alan. 1990. "The Assumption of Equilibrium in the Analysis of Migration and Interregional Differences: A Review of Some Recent Research." *Journal of Regional Science* 30: 515-531.
- Fields, Gary. 1976. "Labor Force Migration, Unemployment and Job Turnover." *The Review of Economic Statistics* 58: 407-415.
- Filer, Randall K., Daniel S. Hammermesh and Albert E. Rees. 1996. *The Economics of Work and Pay, Sixth Edition*. New York: Harper Collins.
- Fischer, Stanley. 1977. "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule." *Journal of Political Economy* 85: 191-205.
- Goetz, Stephan J. 1999. "Migration and Local Labor Market." *Web-Book of Regional Science*. Regional Research Institute, West Virginia University
- Goetz, Stephan J. and David L. Debertin. 1996. "Rural Population Decline in the 1980s: Impacts of Farm Structure and Federal Farm Programs." *American Journal of Agricultural Economics* 78: 517-529.
- Gramlich, Edward M., Deborah Laren, and Naomi Sealand. 1992. "Moving Into and Out of Poor Urban Areas." *Journal of Policy Analysis and Management* 11: 273-287.
- Graves, Philip and Peter Linneman. 1979. "Household Migration: Theoretical and Empirical Results." *Journal of Urban Economics* 6: 383-404.
- Graves, Philip and Peter Mueser. 1993. "The Role of Equilibrium and Disequilibrium in Modeling Regional Growth and Decline: A Critical Reassessment." *Journal of Regional Science* 33: 69-84.
- Greenwood, Michael. 1969. "An Analysis of the Determinants of Geographic Labor Mobility in the United States." *Review of Economics and Statistics* 51: 189-194.

- Greenwood, Michael. 1970. "Lagged Response in the Decision to Migrate." *Journal of Regional Science* 10: 375-84.
- Greenwood, Michael. 1975a. "A Simultaneous-Equations Model of Urban Growth and Migration." *Journal of the American Statistical Association* 70: 797-810.
- Greenwood, Michael. 1975b. "Research on Internal Migration in the United States: A Survey." *Journal of Economic Literature* 13: 397-433.
- Greenwood, Michael. 1981. *Migration and Economic Growth in the United States*. New York: Academic Press.
- Greenwood, Michael. 1997. "Internal Migration in Developed Countries." in Rosenzweig MR and O. Stark (eds.) *Handbook of Population and Family Economics* 1b. Amsterdam: Elsevier.
- Greenwood, Michael, James Chalmers, and Philip Graves. 1989. "Regional Location Patterns in the United States: Recent Changes and Future Prospects." in Van Dijk et al. (eds.) *Migration and Labor Market Adjustments*. Dordrecht: Kluwer Academic.
- Greenwood, Michael and Gary Hunt. 1984. "Migration and Interregional Employment Distribution in the United States." *The American Economic Review* 74: 957-969.
- Gujarati, Damodar N. 1995. *Basic Econometrics, Third Edition*. New York: McGraw-Hill.
- Hammond, George. 2004. "A Time Series Analysis of U.S. Metropolitan and Nonmetropolitan Income Divergence." *Bureau of Business and Economic Research Working Paper: 2004-02*. West Virginia University.
- Hanna, F. A. 1959. *State Income Differentials, 1914-1954*. Durham, NC: Duke University Press.
- Hashimoto, Masanori and Ben T. Yu. 1980. "Specific Capital, Employment Contracts and Wage Rigidity." *The Bell Journal of Economics* 11: 536-549.
- Helwege, Jean. 1992. "Sectoral Shifts and Interindustry Wage Differentials." *Journal of Labor Economics* 10: 55-84.
- Hirschman, Albert O. 1958. *The Strategy of Economic Development*. New Haven, CT: Yale University Press.
- Hoover and Giarratani. 1999. "Introduction to Regional Economics." *Web-Book of Regional Science*. Regional Research Institute, West Virginia University.

- Hunt, Gary. 1994. "Equilibrium and Disequilibrium in Migration Modeling." *Regional Studies* 27: 341-349.
- Ihlanfeldt, Keith R. and David L. Sjoquist. 2000. "The Geographic Mismatch between Jobs and Housing." in *Multi-City Study of Urban Inequality Series*. New York: Russell Sage Foundation.
- Isserman, Andrew. 2005. *Lecture notes for Advanced Regional Economics Class, Spring 2005, and West Virginia University*. Unpublished.
- Kalachek, Edward and Fredric Raines. 1980. "Trade Unions and Hiring Standards." *Journal of Labor Research* 1: 63-75.
- Katz, Lawrence. 1986. "The Efficiency Wage Theories: A Partial Equilibrium." in Stanley Fischer (ed.) *NBER Macroeconomics Annual 1986*. Cambridge, MA: MIT Press.
- Kaufman, Bruce. 1994. *The Economics of Labor Markets: Fourth Edition*. Orlando, FL: The Dryden Press.
- Knapp, Thomas A and Philip E. Graves. 1989. "On the Role of Amenities in Models of Migration and Regional Development." *Journal of Regional Science* 29: 71-87.
- Krueger, Alan and Lawrence Summers. 1987. "Reflections on the Inter-industry Wage Structure." in Kevin Lang and Jonathan (eds.) *Unemployment and the Structure of Labor Markets*. London: Basil Backwell.
- Krugman, Paul. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99: 483-499.
- Krugman, Paul R. and Maurice Obstfeld. 1994. *International Economics: Theory and Policy, Third Edition*. New York: Harper Collins.
- Lowry, I.S. 1966. *Migration and Metropolitan Growth: Two Analytical Models*. San Francisco: Chandra Publishing Co.
- Martin, Richard W. 2004. "Spatial Mismatch and the Structure of American Metropolitan Areas, 1970-2000." *Journal of Regional Science* 44: 467-88.
- Mincer, Jacob. 1978. "Family Migration Decisions." *Journal of Political Economy* 86: 749-773.
- Mortensen, Dale. 2003. *Wage Dispersion: Why Are Similar Workers Paid Differently?* Cambridge, MA: MIT Press.

- Mueller, F. Charles. 1985. *The Economics of Labor Migration, A Behavioral Analysis*. New York: Academic Press.
- Mueser, Peter and Philip Graves. 1995. "Examining the Role of Economic Opportunity and Amenities in Explaining Population Redistribution," *Journal of Urban Economics* 37: 1-25.
- Muth, Richard F. 1971. "Migration: Chicken or Egg?" *Southern Economic Journal* 37: 295-306.
- Myrdal, Gunnar. 1957. *The Economic Theory of and Under-developed Regions*. London: Duckworth.
- Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14: 693-709.
- O'Sullivan, Arthur. 1995. *Urban Economics, Third Edition*. Boston: McGraw-Hill.
- Plane, David. 2004. "Stream, Flows, Currents, Cycles, and Regimes: Proposition on the Spatiotemporal Nature of Migration Systems." *International Journal of Environmental Creation* 4: 60-73.
- Plane, David and Peter Rogerson. 1994. *The Geographical Analysis of Population*. New York: John Wiley and Sons.
- Quah, Danny. 1993. "Galton's Fallacy and Tests of the Convergence Hypothesis." *C.E.P.R. Discussion Papers*: 820.
- Razin, Assaf and Efraim Sadka. 1995. "Resisting Migration: Wage Rigidity and Income Distribution." *The American Economic Review* 85: 312-316.
- Rosen, Sherwin. 1986. "The Theory of Equalizing Differences." in Orley Ashenfelter and Richard Layard (eds.) *Handbook of Labor Economics* 1. Amsterdam: North Holland.
- Schachter, Joseph and Paul G. Althaus. 1989. "An Equilibrium Model of Gross Migration." *Journal of Regional Science* 29: 143-59
- Schaeffer, Peter V. 1985. "Human Capital Accumulation and Job Mobility." *Journal of Regional Science* 25: 103-114.
- Schaeffer, Peter V. 1987. "A Family Model of Migration." *Socio-Economic Planning Sciences* 21: 263-269.
- Schaeffer, Peter V. 2005. "Human Capital, Migration Strategy, and the Brain Drain." *Journal of International Trade and Economic Development* 14: 319-335.

- Schaeffer, Peter V. and Tesfa G. Gebremedhin. 2004. "Labor Market Size and Unemployment Duration: A Theoretical Note." *Regional Research Institute, West Virginia University Working Paper* 2004-8.
- Schlottmann, A. and A. Herzog. 1984. "Career and Geographic Mobility Interactions for the Age Selectivity of Migration." *Journal of Human Resources* 19: 72-86.
- Schwartz, Aba. 1973. "Interpreting the Effect of Distance on Migration." *The Journal of Political Economy* 81: 1153-1169.
- Schwartz, Aba. 1976. "Migration, Age, and Education." *The Journal of Political Economy* 84: 701-720.
- Sjaastad, A. Larry. 1962. "The Costs and Returns of Human Migration." *The Journal of Political Economy* 70: 80-93.
- Stark, Oded, Christian Helmenstein, and Alexia Prskawetz. 1997. "A Brain Gain with a Brain Drain." *Institute for Advanced Studies, Economics Series* 45.
- Stiglitz, Joseph E. 1986. "Theories of Wage Rigidity." in *Keynes' Economic Legacy: Contemporary Economic theories*. Butkiewicz James L., Kenneth J. Koford, and Jeffrey B. Miller (eds.): 153-206. New York: Praeger Publishers.
- Taylor, John B. 1979. "Staggered Wage Setting in a Macro Model." *The American Economic Review* 69: 108-113.
- Vanderkamp, John. 1976. "The Role of Population Size in Migration Studies." *The Canadian Journal of Economics* 19: 508-517.
- Watanabe, S. 1969. "The Brain Drain from Developing to Developed Countries." *International Labour Review* 99: 401-33.
- Wojan, Timothy R. 2000. "The Composition of Rural Employment Growth in the New Economy." *American Journal of Agricultural Economics* 82: 594-605.

Chapter 3

The Interaction between Individuals' Destination Choice and Occupational Choice A Simultaneous Equation Approach

3.1. Introduction

Population migration is the primary determinant of regional population change, dwarfing the small net natural change. Between 1995 and 2000, over 22 million individuals moved across states, representing over 8.5 percent of the U.S. population 5 years old and above. A change of this magnitude has a large impact. For example, migration affects regions' age composition, businesses' location decisions, wage rates, state and local tax revenues, public service allocations, and industrial clustering. Perhaps migration plays its most important role by increasing economic efficiency through allowing the sorting of workers and, indirectly, businesses, as they search for the best fit.

In sorting among labor markets, individuals' occupational skills are a major factor in determining which labor market region fits best. All else equal, utility maximizing individuals are more likely to choose a labor market region in which they can maximize the returns to their occupational skills. The relationship between choice of region and occupation is complex since individuals may change not only region, but also occupation. Individuals could change their occupation based on their choice of a region or *vice versa*, or make both choices simultaneously (e.g., Schaeffer (1985)).

This study examines the relationship between individuals' occupation choice and destination choice. It portrays the relationship as an interaction between the supply of occupational skills by individuals and the demand by different labor market regions. A region's demand is affected by its industrial mix and employment growth by industrial sector. Unlike previous studies, this study recognizes a simultaneous relationship between an individual's occupational and destination choice, therefore, applies two-stage maximum likelihood estimation. The simultaneous system has two equations: (1) a multinomial logit model of occupational choice and (2) a conditional logit model of state destination choice. The unusual merger of multinomial *logit* and conditional *logit* models

in a simultaneous equation framework requires derivation of a unique variance-covariance matrix; otherwise, statistical tests will be incorrect (Greene (2000)).

3.2. Literature Review

3.2.1. Aggregate Model and Individual Model of Migration

Sjaastad (1962) noted the difficulty in explaining the reasons rational people moved to supposedly wrong places, i.e., places economically worse than their origins. Most traditional aggregate models, which assume the same utility function for all individuals, cannot satisfactorily explain it. One of the most frequently noted reasons is that the aggregate models ignore individual factors. Due to differences in individuals' characteristics such as their education, marital status, proximity to home, and occupation, their preference for a destination place can differ even though their states of origin are the same. In other words, different individuals could place different relative value on each place attribute (Train (1986)).

In addition, an individual's current destination is not always his ultimate destination. Da Vanzo (1983) found that those who have moved before are more likely to move than those who have never migrated. First moves, she argued, might fail, thus individuals will move on or return to where they came from. It is likely that moving itself is a learning process, so that repeat-moves become less costly or more effective. Zimmer (1971) found that repeat migrants were more likely in high-skilled occupations (professional and managerial occupations), than first-time movers or nonmigrants. Schaeffer (1985) added that repeat moves might simply be part of a career strategy. Individuals might choose an economically worse state because it provides the best chance for skills advancement. After mastering important skills, they explore other destinations in order to maximize the returns to their skills.

More current migration studies tend to apply individual rather than aggregate migration models. The typical model applied is the conditional logit model, including the early study by Mueller (1985) as well as the later study by Davies, Greenwood, and Li (2001). However, few have incorporated individual factors in the model. This study uses

Mueller's model as its starting point. In this study, the individual's occupational choice is one explanatory variable. Thus, the model examines which destination an individual chooses considering his or her occupational choice.

3.2.2. Migration and Occupational Change

Few studies have examined the relationship between occupation and migration. Some have focused on the relationship between occupational-mobility and migration, and mostly agreed that migration more likely leads to upward occupational mobility (Blau and Duncan (1967), Chattopadhyay (1998), to a certain extent Odland (1988), and Schaeffer (1985)). Other studies have looked at how the migration rate relates to the rate of occupational change. Most studies (Schroeder (1976) and Wilson (1985)) found no evidence of an inverse relationship suggested by Gleave and Palmer (1977), although Schlottmann and Herzog (1984) found that individuals making a career change have a higher likelihood of migrating.

A couple studies have examined whether individuals who simultaneously changed occupation and migrated achieved higher earnings. Krieg (1997) found little evidence of this. Bartel (1979) concluded that job-transfer initiated moves led to higher earnings, while other kinds of job-change initiated moves did not.

A few scholars have studied how an individual's occupation affects migration decisions. Most focused on the migration behavior of skilled individuals (Gani and Ward (1995), Pashigian (1979), Comay (1972)), although some of them looked at both skilled and unskilled workers (Ellis (1993) and Kleiner (1976)). Ellis and Kleiner concluded that individuals in high-skill occupations are more likely to move and more willing to move long distances than those in unskilled occupations (for a theoretical explanation see Schaeffer (1985)).

A major shortcoming of this last group of studies is that they disaggregated the sample by occupation category. This disaggregation raises two main problems. First, it assumes that individuals' occupations do not change, whereas unlike sex or race, individuals'

occupations can change during the migration period. Secondly, individuals likely make decisions on occupational choice and destination choice simultaneously. Several studies (e.g., Bartel (1979); Krieg (1997); and Schroeder (1976)) found that a significant number of migrants indeed changed their occupation when they migrated.

3.2.3. Modeling Simultaneous Discrete Choice Models

Simultaneity means individuals do not face two separate sets of occupational choice and destination choice, but a set of joint choices. Under this setting, variables of occupational and destination choice become stochastic and failure to account for this property could yield incorrect statistical tests (Greene (2000)).

Schmidt and Strauss (1975b) offered a method to estimate two simultaneous dependent qualitative variables. Their method operates by jointly solving the probability functions for the location and the occupational choice, respectively. The symmetry of the two probability functions allows the method to generate a combined likelihood function that incorporates both functions, so that unbiased and consistent parameter estimates can be obtained simply by maximizing the likelihood function.

Although our model is also dealing with two probabilistic discrete choice functions, we cannot apply the Schmidt and Strauss method because this study includes two different kinds of discrete choice models. We simultaneously use a multinomial *logit* model and conditional *logit* model for which the symmetry condition does not hold. The multinomial *logit* model estimates individuals' choice based on individual characteristics, whereas the conditional *logit* model estimates the choice based on the choice characteristics (place and occupational). The Schmidt and Strauss method was applicable because both models were multinomial *logit* models, which is not the case here.

This study applies two-step maximum likelihood estimation, where the conditional *logit* model of destination choice is estimated using the predicted values of the occupational choice generated by the multinomial *logit* model of occupational choice. Such a method requires that the variance-covariance matrix of the parameter estimates be adjusted. This

study applies the method developed by Murphy and Topel (1985), in Greene (2000. p. 134), to compute the adjustments.

3.3. The Basic Models

This study defines that a two-equation simultaneous system: one equation for occupational choice; the other for destination choice. The two equations take the following general forms:

$$P_{ik} = f(X_{1i}, X_{2i}, P_{ij}) \quad (1)$$

$$P_{ij} = g(Z_i, X_{1i}, P_{ik}) \quad (2)$$

Equation (1) represents the occupational-choice model. P_{ik} measures the probability of individual i choosing occupation k , which is a function of both the individual's characteristics (X_{1i} and X_{2i}) as well as the individual's destination choice P_{ij} . Equation (2) represents the destination choice model, where P_{ij} measures the probability of individual i choosing state j , which is a function of not only the place attributes Z_i and the individual characteristics X_{1i} , but also the individual's occupational choice P_{ik} .

To solve the system, this study first regresses P_{ik} on all predetermined variables to obtain the predicted values of P_{ik} . Then, using these predicted values, $\hat{P}_{ik}$, in place of P_{ik} the study regresses equation (2). Due to the simultaneity in this two-stage procedure, the estimated standard errors of the parameter estimates must be adjusted. The adjustment is described in Appendix 3.1.

Multinomial Logit Model of Occupational Choice - To estimate equation (1), I apply the typical multinomial-logit (MNL) model of occupational choice as used by Schmidt and Straus (1985). This study extends the model by incorporating variables related to migration. Unlike typical MNL models, the individual's occupational choice is not a once-in-a-lifetime decision. Instead, it is a one-event decision, a decision made in the event of choosing a destination. In the initial period, the study assumes that individuals (except new entrants) have made their occupational choices. Now, facing a destination

choice decision, they have to choose their occupation again, whether to stay or to move to another occupation. For that reason the model includes previous occupation, representing previous choice of occupation, as an explanatory variable.

To represent different types of occupation skills, this study uses occupation categories defined by the U.S. Bureau of the Census 1980 Standard Occupational Code, but reclassifies them into 5 categories:

1. Managerial and Professional Occupations
2. Sales Occupations
3. Operator Occupations
4. Laborer Occupations
5. Other occupations: Production-Precision, Clerical, and Service Worker Occupations.

Let the utility of choosing occupation k for individual i be:

$$U_{ik} = \beta_k' \mathbf{X}_i + \varepsilon_{ik} \quad \text{where } i = 1, 2, \dots, N, \text{ and } k = 1, 2, \dots, 5. \quad (3)$$

The subscript k indexes the types of occupation in the choice set, ranging from 1 to 5, and variable $\mathbf{X}$ represents the individual characteristics. Assuming that the disturbances ε_{ik} are independent and identically distributed with extreme value distribution (McFadden (1973) and Greene (2000)), the probability of individual i choosing occupation k can be represented as:

$$P(Y=k) = P_{ik} = \frac{e^{\beta_k' \mathbf{X}_i}}{\sum_k e^{\beta_k' \mathbf{X}_i}} \quad k = 1, \dots, 5. \quad (4)$$

The model's log likelihood function is:

$$\text{Log } L_{MNL} = \sum_i \sum_k d_{ik} \text{Log } P_{ik} \quad \text{where } d_{ik} = 1 \text{ if the occupation is chosen, and } 0$$

otherwise. Applying maximum likelihood estimation yields parameter estimates, $\hat{\beta}_k$.

Using the values of $\hat{\beta}_k$ and the fact that $\sum_k P_k = 1$, I can compute each of the five P_k s.

Conditional Logit Model of State Destination Choice - To estimate the second model, the state destination choice model, this study applies the conditional *logit* model, as used by Mueller (1985). The choice set consists of all possible state destinations. The analysis excludes Hawaii and Alaska, and combines the District of Columbia and Maryland. Thus, each individual chooses among 48 possible destinations, including his or her current state. Migration is defined as a change of residence during the beginning (1993) and ending period (1998).

Let the utility of choosing state j for individual i be:

$$U_{ij} = \alpha' \mathbf{Z}_{ij} + \varepsilon_{ij} \quad i = 1, 2, \dots, N, \text{ and } j = 1, 2, \dots, 48. \quad (5)$$

The subscript j indexes the states in the choice set, and variable $\mathbf{Z}$ represents the place attributes and individual characteristics. Assuming the disturbances ε_{ij} are independent and identically distributed with *Weibull* distribution, the probability of individual i choosing state j can specifically be represented as:

$$P(Y=j) = P_{ij} = \frac{e^{\alpha' \mathbf{Z}_{ij}}}{\sum_j e^{\alpha' \mathbf{Z}_{ij}}} \quad i = 1, 2, \dots, N, \text{ and } j = 1, 2, \dots, 48. \quad (6)$$

In order to incorporate the individual characteristics, including $\hat{\mathbf{P}}_k$, the model creates interaction terms between the individual characteristics and place attributes. In this case, the complete conditional *logit* model is:

$$P(Y=j) = P_{ij} = \frac{e^{\bar{\alpha}' \bar{\mathbf{Z}}_{ij} + \tilde{\alpha}' \tilde{\mathbf{Z}}_{ij} * \mathbf{X}_i + \overset{\circ}{\alpha}' \overset{\circ}{\mathbf{Z}}_{ij} * \hat{\mathbf{P}}_k}}{\sum_j e^{\bar{\alpha}' \bar{\mathbf{Z}}_{ij} + \tilde{\alpha}' \tilde{\mathbf{Z}}_{ij} * \mathbf{X}_i + \overset{\circ}{\alpha}' \overset{\circ}{\mathbf{Z}}_{ij} * \hat{\mathbf{P}}_k}} \quad \text{where } i = 1, \dots, N, \text{ and } j = 1, \dots, 48. \quad (7)$$

Note that $\mathbf{Z}$ consists of $\bar{\mathbf{Z}}$, $\tilde{\mathbf{Z}} * \mathbf{X}$ and $\overset{\circ}{\mathbf{Z}} * \hat{\mathbf{P}}_k$. Of all the state-attribute variables in $\mathbf{Z}$, only $\bar{\mathbf{Z}}$ do not interact with any of the individual characteristics. Other place attributes $\overset{\circ}{\mathbf{Z}}$ interact

with the predicted probabilities of occupational choice, $\hat{\mathbf{P}}_k$, and $\tilde{\mathbf{Z}}$ interact with the other individual characteristic variables, $\mathbf{X}$. This study employs certain individual characteristics to explain individuals' likelihood to choose between origin versus non-origin and between close versus distant states, including education, age, race, marital status, family size, and a dummy for the presence of a working spouse. In addition, the study employs two spatially related individual characteristics, house-tenure and the frequency of previous interstate moves, which captures the greater propensity for previous migrants to move again. In equation (7) this effect is shown as the interactions between $\tilde{\mathbf{Z}}$, which include a distance variable, a non-origin dummy variable, and $\mathbf{X}$, representing individuals' education, age, race, marital status, family size, etc. Including those individual characteristics should enable the model to explain aspects of migration choices, such as moving vs. staying or short distance vs. long distance, that are not explained by destination attributes.

The main focus of this essay is on the interactions between $\tilde{\mathbf{Z}}$ and $\hat{\mathbf{P}}_k$. This study hypothesizes that utility maximizing individuals will choose a destination where they can maximize the returns to their occupational skills. The labor market region with the strongest demand for their skills promises the highest returns to the skills. To indicate the strength of the state's demand for certain occupational skills, this study uses the state's industrial mix. Goetz (1999) and Schmidt and Strauss (1975a) showed that industrial sectors correlate strongly with types of occupation.

Tables 3.1 through 3.4 illustrate the patterns of association between types of occupation and industrial sectors. More specifically, they show that each state exhibits similar patterns of association. The patterns demonstrate that certain industrial sectors dominantly absorbed certain types of occupations. More particularly, they show that professional and managerial occupations are dominantly absorbed by the service sector, sales occupations by the retail sector, and operator occupations by the manufacturing sector. Some other occupations are dominated by two or a few sectors. On the other

hand, no sector seems to dominantly absorb laborer occupations, suggesting the demand for laborer employment is widely dispersed across industrial sectors.

Those patterns should represent the association between the supply of occupational skills from the individuals and the demand from the state. According to these particular patterns, utility-maximizing individuals choosing professional or managerial skills will more likely choose a state with strong service sector demand. Similarly, individuals with operator occupation skills will more likely choose a state with strong manufacturing sector demand.

To measure the strength of demand by sector, the analysis could use either the size of employment in the sector or the employment growth in that sector. The employment size of the sector, however, correlates strongly with the size of the state. In other words, big states tend to always have greater total employment compared with small states. Thus, still controlling for the employment size, this study uses employment growth by sector in each state as a measure of the strength of the state's demand for certain occupations.

This means the variable $\hat{\mathbf{Z}}^{\circ} \hat{\mathbf{P}}_k$ in equation (7) represents the interaction between employment growth (by sector) and predicted probability of choosing certain occupations.

3.4. Data

The individual data come from the 1993 and 1998 panels of the National Longitudinal Survey of Youth (NLSY) 1979. The 1993 data represent initial characteristics, while the 1998 data represent end-of-period characteristics. Typically individual data would come from the Census PUMS (Public Use Microdata Sample) or CPS (Current Population Survey), which have larger samples. However, they do not have information on the individuals' previous occupation, which is crucial for this study. The NLSY data has 6,359 observations, after excluding individuals enrolled in school, out of the labor force, or with incomplete records. Data on state attributes come from several sources. Employment data come from the Regional Economic Information System. Other state

attributes such as distance, temperature, state adjacencies, and topography can be obtained from U.S. Statistical Abstract and other standard government sources.

3.5. Empirical Results

3.5.1. Multinomial *Logit* Model of Occupational Choice

Table 3.5 presents the reduced from regression results of the multinomial *logit* model of occupational choice, which includes all predetermined variables in the system. Since place attributes do not originally belong to the MNL model, the table shows only the parameter estimates of the relevant individual characteristics. Although this study focuses more on an individual's migration behavior, it is useful to see whether the occupational choice model yields reasonable results.

The result shows that previous occupations strongly determine the individual's current occupational choice. Individuals are more likely to choose the same occupation as they had in the initial period. Education also significantly affects occupation choice. The more educated an individual the more likely she or he is to choose professional & managerial occupations and less likely to choose operator and laborer occupations.

Interestingly, race significantly affects the choice of professional-managerial or sales occupations. Being white increases the likelihood of choosing professional-managerial or sales occupations. Race does not significantly affect the likelihood of choosing an operator or laborer occupation. On the other hand, being male increases the likelihood of choosing an operator or laborer occupation, but does not significantly affect the choice of professional-managerial or sales occupations.

3.5.2. Conditional *Logit* Model of State Destination Choice

Table 3.6 presents the results of the conditional *logit* model of state destination choice. The results, in most cases, agree with those of previous migration studies. The first set of variables interact the dummy of non-origin with individual characteristics: Age, White, House Tenure, Male, Have a working spouse, Family size, and Frequency of previous interstate moves. These interaction terms represent the effect of those individual

characteristics on the likelihood of choosing non-origin states, or in other words, the likelihood of moving. Age, House tenure, and Frequency of previous interstate moves significantly affect the likelihood of moving. Consistent with the migration literature, the older an individual and the longer an individual resides in a given house (stays in the origin) the less likely that he or she moves to another state. Similarly, the more frequently an individual has made interstate moves previously, the more likely he or she will move again. Other individual characteristics do not significantly affect the likelihood of moving.

Another variable interacts educational attainment and moving distance. This term tests whether, as suggested, the higher the education, the more willing to move long distance. The parameter estimate is positive, as hypothesized, but of borderline statistical significance (with a t-value of 1.6).

All parameter estimates for place attribute variables are significant, except January temperature. Consistent with the migration literature, the results indicate that people are more likely to choose states with large employment size (Share), destinations that are closer (Distance), adjacent states (Adjacent), and warmer climates (July temperature). Surprisingly, however, the results show that individuals are more likely to choose states with higher humidity. This finding differs from the conventional wisdom that people are attracted to places with low humidity.

The focus of this study is on the parameter estimates of the last interaction terms. The last three variables represent interactions between employment growth of certain sectors and the individual's predicted probability of choosing certain occupations. As discussed, this study perceives that some pairs of interaction (e.g. professional-managerial and services sector; sales and retail sector; operator and manufacturing sector; and laborer and overall sectors) show strong association between supply of and demand for occupational skills. The first of the three variables (*Emp-Growth*Pk_Strong Association*) represents the collection of these pairs that have a strong association. To check for robustness of results, the model also tests those pairs with other degree of associations. For that reason,

the model also includes *Emp-Growth*Pk_Medium Association*, which represents the collection of the pairs with medium degree of association. This collection includes the pairings of professional-managerial and manufacturing sector; operator and services sector; laborer and retail sector; and laborer and manufacturing sector. Finally, for the same reason, the model includes a variable representing the collection of pairs with weak association (*Emp-Growth*Pk_Weak Association*). This collection includes other possible pairs that are not included in the first two variables.

Table 3.6 shows that the parameter estimate for the collection of pairs with strong association is positive and significant, of pairs with medium association is insignificant, and of pairs with weak association is negative and significant. This result confirms the study hypothesis, finding that individuals are indeed more likely to choose states where the demand for their occupational skill is strong. The result also suggests that individuals are less likely to choose states where the demand for their occupational skill is weak.

Table 3.6 also shows how the regression result is affected by the variance-covariance adjustment required due to the simultaneity problem. It shows, however, that the adjusted standard errors are almost the same as the non-adjusted standard errors. Thus, the results of significance tests are unaffected.

3.6. Notes on the Independence from Irrelevant Alternatives (IIA) Assumption

A typical conditional logit model works under the assumption that the error terms, ε_{ij} , are independent, which is called the Independence from Irrelevant Alternatives (IIA) assumption. This assumption yields statistical properties that are computationally convenient, but it also implies restrictions on the individual's decision-making behavior. More specifically, the assumption implies that the relative probability of choosing between two alternatives is independent of any other alternatives. For example, under the IIA assumption, the relative probability of choosing between Utah and Colorado does not change whether or not California is available as a third choice. This assumption might be appropriate in some cases. For example, some retirees might choose Florida as their retirement destination regardless of which other states are available. Also many "job-

transfer” migrations may not consider alternatives. As a general rule, however, it does not seem like a sensible assumption.

To see the IIA property, one can look at the ratio of P_{ij} over P_{il} (where $j \neq l$). Let the conditional logit model be in its simplest form:

$$P_{ij} = \frac{e^{\alpha'Z_{ij}}}{\sum_j e^{\alpha'Z_{ij}}}; \text{ and } P_{il} = \frac{e^{\alpha'Z_{il}}}{\sum_j e^{\alpha'Z_{ij}}}$$

$$\Rightarrow \frac{P_{ij}}{P_{il}} = \frac{e^{\alpha'Z_{ij}}}{\sum_j e^{\alpha'Z_{ij}}} / \frac{e^{\alpha'Z_{il}}}{\sum_j e^{\alpha'Z_{ij}}}$$

$$\Rightarrow \frac{P_{ij}}{P_{il}} = \frac{e^{\alpha'Z_{ij}}}{e^{\alpha'Z_{il}}}$$

As shown, the relative probability of choosing between j and l depends only on the place attributes of j and l , while place attributes of other destinations (represented in the term $\sum_j e^{\alpha'Z_{ij}}$) fall out of the equation.

Violating the IIA assumption implies that the model either overestimates or underestimates P_{ij} (Train, 1986). On the other hand, had the model been able to take into account all relevant explanatory variables, the IIA assumption would not be violated. The residuals would no longer contain elements that significantly affect the relative choice among the alternatives. Dahlberg and Eklöf (2003) demonstrate, for a local residential choice model, that a well specified conditional logit model might in fact provide qualitatively similar results compared with a model that does not violate the IIA assumption.

This study applies the Small-Hsiao (SH)⁵ method (1985) to conduct the test on the IIA assumption. This SH method compares the likelihood values of two models: the

⁵ The traditional Hausman-McFadden (1984) IIA test was not applicable because the difference matrices were not positive definite.

unrestricted model, which includes all the alternatives in the choice set, and the restricted model, which excludes one or more alternatives from the choice set. If the model does not violate the IIA assumption then the parameter estimates obtained from the models will not be significantly different. According to the SH method, this implies that the log-likelihood values for both models are not significantly different either. The SH-statistics compute $-2*(L_{Unrestricted} - L_{Restricted})$ which follows the χ^2 distribution with degrees of freedom equal to the number of restrictions.

With 48 destination alternatives in the choice set, however, hundreds of restricted models are available. Testing all possible restricted models is not practical. Instead, the study tests only a few restricted models where highly populated states are excluded from the choice set. Excluding the largest states seems most likely to reveal violation of IIA.

The SH test is applied to seven restricted models that exclude the following states from the choice set:

- California
- New York
- Florida
- California and New York
- California and Florida
- New York and Florida
- California, New York, and Florida

Table 3.7 shows that the SH-statistics get larger as more states are excluded from the choice set. As more states are excluded, the restricted models become more different from the unrestricted model, causing the likelihood values between the two models to move further apart. As shown, all SH-statistics exceed 190, much larger than the critical values. Thus, the unrestricted model strongly violates the IIA assumption.

3.7. Concluding Remarks

Considering the significant impact of migration on the economy of a region, it is important to better understand the migration process. A potentially important part of this process involves the possible interaction of residential choice and occupational choice. This study shows a strong association between the supply of (migration) and the demand for individual's (industry mix) occupational skills. This study found that individuals are indeed more likely to choose states where the demand for their occupational skill is the strongest, and *vice versa*. This agrees with the underlying principle of migration as a human capital investment, controlling for amenity variables.

Another main contribution of this study is the application of a two-step maximum likelihood model, which has not been applied or developed in previous migration studies. This method offers a way to examine cases where an individual simultaneously makes two choices, in this case occupational choice and destination choice. It seems reasonable that individuals frequently face simultaneous choices. Thus, this methodology can be very useful for similar studies.

However, the migration model estimated here violates the IIA assumption. At least two options are available to deal with the violation of IIA. First, respecify the model such that it yields white noise error terms or second, apply a model that relaxes the IIA assumption. The next essay of this dissertation will develop models that relax the IIA assumption: a nested *logit* model and a mixed logit model. It will then compare results with those of the conditional logit model in order to determine if results are qualitatively the same, i.e., whether the conditional logit model is specified well enough that, from a practical perspective, the errors are effectively white noise and violation of IIA by this conditional logit model is inconsequential.

Table 3.1
Distribution of Managerial and Professional Occupations across Industrial Sectors by
State, 1990

State	Industry									Total
	Agri.	Construct.	Finance	Manufact.	Mining	Retail-Tr.	Service	Transport.	Wholesale	
Maine	3.0	2.7	5.6	12.4	0.1	5.4	64.9	3.8	2.1	100.0
New Hampshire	1.8	2.6	6.8	25.1	0.0	5.5	52.7	3.7	1.8	100.0
Vermont	8.4	2.6	4.4	10.4	0.6	5.7	60.8	3.8	3.3	100.0
Massachusetts	1.0	2.9	7.7	17.3	0.1	6.0	58.7	3.7	2.6	100.0
Rhode Island	1.1	2.5	7.3	16.3	0.1	8.0	58.7	3.4	2.6	100.0
Connecticut	1.1	2.7	10.9	19.5	0.2	5.6	52.0	4.9	3.1	100.0
New York	1.4	2.4	9.5	12.8	0.1	5.9	60.6	4.8	2.6	100.0
New Jersey	1.1	3.3	9.8	14.7	0.2	6.6	54.4	6.5	3.5	100.0
Pennsylvania	2.8	3.5	6.3	14.9	0.4	6.7	58.9	4.3	2.2	100.0
Ohio	3.5	2.8	5.7	16.6	0.4	8.0	56.2	4.1	2.8	100.0
Indiana	5.8	3.0	5.7	17.1	0.2	7.5	54.4	4.2	2.1	100.0
Illinois	4.2	2.7	7.9	14.8	0.2	7.0	55.1	4.9	3.1	100.0
Michigan	3.1	2.8	5.2	19.1	0.1	7.7	55.9	3.5	2.4	100.0
Wisconsin	10.4	2.6	5.7	15.8	0.1	6.1	53.5	3.7	2.0	100.0
Minnesota	9.6	2.7	6.2	15.7	0.1	6.2	52.1	4.2	3.3	100.0
Iowa	20.9	1.8	5.8	10.9	0.0	5.9	49.5	3.3	1.9	100.0
Missouri	7.5	2.7	5.5	14.1	0.1	7.3	54.7	5.4	2.6	100.0
North Dakota	29.3	4.0	4.2	3.9	1.0	5.5	47.1	4.0	1.0	100.0
South Dakota	32.1	2.7	5.1	6.1	0.3	5.1	43.4	4.3	1.0	100.0
Nebraska	20.3	2.8	5.6	6.8	0.2	6.4	50.1	5.6	2.1	100.0
Kansas	12.4	2.7	5.7	11.6	0.8	6.1	52.7	5.6	2.5	100.0
Delaware	9.4	5.5	4.5	10.2	0.0	7.1	57.7	5.5	0.1	100.0
Maryland & DC	1.8	4.5	8.2	8.6	0.1	6.1	64.0	4.4	2.3	100.0
Virginia	2.9	5.0	7.0	10.4	0.2	7.1	59.9	5.3	2.2	100.0
West Virginia	4.6	3.3	3.4	8.0	1.9	6.3	65.4	6.2	0.9	100.0
North Carolina	5.1	3.4	5.1	15.6	0.1	7.4	56.1	4.6	2.5	100.0
South Carolina	3.2	5.1	5.5	15.0	0.1	7.3	56.5	5.2	2.1	100.0
Georgia	3.6	3.7	6.5	11.8	0.2	8.1	54.9	7.6	3.6	100.0
Florida	2.7	4.2	8.2	10.7	0.2	8.2	56.9	5.9	3.1	100.0
Kentucky	10.6	2.8	5.8	11.1	0.8	8.2	54.7	4.3	1.7	100.0
Tennessee	4.5	3.2	5.9	14.1	0.2	7.0	56.0	6.6	2.4	100.0
Alabama	4.2	4.1	5.3	11.5	0.2	7.7	59.2	5.9	1.9	100.0
Mississippi	7.8	3.2	4.3	11.4	0.9	6.6	59.9	4.0	1.8	100.0
Arkansas	12.0	2.6	5.3	10.1	0.3	7.7	54.8	5.1	1.9	100.0
Louisiana	3.9	3.8	6.0	7.7	2.3	7.0	61.5	5.1	2.7	100.0
Oklahoma	7.7	2.9	6.1	9.6	3.7	6.2	56.2	5.6	2.1	100.0
Texas	4.2	3.9	6.8	12.3	3.0	7.9	53.5	5.5	3.0	100.0
Montana	18.8	3.4	3.8	4.2	0.6	5.9	56.1	5.7	1.5	100.0
Idaho	15.0	2.8	4.8	10.3	0.6	6.1	53.9	3.7	2.8	100.0
Wyoming	13.2	4.0	3.8	4.7	5.3	5.1	57.9	5.9	0.1	100.0
Colorado	4.6	3.7	6.6	13.2	1.9	6.3	54.7	6.6	2.4	100.0
New Mexico	3.3	4.3	5.0	7.9	1.9	6.7	65.7	3.9	1.4	100.0
Arizona	2.3	3.9	7.3	14.2	0.5	7.0	57.0	5.7	2.0	100.0
Utah	3.8	3.0	5.4	13.9	0.7	7.1	59.4	4.8	1.8	100.0
Nevada	2.6	6.0	7.9	5.4	1.5	8.1	61.3	5.3	2.0	100.0
Washington	4.4	3.7	6.5	15.7	0.1	6.9	55.3	5.1	2.5	100.0
Oregon	6.5	3.3	7.3	12.5	0.2	7.7	55.5	4.5	2.7	100.0
California	2.4	3.8	8.0	17.1	0.3	6.3	54.8	4.5	2.8	100.0
Average	7.2	3.4	6.2	12.4	0.7	6.7	56.4	4.9	2.2	100.0

Source: 1990 5% PUMS, U.S. Census Bureau.

Table 3.2
Distribution of Sales Occupations across Industrial Sectors by State, 1990

State	Industry									Total
	Agri.	Construct.	Finance	Manufact.	Mining	Retail-Tr.	Service	Transport.	Wholesale	
Maine	0.4	0.8	13.5	4.0	0.0	67.6	3.7	1.3	8.7	100.0
New Hampshire	0.5	0.6	14.0	7.0	0.0	59.2	6.7	1.7	10.3	100.0
Vermont	1.1	0.0	12.4	4.0	0.0	63.2	6.6	0.8	11.9	100.0
Massachusetts	0.4	0.5	13.9	6.4	0.0	58.6	8.1	2.0	10.1	100.0
Rhode Island	0.7	0.3	12.2	5.9	0.0	64.6	4.2	1.3	10.8	100.0
Connecticut	0.4	0.6	17.5	6.3	0.0	56.6	7.7	1.1	9.8	100.0
New York	0.2	0.5	14.9	6.3	0.0	58.5	7.1	1.9	10.5	100.0
New Jersey	0.2	0.6	16.2	7.0	0.0	53.9	7.2	2.0	12.8	100.0
Pennsylvania	0.3	0.7	11.1	6.1	0.0	63.1	6.9	1.5	10.2	100.0
Ohio	0.4	0.8	10.9	7.1	0.0	62.3	6.8	1.5	10.1	100.0
Indiana	0.3	0.7	10.4	6.3	0.1	63.3	6.8	1.7	10.5	100.0
Illinois	0.3	0.7	13.8	7.0	0.0	57.2	7.7	2.1	11.3	100.0
Michigan	0.4	0.6	10.9	6.3	0.0	64.3	6.6	1.3	9.5	100.0
Wisconsin	0.4	1.0	12.6	7.4	0.0	61.7	6.4	1.0	9.6	100.0
Minnesota	0.5	0.6	13.5	7.9	0.0	56.0	7.4	1.5	12.6	100.0
Iowa	1.0	0.7	11.5	6.0	0.2	61.0	5.7	2.3	11.7	100.0
Missouri	0.4	0.8	11.4	6.3	0.0	60.4	7.4	2.6	10.6	100.0
North Dakota	0.7	0.2	11.3	2.8	0.6	68.5	5.7	0.8	9.6	100.0
South Dakota	0.5	0.9	9.3	6.4	0.0	62.9	4.5	1.8	13.7	100.0
Nebraska	0.7	0.5	13.4	5.6	0.0	58.1	7.7	1.3	12.7	100.0
Kansas	0.5	0.4	12.2	4.8	0.0	62.2	5.7	2.3	11.8	100.0
Delaware	0.0	0.0	11.7	1.7	0.0	69.9	8.6	2.4	5.8	100.0
Maryland & DC	0.3	0.8	14.4	4.5	0.0	60.0	7.2	2.2	10.5	100.0
Virginia	0.2	0.5	12.3	4.3	0.0	65.4	7.6	1.5	8.2	100.0
West Virginia	0.1	0.1	9.9	4.5	0.3	72.1	5.1	0.5	7.5	100.0
North Carolina	0.1	0.5	10.6	5.8	0.1	65.2	6.0	1.3	10.4	100.0
South Carolina	0.2	0.9	9.9	4.2	0.0	66.7	6.6	1.2	10.2	100.0
Georgia	0.3	0.6	11.7	5.7	0.0	60.5	7.3	2.2	11.7	100.0
Florida	0.3	0.8	14.7	4.5	0.0	58.9	7.7	2.0	11.0	100.0
Kentucky	0.2	0.7	10.3	4.5	0.3	67.4	6.6	1.9	8.1	100.0
Tennessee	0.2	0.5	11.1	5.0	0.1	63.5	6.7	1.7	11.2	100.0
Alabama	0.2	0.6	9.0	4.8	0.0	67.9	6.4	1.3	9.6	100.0
Mississippi	0.3	0.4	8.2	3.7	0.0	70.6	5.9	1.4	9.4	100.0
Arkansas	0.3	0.8	10.0	5.5	0.2	64.3	6.5	2.3	10.3	100.0
Louisiana	0.2	0.6	9.7	3.7	0.7	67.0	7.1	1.8	9.2	100.0
Oklahoma	0.3	0.5	10.6	4.6	0.4	64.2	7.8	1.7	10.0	100.0
Texas	0.3	0.6	11.5	4.7	0.4	62.0	7.3	1.9	11.3	100.0
Montana	0.7	0.4	9.4	3.7	0.1	72.4	3.2	1.5	8.5	100.0
Idaho	0.8	1.2	11.3	3.2	0.0	62.1	6.7	2.3	12.4	100.0
Wyoming	0.0	1.0	11.4	3.4	0.6	65.6	6.7	3.3	8.1	100.0
Colorado	0.1	0.5	15.6	4.7	0.2	59.0	7.5	1.7	10.8	100.0
New Mexico	0.1	0.3	9.5	4.6	0.1	66.8	7.8	1.6	9.1	100.0
Arizona	0.2	0.3	14.7	5.7	0.0	58.6	8.7	1.5	10.3	100.0
Utah	0.2	0.4	10.8	5.3	0.1	61.0	9.7	2.6	9.9	100.0
Nevada	0.0	0.3	11.4	2.5	0.1	54.6	22.2	1.5	7.5	100.0
Washington	0.6	0.5	13.3	5.4	0.0	58.2	7.7	2.0	12.3	100.0
Oregon	0.4	0.4	11.1	6.2	0.0	61.9	6.9	1.3	11.8	100.0
California	0.3	0.6	15.5	6.0	0.0	57.3	7.6	1.7	11.0	100.0
Average	0.4	0.6	12.0	5.2	0.1	62.6	7.1	1.7	10.3	100.0

Source: 1990 5% PUMS, U.S. Census Bureau.

Table 3.3
Distribution of Operator Occupations across Industrial Sectors by State, 1990

State	Industry									Total
	Agri.	Construct.	Finance	Manufact.	Mining	Retail-Tr.	Service	Transport.	Wholesale	
Maine	2.3	7.3	0.6	56.4	0.1	7.7	7.8	12.2	5.6	100.0
New Hampshire	0.4	7.3	0.3	63.1	0.0	5.1	7.5	11.3	5.0	100.0
Vermont	0.3	6.8	0.2	61.4	1.0	7.1	9.8	7.2	6.1	100.0
Massachusetts	0.7	4.3	0.4	55.2	0.1	7.2	10.1	16.0	5.9	100.0
Rhode Island	0.2	3.7	0.7	64.8	0.0	7.4	9.2	9.5	4.6	100.0
Connecticut	0.5	5.0	0.7	59.0	0.2	6.7	10.7	11.2	5.8	100.0
New York	0.3	4.6	0.7	50.3	0.2	5.9	11.9	20.0	6.0	100.0
New Jersey	0.4	4.1	0.6	52.1	0.2	6.8	11.3	17.8	6.8	100.0
Pennsylvania	0.5	4.9	0.3	58.7	1.1	6.4	8.8	13.9	5.5	100.0
Ohio	0.5	3.2	0.4	62.5	0.9	5.2	9.9	12.9	4.6	100.0
Indiana	0.6	3.6	0.2	64.2	0.8	5.2	8.5	12.8	4.1	100.0
Illinois	0.8	3.4	0.4	56.0	0.8	5.8	10.0	17.1	5.8	100.0
Michigan	0.5	3.6	0.3	65.5	0.4	5.4	9.7	10.3	4.2	100.0
Wisconsin	0.7	4.4	0.3	63.3	0.3	5.1	7.8	13.5	4.5	100.0
Minnesota	1.1	5.1	0.5	53.5	0.4	7.3	11.8	15.8	4.5	100.0
Iowa	1.5	6.5	0.8	53.2	0.6	6.2	12.4	13.8	5.2	100.0
Missouri	1.3	4.2	0.1	54.7	0.5	6.2	9.4	18.2	5.3	100.0
North Dakota	9.3	10.8	0.6	20.1	4.9	10.6	10.4	26.6	6.6	100.0
South Dakota	3.3	6.4	0.4	45.9	1.0	7.5	10.8	18.3	6.4	100.0
Nebraska	2.4	6.0	0.4	44.8	0.8	7.8	10.9	19.7	7.0	100.0
Kansas	2.1	7.0	0.6	46.5	2.6	7.2	12.0	17.2	4.9	100.0
Delaware	1.5	5.5	0.2	50.4	0.0	11.4	6.8	15.4	8.8	100.0
Maryland & DC	0.6	8.0	0.7	40.9	0.8	8.3	15.1	19.9	5.7	100.0
Virginia	0.7	6.1	0.4	54.6	1.8	6.2	11.4	13.8	4.9	100.0
West Virginia	0.2	8.8	0.1	41.3	11.7	5.1	11.9	15.6	5.3	100.0
North Carolina	1.0	3.9	0.2	70.4	0.3	4.5	6.8	9.6	3.3	100.0
South Carolina	1.0	4.5	0.2	70.0	0.2	4.8	7.0	8.2	4.1	100.0
Georgia	1.0	4.9	0.2	60.0	0.4	6.4	8.4	13.9	4.9	100.0
Florida	1.5	7.2	0.6	40.3	0.5	10.0	15.0	18.2	6.7	100.0
Kentucky	0.4	4.9	0.2	56.8	4.7	5.5	9.4	14.1	4.0	100.0
Tennessee	0.5	4.4	0.1	66.4	0.5	4.3	7.2	13.0	3.7	100.0
Alabama	0.9	4.7	0.2	65.5	1.2	4.3	7.6	11.7	3.9	100.0
Mississippi	1.6	4.7	0.3	65.1	1.3	3.4	6.3	13.6	3.7	100.0
Arkansas	1.0	5.9	0.3	62.1	0.7	5.1	6.4	13.4	5.0	100.0
Louisiana	1.1	9.5	0.4	37.9	4.7	6.7	12.1	21.5	6.1	100.0
Oklahoma	1.1	7.4	0.4	45.2	4.4	6.7	12.4	17.5	4.8	100.0
Texas	1.1	7.4	0.3	43.8	2.9	7.8	12.6	17.6	6.4	100.0
Montana	3.4	12.8	0.1	24.5	6.1	9.7	12.2	26.8	4.4	100.0
Idaho	5.9	9.5	0.7	42.4	1.9	5.9	11.4	16.4	5.9	100.0
Wyoming	1.1	12.3	0.0	20.4	18.0	10.3	12.2	24.0	1.7	100.0
Colorado	1.2	7.6	0.3	43.4	2.0	8.3	14.3	16.6	6.2	100.0
New Mexico	2.2	11.1	0.2	33.2	4.3	10.6	13.7	17.5	7.2	100.0
Arizona	0.9	8.5	0.3	38.9	3.1	9.0	15.2	18.4	5.6	100.0
Utah	0.5	5.8	0.3	50.1	2.3	7.2	11.0	16.4	6.3	100.0
Nevada	0.1	8.0	0.3	23.6	6.9	11.1	22.0	22.2	5.9	100.0
Washington	1.7	6.3	0.3	48.8	0.3	7.5	10.8	18.1	6.2	100.0
Oregon	2.0	5.5	0.2	52.9	0.5	5.2	9.9	17.5	6.3	100.0
California	1.7	4.9	0.4	51.3	0.7	7.3	12.9	14.3	6.7	100.0
Average	1.4	6.3	0.4	51.1	2.1	6.9	10.7	15.8	5.4	100.0

Source: 1990 5% PUMS, U.S. Census Bureau.

Table 3.4
Distribution of Laborer Occupations across Industrial Sectors by State, 1990

State	Industry									Total
	Agri.	Construct.	Finance	Manufact.	Mining	Retail-Tr.	SeNice	Transport.	Wholesale	
Maine	26.3	13.0	0.4	25.8	0.1	19.7	6.3	3.8	4.6	100.0
New Hampshire	14.0	19.5	2.1	18.1	0.0	28.7	7.2	6.9	3.5	100.0
Vermont	29.9	15.6	5.1	13.7	0.0	19.4	9.9	3.3	3.1	100.0
Massachusetts	14.6	16.6	1.1	16.7	0.1	29.5	7.9	7.2	6.3	100.0
Rhode Island	21.3	12.6	1.6	20.4	0.4	22.6	10.6	4.1	6.3	100.0
Connecticut	18.9	16.8	1.3	12.0	0.4	30.6	9.5	5.1	5.4	100.0
New York	13.4	19.4	1.6	15.8	0.1	23.4	11.1	10.0	5.3	100.0
New Jersey	13.6	16.6	1.4	18.3	0.1	24.5	8.1	10.0	7.4	100.0
Pennsylvania	12.8	15.1	0.9	26.2	0.6	23.6	7.9	7.1	5.8	100.0
Ohio	13.9	13.3	1.2	24.4	0.3	25.5	9.8	5.5	6.1	100.0
Indiana	17.1	16.1	0.9	24.5	0.4	20.7	9.1	5.6	5.6	100.0
Illinois	16.2	13.5	1.3	20.8	0.6	22.8	9.8	8.6	6.6	100.0
Michigan	16.3	12.9	1.1	21.5	0.3	27.5	10.1	4.6	5.7	100.0
Wisconsin	23.3	10.8	1.0	27.1	0.0	19.8	8.3	4.5	5.1	100.0
Minnesota	24.7	10.6	1.0	19.3	0.2	22.5	9.0	5.9	6.8	100.0
Iowa	32.6	11.9	0.4	19.4	0.2	19.3	5.3	3.1	7.8	100.0
Missouri	21.6	15.7	0.9	18.6	0.3	21.8	9.4	6.7	5.1	100.0
North Dakota	45.9	12.3	0.2	7.9	2.3	15.9	5.6	6.5	3.4	100.0
South Dakota	42.2	9.5	0.5	8.8	1.1	20.0	7.2	5.1	5.7	100.0
Nebraska	34.4	12.1	1.5	14.9	0.4	16.5	6.5	7.5	6.1	100.0
Kansas	24.5	14.4	1.0	16.7	1.0	22.3	7.9	5.9	6.3	100.0
Delaware	18.7	13.1	0.3	34.0	0.0	15.6	9.2	2.7	6.5	100.0
Maryland & DC	17.8	22.7	1.1	9.4	0.0	23.2	10.1	9.0	6.7	100.0
Virginia	20.9	21.7	1.2	17.2	0.3	21.0	8.0	5.2	4.7	100.0
West Virginia	10.9	20.8	2.1	19.5	7.4	22.7	6.4	7.0	3.1	100.0
North Carolina	24.4	14.2	1.0	26.3	0.1	17.7	7.1	5.5	3.7	100.0
South Carolina	21.2	16.3	0.5	24.5	0.2	19.1	8.8	4.4	4.9	100.0
Georgia	21.0	17.8	0.8	21.0	0.2	18.0	8.0	7.1	6.2	100.0
Florida	31.8	16.7	1.0	8.0	0.1	20.9	10.7	6.0	4.7	100.0
Kentucky	24.9	15.7	1.0	21.0	1.6	19.0	6.5	6.0	4.2	100.0
Tennessee	17.7	15.5	1.0	25.0	0.5	19.9	7.7	8.0	4.9	100.0
Alabama	20.0	16.5	0.5	25.4	0.5	18.7	7.1	5.6	5.8	100.0
Mississippi	23.8	14.6	0.7	28.1	1.1	17.3	5.8	4.5	4.3	100.0
Arkansas	28.3	12.3	0.5	28.8	0.2	16.3	4.5	5.1	4.2	100.0
Louisiana	21.8	18.2	0.7	15.8	2.0	20.0	7.2	7.4	7.0	100.0
Oklahoma	26.9	15.3	1.0	14.6	0.9	23.9	8.0	5.2	4.3	100.0
Texas	27.8	17.4	1.1	11.8	0.8	20.3	8.7	6.3	5.9	100.0
Montana	38.2	12.4	1.1	18.5	0.7	11.1	6.2	6.7	5.1	100.0
Idaho	44.9	9.6	1.1	16.3	0.4	10.5	6.7	4.3	6.3	100.0
Wyoming	31.9	18.2	1.5	7.9	5.3	17.9	9.7	5.6	2.0	100.0
Colorado	25.1	15.6	1.3	10.2	0.5	22.2	12.1	7.1	6.0	100.0
New Mexico	26.1	19.8	0.6	9.0	1.3	19.9	10.8	7.0	5.5	100.0
Arizona	27.3	16.7	1.5	7.4	0.4	23.0	13.7	5.6	4.3	100.0
Utah	19.3	17.6	0.8	12.5	0.8	26.0	13.2	5.5	4.2	100.0
Nevada	18.2	21.9	3.4	6.3	1.4	21.7	17.3	5.8	4.0	100.0
Washington	33.2	13.0	1.0	15.3	0.1	18.9	6.9	5.2	6.3	100.0
Oregon	33.0	8.5	0.5	23.1	0.1	18.5	6.9	4.0	5.4	100.0
California	34.3	18.3	0.7	10.9	0.2	16.3	8.5	5.2	5.5	100.0
Average	24.3	15.4	1.1	17.9	0.7	20.8	8.6	5.9	5.3	100.0

Source: 1990 5% PUMS , U.S. Census Bureau.

Table 3.5
Multinomial Logit Model of Occupational Choice
Selected Variables (N=6359)

Choice of Occupation in 1998	Explanatory Variables (Characterisrics in 1993)	Parameter Estimates	Standard Error	t-Value
Professional_ Managerial	Prof_Mgr	1.905	0.088	21.7 ***
	Sales	1.261	0.127	9.9 ***
	Laborer	-0.296	0.189	-1.6
	Operator	-0.187	0.156	-1.2
	Educational Attainment	0.323	0.018	18 ***
	Age	-0.017	0.014	-1.2
	White	0.314	0.073	4.3 ***
	Male	0.106	0.071	1.5
	Move	-0.093	0.213	-0.4
	House Tenure	-0.011	0.007	-1.5
	Having a working spouse	0.311	0.080	3.9 ***
Sales	Prof_Mgr	0.604	0.217	2.8 ***
	Sales	2.619	0.171	15.3 ***
	Laborer	0.569	0.313	1.8 *
	Operator	-0.254	0.378	-0.7
	Educational Attainment	0.194	0.035	5.6 ***
	Age	-0.037	0.028	-1.3
	White	0.406	0.144	2.8 ***
	Male	-0.098	0.141	-0.7
	Move	-0.924	0.607	-1.5
	House Tenure	-0.004	0.014	-0.3
	Having a working spouse	0.255	0.159	1.6 *
Operator Occupations	Prof_Mgr	0.533	0.171	3.1 ***
	Sales	0.256	0.239	1.1
	Laborer	1.042	0.152	6.9 ***
	Operator	2.111	0.112	18.8 ***
	Educational Attainment	-0.205	0.025	-8.4 ***
	Age	-0.017	0.018	-1.0
	White	-0.160	0.100	-1.6
	Male	0.756	0.096	7.8 ***
	Move	0.157	0.302	0.5
	House Tenure	-0.003	0.009	-0.3
	Having a working spouse	-0.188	0.099	-1.9 *
Laborer Occupations	Prof_Mgr	-0.025	0.264	-0.1
	Sales	0.185	0.309	0.6
	Laborer	1.562	0.158	9.9 ***
	Operator	0.546	0.190	2.9 ***
	Educational Attainment	-0.193	0.032	-6.1 ***
	Age	0.017	0.024	0.7
	White	-0.042	0.136	-0.3
	Male	1.627	0.155	10.5 ***
	Move	-0.323	0.409	-0.8
	House Tenure	-0.001	0.012	-0.1
	Having a working spouse	-0.507	0.132	-3.8 ***

Log Likelihood value = -6264

Note: *** indicates significance at the one-percent level; ** indicates significance at the five-percent level; and * indicates significance at the ten percent level.

Table 3.6
Conditional Logit Model of State Destination Choice (N=6359)

Explanatory Variables	Parameter Estimates	Non-Adjusted		Adjusted	
		Standard Error	t-Value	Standard Error	t-Value
Age*Non_Origin	-0.061	0.021	-2.88 ***	0.021	-2.88 ***
White*Non_Origin	-0.047	0.096	-0.49	0.097	-0.48
House_Tenure*Non_Origin	-0.089	0.011	-8.21 ***	0.011	-8.25 ***
Male*Non_Origin	0.129	0.095	1.36	0.094	1.37
Spouse_Working*Non_Origin	-0.079	0.107	-0.74	0.107	-0.74
Family_Size*Non_Origin	-0.041	0.035	-1.17	0.035	-1.17
Freq_of_Previous Interstate Moves*Non_Origin	0.324	0.032	10.21 ***	0.032	10.22 ***
Education*Distance	1.726	1.072	1.61	1.071	1.61
Distance	-66.869	18.430	-3.63 ***	18.308	-3.65 ***
Dummy of Non_Origin	-3.112	0.759	-4.10 ***	0.759	-4.10 ***
Dummy of Adjacent	1.195	0.128	9.34 ***	0.128	9.35 ***
State's Employment Share	0.086	0.017	5.00 ***	0.017	5.00 ***
January Temperature	0.008	0.006	1.34	0.006	1.34
July Temperature	0.025	0.012	2.05 **	0.012	2.05 **
Humidity	0.022	0.005	4.30 ***	0.005	4.29 ***
Emp-Growth*Pk_Strong Association	0.456	0.089	5.14 ***	0.091	5.01 ***
Emp-Growth*Pk_Medium Association	0.029	0.060	0.49	0.060	0.49
Emp-Growth*Pk_Weak Association	-0.112	0.067	-1.66 *	0.067	-1.67 *

Note: *** indicates significance at the one-percent level; ** indicates significance at the five-percent level; and * indicates significance at the ten percent level.

Table 3.7
Small-Hsiao IIA Test Results

Statistics	Un-restricted Model	Restricted models, excluding:						
		CA	FL	NY	CA & FL	CA & NY	FL & NY	CA, FL, & NY
Log-Likelihood Value	-3,655	-3,473	-3,324	-3,560	-3,114	-3,362	-3,238	-3,012
SH-Statistics Value	--	364	662	190	1,082	586	834	1,286

Appendix 3.1. Adjusting the Variance

Let β , $\mathbf{D}_{MNL}$ and $\mathbf{V}_{MNL}$ be the parameter, the first derivative and the variance of the log likelihood function of the occupational choice multinomial logit model and α , $\mathbf{D}_{CL}$ and $\mathbf{V}_{CL}$ be the parameter, first derivative and variance of the destination choice conditional logit model, then the correct asymptotic variance of the system, $\mathbf{V}_S$ can be computed as:

$$\mathbf{V}_S = \mathbf{V}_{CL} + \mathbf{V}_{CL}(\mathbf{C}\mathbf{V}_{MNL}\mathbf{C}' - \mathbf{R}\mathbf{V}_{MNL}\mathbf{C}' - \mathbf{C}\mathbf{V}_{MNL}\mathbf{R}')\mathbf{V}_{CL}.$$

where the estimates of:

$$\begin{aligned}\mathbf{R} &= \mathbf{D}_{CL} \mathbf{D}_{MNL}' \\ \mathbf{C} &= \mathbf{D}_{CL} \mathbf{D}_{Cross}'\end{aligned}$$

We have found that:

$$\mathbf{D}_{MNL} = \frac{\partial \text{Ln } L_{MNL}}{\partial \beta_k} = \sum_i \mathbf{x}_i [\mathbf{d}_{ik} - \mathbf{P}_{ik}]$$

$$\text{and } \mathbf{D}_{CL} = \frac{\partial \text{Ln } L_{CL}}{\partial \alpha} = \sum_i \sum_j \mathbf{d}_{ij} [\mathbf{z}_{ij} - \sum_j \mathbf{z}_{ij} \mathbf{P}_{ij}]$$

We still need to find out:

$$\mathbf{D}_{Cross} = \frac{\partial \text{Ln } L_{CL}}{\partial \beta_k}$$

Deriving $\mathbf{D}_{Cross}$

$$\log L_{CL} = \sum_i \sum_j \mathbf{d}_{ij} \log \mathbf{P}_{ij}. \quad \mathbf{P}_{ij} \text{ is a function of } \beta \text{ since } \mathbf{Z} \text{ in } \mathbf{P}_{ij} \text{ contain } \bar{\mathbf{Z}}, \tilde{\mathbf{Z}}^* \mathbf{X} \text{ and } \overset{\circ}{\mathbf{Z}}^* \hat{\mathbf{P}}_k.$$

For each individual i :

$$\frac{\partial \log L_{CL}}{\partial \beta_k} = \sum_j d_j \frac{1}{P_j} \frac{\partial P_j}{\partial \beta_k} \quad (i)$$

$$\begin{aligned} \frac{\partial P_j}{\partial \beta_k} &= \frac{\frac{\partial}{\partial \beta_k} \frac{e^{\alpha' z_j \hat{P}_k}}{\sum_j e^{\alpha' z_j \hat{P}_k}}}{\sum_j e^{\alpha' z_j \hat{P}_k}} \\ &= \frac{[\Sigma] \frac{\partial e^{\alpha' z_j \hat{P}_k}}{\partial \beta_k} - e^{\alpha' z_j \hat{P}_k} \frac{\partial [\Sigma]}{\partial \beta_k}}{[\Sigma]^2}, \text{ where } [\Sigma] = \sum_j e^{\alpha' z_j \hat{P}_k} \end{aligned} \quad (ii)$$

$$\frac{\partial e^{\alpha' z_j \hat{P}_k}}{\partial \beta_k} = \alpha' z_j \frac{\partial \hat{P}_k}{\partial \beta_k} e^{\alpha' z_j \hat{P}_k} \quad (iii)$$

let $j=1,2,3$:

$$\begin{aligned} \frac{\partial [\Sigma]}{\partial \beta_k} &= \frac{\partial e^{\alpha' z_1 \hat{P}_k}}{\partial \beta_k} + \frac{\partial e^{\alpha' z_2 \hat{P}_k}}{\partial \beta_k} + \frac{\partial e^{\alpha' z_3 \hat{P}_k}}{\partial \beta_k} \\ &= \alpha' z_1 \frac{\partial \hat{P}_k}{\partial \beta_k} e^{\alpha' z_1 \hat{P}_k} + \alpha' z_2 \frac{\partial \hat{P}_k}{\partial \beta_k} e^{\alpha' z_2 \hat{P}_k} + \alpha' z_3 \frac{\partial \hat{P}_k}{\partial \beta_k} e^{\alpha' z_3 \hat{P}_k} \\ &= \sum_j \alpha' z_j \frac{\partial \hat{P}_k}{\partial \beta_k} e^{\alpha' z_j \hat{P}_k} \end{aligned} \quad (iv)$$

To solve $\frac{\partial \hat{P}_k}{\partial \beta_k}$, let's suppose two possible different occupations to be labeled k and l .

For k equal l :

$$\begin{aligned}\frac{\partial \hat{P}_l}{\partial \beta_k} &= \frac{\partial e^{\beta_l' X}}{[\Sigma_k]} , \text{ where } [\Sigma_k] = \Sigma_k e^{\beta_k' X} \\ &= \frac{([\Sigma_k] X e^{\beta_l' X} - e^{\beta_l' X} X e^{\beta_k' X})}{[\Sigma_k]^2} \\ &= X(\hat{P}_l - \hat{P}_l \hat{P}_k) ,\end{aligned}$$

For k equal l :

$$\begin{aligned}\frac{\partial \hat{P}_l}{\partial \beta_k} &= \frac{-e^{\beta_l' X} X e^{\beta_k' X}}{[\Sigma_k]^2} \\ &= X(0 - \hat{P}_l \hat{P}_k) ,\end{aligned}$$

In general:

$$\frac{\partial \hat{P}_l}{\partial \beta_k} = X(d_k \hat{P}_l - \hat{P}_l \hat{P}_k) , \quad (v)$$

where $d_k = 1$ if l equals k , and 0 otherwise.

Substituting (v) into (iv) yields:

$$\frac{\partial [\Sigma]}{\partial \beta_k} = \Sigma_j \alpha' z_j (X(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) e^{\alpha' z_j \hat{P}_k} \quad (iv^*)$$

Substituting (v) into (iii) yields:

$$\frac{\partial e^{\alpha' z_j \hat{P}_k}}{\partial \beta_k} = \alpha' z_j (X(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) e^{\alpha' z_j \hat{P}_k} \quad (iii^*)$$

Substituting (iii*) and (iv*) into (ii) yields:

$$\begin{aligned}
\frac{\partial P_j}{\partial \beta_k} &= \frac{[\Sigma] \alpha' z_j (x(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) e^{\alpha' z_j \hat{P}_k} - e^{\alpha' z_j \hat{P}_k} \Sigma_j \alpha' z_j (x(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) e^{\alpha' z_j \hat{P}_k}}{[\Sigma]^2} \\
&= \frac{\alpha' z_j (x(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) e^{\alpha' z_j \hat{P}_k}}{[\Sigma]} - \frac{P_j \Sigma_j \alpha' z_j (x(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) e^{\alpha' z_j \hat{P}_k}}{[\Sigma]} \\
&= \alpha' z_j ((x(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) P_j - P_j \Sigma_j \alpha' z_j ((x(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) P_j) \\
&= P_j (\alpha' z_j - \Sigma_j \alpha' z_j P_j) (x(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) \quad (ii*)
\end{aligned}$$

Substituting (ii*) into (i) yields:

$$\begin{aligned}
\frac{\partial \log L_{CL}}{\partial \beta_k} &= \sum_j d_j \frac{1}{P_j} \{ P_j (\alpha' z_j - \Sigma_j \alpha' z_j P_j) (x(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) \} \\
&= \sum_j d_j (\alpha' z_j - \Sigma_j \alpha' z_j P_j) (x(d_k \hat{P}_k - \hat{P}_k \hat{P}_k)) \quad (i*)
\end{aligned}$$

Thus, the solution for all individuals:

$$\frac{\partial \log L_{CL}}{\partial \beta_k} = \sum_i \sum_j d_j (\alpha' z_j - \Sigma_j \alpha' z_j P_j) (x(d_k \hat{P}_k - \hat{P}_k \hat{P}_k))$$

The solution above applies to the general form of the conditional *logit* model equation.

Appendix 3.2 describes the solution applied to the real form of the conditional *logit* model equation applied in this study.

Appendix 3.2. Adjusting the Variance: The Real Form of the Conditional *Logit* Model Equation

For simplicity, let's leave out the subscript i denoting the individual. In this particular study, the real P_j takes the following form:

$$P_j = \frac{e^{\alpha_A'Z_A + \alpha_H(Z_{1j}*\hat{P}_{1j} + Z_{2j}*\hat{P}_{2j} + Z_{3j}*\hat{P}_{3j} + Z_{4j}*\hat{P}_{4j}) + \alpha_M(Z_{1j}*\hat{P}_{3j} + Z_{2j}*\hat{P}_{4j} + Z_{3j}*\hat{P}_{1j} + Z_{4j}*\hat{P}_{2j}) + \alpha_L(Z_{1j}*\hat{P}_{2j} + Z_{1j}*\hat{P}_{4j} + Z_{2j}*\hat{P}_{1j} + Z_{2j}*\hat{P}_{3j} + Z_{3j}*\hat{P}_{2j} + Z_{4j}*\hat{P}_{1j} + Z_{4j}*\hat{P}_{2j} + Z_{4j}*\hat{P}_{3j})}}{\sum_j e^{\alpha'(Z_j \hat{P}_k)}}$$

where α consists of α_A , α_H , α_M , and α_L . α_A is the coefficient for variable Z_A , α_H is the coefficient for the variable representing the collective interactions of pairs with strong association; α_M is for those with medium degree of association; and α_L is for those with weak association. Z consists of Z_A , representing place attributes that do not interact with $\hat{P}_k$, as well as Z_1 , Z_2 , Z_3 , and Z_4 , representing the employment growth for each sector. Each of Z_1 , Z_2 , Z_3 , and Z_4 , interacts with $\hat{P}_k$.⁶

Let's simplify the above form into:

$$P_j = \frac{e^{[ZP]}}{[\Sigma]}$$

Given $\text{Log } L_{CL} = \sum_i \sum_j d_{ij} \log P_{ij}$, then for each individual i :

$$\frac{\partial \text{Log } L_{CL}}{\partial \beta_k} = \sum_j d_j \frac{1}{P_j} \frac{\partial P_j}{\partial \beta_k}$$

⁶ Note that the number of k (occupation categories) in $\hat{P}_k$ is four, with the 5th category as the reference. The model also includes four sectors: services, manufacturing, retail-trade, and all other.

For $k = 1$;

$$\begin{aligned}\frac{\partial P_i}{\partial \beta_l} &= \frac{[\Sigma] \frac{\partial e^{[\hat{ZP}_k]}]}{\partial \beta_l} - e^{[Z\hat{P}_k]} \frac{\partial [\Sigma]}{\partial \beta_l}}{[\Sigma]^2} \\ &= \frac{[\Sigma] [\alpha'z [\frac{\partial \hat{P}_k}{\partial \beta_l}]] e^{[Z\hat{P}_k]} - e^{[Z\hat{P}_k]} \Sigma_j [\alpha'z [\frac{\partial \hat{P}_k}{\partial \beta_l}]] e^{[Z'P]}}{[\Sigma]^2}\end{aligned}$$

Note that in this case $\alpha'z$ refers to the multiplication between $\alpha_H, \alpha_M, \alpha_W$ and Z_1, Z_2, Z_3 , and Z_4 only. This is because the derivative of the term $\alpha_A'Z_A$ with respect to β_k is zero.

$$= [\alpha'z [\frac{\partial \hat{P}_k}{\partial \beta_l}]] P_j - P_j \Sigma_j [\alpha'z [\frac{\partial \hat{P}_k}{\partial \beta_l}]] P_j$$

Thus,

$$\frac{\partial \log L_{CL}}{\partial \beta_l} = d_j \left\{ [\alpha'z [\frac{\partial \hat{P}_k}{\partial \beta_l}]] - \Sigma_j [\alpha'z [\frac{\partial \hat{P}_k}{\partial \beta_l}]] P_j \right\}$$

where in this case:

$$\begin{aligned}[\alpha'z [\frac{\partial \hat{P}_k}{\partial \beta_l}]] &= \alpha_H (Z_1 \frac{\partial \hat{P}_1}{\partial \beta_l} + Z_2 \frac{\partial \hat{P}_2}{\partial \beta_l} + Z_3 \frac{\partial \hat{P}_3}{\partial \beta_l} + Z_4 \frac{\partial \hat{P}_4}{\partial \beta_l}) + \\ &\quad \alpha_M (Z_1 \frac{\partial \hat{P}_3}{\partial \beta_l} + Z_2 \frac{\partial \hat{P}_4}{\partial \beta_l} + Z_3 \frac{\partial \hat{P}_1}{\partial \beta_l} + Z_4 \frac{\partial \hat{P}_2}{\partial \beta_l}) + \\ &\quad \alpha_L (Z_1 \frac{\partial \hat{P}_2}{\partial \beta_l} + Z_2 \frac{\partial \hat{P}_4}{\partial \beta_l} + Z_3 \frac{\partial \hat{P}_1}{\partial \beta_l} + Z_4 \frac{\partial \hat{P}_3}{\partial \beta_l} + Z_3 \frac{\partial \hat{P}_2}{\partial \beta_l} + Z_4 \frac{\partial \hat{P}_1}{\partial \beta_l} + Z_4 \frac{\partial \hat{P}_2}{\partial \beta_l} + \\ &\quad Z_4 \frac{\partial \hat{P}_3}{\partial \beta_l}) \\ &= \alpha_H \times \{(Z_1(\hat{P}_1 - \hat{P}_1\hat{P}_1) + Z_2(0 - \hat{P}_2\hat{P}_1) + Z_3(0 - \hat{P}_3\hat{P}_1) + Z_4(0 - \hat{P}_4\hat{P}_1))\} + \\ &\quad \alpha_M \times \{(Z_1(0 - \hat{P}_3\hat{P}_1) + Z_2(0 - \hat{P}_4\hat{P}_1) + Z_3(\hat{P}_1 - \hat{P}_1\hat{P}_1) + Z_4(0 - \hat{P}_4\hat{P}_1))\} + \\ &\quad \alpha_L \times \{(Z_1(0 - \hat{P}_2\hat{P}_1) + Z_2(0 - \hat{P}_4\hat{P}_1) + Z_3(\hat{P}_1 - \hat{P}_1\hat{P}_1) + Z_4(0 - \hat{P}_3\hat{P}_1) + \\ &\quad (Z_3(0 - \hat{P}_2\hat{P}_1) + Z_4(\hat{P}_1 - \hat{P}_1\hat{P}_1) + Z_4(0 - \hat{P}_2\hat{P}_1) + Z_4(0 - \hat{P}_3\hat{P}_1))\}\end{aligned}$$

Similarly, for k = 2:

$$\frac{\partial \text{Log } L_{CL}}{\partial \beta_2} = d_j \left\{ \left[\alpha'z \left[\frac{\partial \hat{P}_k}{\partial \beta_2} \right] \right] - \sum_j \left[\alpha'z \left[\frac{\partial \hat{P}_k}{\partial \beta_2} \right] \right] P_j \right\}, \text{ where:}$$

$$\begin{aligned} \left[\alpha'z \left[\frac{\partial \hat{P}_k}{\partial \beta_2} \right] \right] = & \alpha_H \times \{ (z_I(0 - \hat{P}_I \hat{P}_2) + z_2(\hat{P}_2 - \hat{P}_2 \hat{P}_2) + z_3(0 - \hat{P}_3 \hat{P}_2) + z_4(0 - \hat{P}_4 \hat{P}_2)) \} + \\ & \alpha_M \times \{ (z_I(0 - \hat{P}_3 \hat{P}_2) + z_2(0 - \hat{P}_4 \hat{P}_2) + z_3(\hat{P}_I - \hat{P}_I \hat{P}_2) + z_3(0 - \hat{P}_4 \hat{P}_2)) \} + \\ & \alpha_L \times \{ (z_I(\hat{P}_2 - \hat{P}_2 \hat{P}_2) + z_I(0 - \hat{P}_4 \hat{P}_2) + z_2(0 - \hat{P}_I \hat{P}_2) + \\ & z_2(0 - \hat{P}_3 \hat{P}_2) + (z_3(\hat{P}_2 - \hat{P}_2 \hat{P}_2) + z_4(0 - \hat{P}_I \hat{P}_2) + z_4(\hat{P}_2 - \hat{P}_2 \hat{P}_2) + z_4(0 - \hat{P}_3 \hat{P}_2)) \} \end{aligned}$$

For k=3:

$$\frac{\partial \text{Log } L_{CL}}{\partial \beta_3} = d_j \left\{ \left[\alpha'z \left[\frac{\partial \hat{P}_k}{\partial \beta_3} \right] \right] - \sum_j \left[\alpha'z \left[\frac{\partial \hat{P}_k}{\partial \beta_3} \right] \right] P_j \right\}, \text{ where:}$$

$$\begin{aligned} \left[\alpha'z \left[\frac{\partial \hat{P}_k}{\partial \beta_3} \right] \right] = & \alpha_H \times \{ (z_I(0 - \hat{P}_I \hat{P}_3) + z_2(0 - \hat{P}_2 \hat{P}_3) + z_3(\hat{P}_3 - \hat{P}_3 \hat{P}_3) + z_4(0 - \hat{P}_4 \hat{P}_3)) \} + \\ & \alpha_M \times \{ (z_I(\hat{P}_3 - \hat{P}_3 \hat{P}_3) + z_2(0 - \hat{P}_4 \hat{P}_3) + z_3(0 - \hat{P}_I \hat{P}_3) + z_3(0 - \hat{P}_4 \hat{P}_3)) \} + \\ & \alpha_L \times \{ (z_I(0 - \hat{P}_2 \hat{P}_3) + z_I(0 - \hat{P}_4 \hat{P}_3) + z_2(0 - \hat{P}_I \hat{P}_3) + z_2(\hat{P}_3 - \hat{P}_3 \hat{P}_3) + \\ & (z_3(0 - \hat{P}_2 \hat{P}_3) + z_4(0 - \hat{P}_I \hat{P}_3) + z_4(0 - \hat{P}_2 \hat{P}_3) + z_4(\hat{P}_3 - \hat{P}_3 \hat{P}_3)) \} \end{aligned}$$

Finally, for k=4:

$$\frac{\partial \text{Log } L_{CL}}{\partial \beta_4} = d_j \left\{ \left[\alpha'z \left[\frac{\partial \hat{P}_k}{\partial \beta_4} \right] \right] - \sum_j \left[\alpha'z \left[\frac{\partial \hat{P}_k}{\partial \beta_4} \right] \right] P_j \right\}, \text{ where:}$$

$$\begin{aligned}
[\alpha'z [\frac{\partial \hat{P}_k}{\partial \beta_4}]] = & \alpha_H \times \{ (z_I(0 - \hat{P}_I \hat{P}_4) + z_2(0 - \hat{P}_2 \hat{P}_4) + z_3(0 - \hat{P}_3 \hat{P}_4) + z_4(\hat{P}_4 - \hat{P}_4 \hat{P}_4)) \} + \\
& \alpha_M \times \{ (z_I(0 - \hat{P}_3 \hat{P}_4) + z_2(\hat{P}_4 - \hat{P}_4 \hat{P}_4) + z_3(0 - \hat{P}_I \hat{P}_4) + z_3(\hat{P}_4 - \hat{P}_4 \hat{P}_4)) \} + \\
& \alpha_L \times \{ (z_I(0 - \hat{P}_2 \hat{P}_4) + z_I(\hat{P}_4 - \hat{P}_4 \hat{P}_4) + z_2(0 - \hat{P}_I \hat{P}_4) + z_2(0 - \hat{P}_3 \hat{P}_4) + \\
& (z_3(0 - \hat{P}_2 \hat{P}_4) + z_4(0 - \hat{P}_I \hat{P}_4) + z_4(0 - \hat{P}_2 \hat{P}_4) + z_4(0 - \hat{P}_3 \hat{P}_4)) \}
\end{aligned}$$

References

- Bartel, A. P. 1979. "The Migration Decision: What Role Does Job Mobility Play." *American Economic Review* 69: 775-786.
- Blau, Peter and Otis D. Duncan. 1967. *American Occupational Structure*. New York: Wiley.
- Chattopadhyay, Arpita. 1998. "Gender, Migration, and Career Trajectories in Malaysia." *Demography* 35: 335-44.
- Comay, Yochanan. 1972. "The Migration of Professionals: An Empirical Analysis." *Canadian Journal of Economics* 5: 419-429.
- Dahlberg, Matz and Matias Eklöf. 2003. "Relaxing the IIA Assumption in Locational Choice Models: A Comparison between Conditional Logit Model, Mixed Logit, and Multinomial Probit Models." *Working Paper 2003: 9, Department of Economics Uppsala University*.
- Da Vanzo, Julie. 1983. "Repeat Migration in the United States: Who Moves Back and Who Moves on? Evidence from Micro Data." *Review of Economics and Statistics* 65: 552-559.
- Davies, Paul, Michael J. Greenwood, and Haizheng Li. 2001. "A Conditional Logit Approach to U.S. State-to-State Migration." *Journal of Regional Science* 41: 337-360.
- Ellis, Mark et al. 1993. "Migration Regions and Interstate Labor Flows by Occupation in the United States." *Growth and Change* 24: 166-190.
- Gani, Azmat and Bert D. Ward. 1995. "Migration of Professionals from Fiji to New Zealand: A Reduced Form Supply Demand Model." *World Development* 23(9): 1633-1637.
- Gleave, David and Derek Palmer. 1977. "The Relationship between Geographic and Occupational Mobility in the Context of Regional Economic Growth." in Hobcraft, John and Rees, Philip, ed., *Regional Demographic Development*. London: Billings and Sons.
- Goetz, Stephan J. 1999. Migration and Local Labor Market. *Web-Book of Regional Science*. Regional Research Institute, West Virginia University
- Greene, William H. 2000. *Econometric Analysis, Fourth Edition*. New Jersey: Prentice-Hall Inc.

- Hausman, J. A. and D. McFadden. 1984. "Specification Tests for the Multinomial Logit Model." *Econometrica* 52: 1219–1240.
- Kleiner, Morris. 1976. "The Interstate Migration of Professional and Technical Workers: A Cross-Sectional Analysis." *Review of Business and Economic Research* 12: 27-34.
- Krieg, R. G. 1997. "Occupational Change, Employer Change, Internal Migration, and Earnings." *Regional Science and Urban Economics* 27: 1-15.
- McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." in P. Zarembka, ed. *Frontiers in Econometrics*. New York: Academic Press.
- Mueller, Charles F. 1985. *The Economics of Labor Migration: A Behavioral Analysis*. New York: Academic Press.
- Murphy, K. and R. Topel. 1985. "Estimation and Inference in Two Step Econometric Models." *Journal of Business and Economic Statistics* 3: 370-379.
- Odland, John. 1996. "Migration and Occupational Choice Among Young Labor Force Entrants: A Human Capital Model." *Modern Classics in Regional Science* 5.
- Pashigian, Peter B. 1979. "Occupational Licensing and the Interstate Mobility of Professionals." *Journal of Law Economics* 22: 1–25.
- Schaeffer, Peter V. 1985. "Human Capital Accumulation and Job Mobility." *Journal of Regional Science* 25: 103–114.
- Schlottmann, A. and A. Herzog. 1984. "Career and Geographic Mobility Interactions for the Age Selectivity of Migration." *Journal of Human Resources* 19: 72–86.
- Schmidt and Strauss. 1975a. "The Prediction of Occupation Using Multinomial Logit Models." *International Economic Review* 16: 471–486.
- Schmidt and Strauss. 1975b. "Estimations of Models with Jointly Dependent Qualitative Variables: A Simultaneous Approach." *Econometrica*, 43: 745–755.
- Schroeder, Larry D. 1976. "Interrelatedness of Occupational and Geographical Mobility." *Industrial and Labor Relations Review* 29: 405–11.
- Sjaastad, Larry A. 1962. "The Costs and Returns of Human Migration." *The Journal of Political Economy* 70: 80–93.
- Train, K. 1986. *Qualitative Choice Analysis*. Cambridge: MIT Press.

Wilson, F. 1985. "Migration and Occupational Mobility: A Research Note." *International Migration Review* 19: 278–92.

Zimmer, G. Basil. 1973. "Migration and Changes in Occupational Compositions." *The International Migration Review* 7: 437-448.

Chapter 4

Comparing Conditional Logit, Nested Logit, and Mixed Logit State Destination Choice Models

4.1. Introduction

The increasing availability of individual data and the rapid advancement in computer technology have permitted researchers to analyze migration in new ways (for a detailed review, see Cushing and Poot (2004)). In this respect, one of the more recent models used in the migration literature is a conditional logit model, which examines migration choice as a multinomial discrete choice. Unlike an aggregate migration model, this conditional logit model is an individual model, thus it better represents migration as an individual's utility maximization decision. Moreover, this model can be manipulated for analyses that are not possible with aggregate models, such as incorporating individual characteristics as explanatory variables and computing cross elasticities of choosing among alternatives.

Application of a conditional logit model in the migration literature, however, has been very limited. Mueller (1985) was among the first to apply a conditional logit model to examine an individual's destination choice among states. Considering the state of the computer technology at that time, Mueller's contribution should be considered exceptional. Unfortunately, the migration literature did not immediately recognize the value of this model. Only in recent years has migration research, such as Davies et al. (2001), made good use of the conditional logit model.

Applying a conditional logit model, however, imposes the independence from irrelevant alternatives (IIA) assumption. This assumption implies that the probability ratio of individuals choosing between two alternatives does not depend on the availability or attributes of the other alternatives. This assumption can be realistic in some situations. For example, people who move for a job transfer typically have fixed their destination, and retirees may consider only one or two possible destinations in which they want to

live. For these people, any changes in the other destinations will not significantly affect their choice decision. In general, however, the IIA assumption is too restrictive, especially when the number of alternatives in the choice set is large (e.g., a model of state destination choice for the United States has about 50 alternatives).⁷

Violating IIA may lead a model to incorrectly predict the probability of destinations being chosen. The model may overestimate the probability of choosing California, while at the same time underestimating the probability of choosing another state. In light of this problem, several models have been developed to relax the IIA assumption, including nested logit, mixed logit, multinomial probit, and heteroscedastic extreme value models. They are more computationally complex than the conditional logit model. One direct consequence of relaxing the IIA assumption is additional time/computational cost for estimation.

This study applies two of the above models: nested logit and mixed logit. Even in this computer age, estimating these two models takes considerable time. In this study, while it took about 1.5 minutes for the conditional logit model to converge, it took more than 30 minutes to run the nested logit model, and nearly 10 hours to run the mixed logit model.⁸ This essay examines to what extent the outcomes of these two models differ from those of the conditional logit model. Based on the comparison, this study then assesses whether relaxing the IIA assumption warrants the application of the more complex nested logit or mixed logit models.

The next section gives a general introduction of the nested logit and mixed logit models. A more formal discussion on all discrete choice models follows. After a description of the econometric specification applied in this study, I analyze the results, comparing how

⁷ Statistically, the larger the number of alternatives, the higher the likelihood of finding at least one restricted model, which excludes one or more alternatives, that is significantly different from the unrestricted model, which includes all alternatives. Thus the easier it is to violate the IIA assumption.

⁸ Models that relax the IIA assumption other than these two are computationally more burdensome, thus have a significantly longer convergence time. Dahlberg and Eklöf (2003) found the convergence time for their multinomial probit models to be significantly longer than for their mixed logit models. This study attempted to apply a heteroscedastic extreme value model but it continually failed to converge.

the outcomes from the nested logit and mixed logit models differ from those of the conditional logit model.

4.2. Nested Logit and Mixed Logit Models

A nested logit model has a few desirable properties: (1) it has a closed form solution; (2) it is computationally easier than other models that relax the IIA assumption; and (3) its extension from the conditional logit model can reveal an appealing story on how individuals make choices. Newbold (1997), Frey et al. (1996), and White and Liang (1998) were among the first to apply nested logit models. However, they applied a limited instead of a full information nested logit model, by estimating the model sequentially. White and Liang (1998), for instance, first estimated a binomial logit model of move versus don't-move, and then estimated the movers' destination choice using a conditional logit model. This sequential nested logit model yields consistent but inefficient estimates (Green (2000)). Moreover, Hensher (1986) showed that its results are not comparable with those of the conditional logit model because they are derived from different sample sets. For example, in White and Liang's (1988) study, they estimate the destination choice model only for those who move, thus omitting nonmovers. A full information nested logit model would estimate the two decisions simultaneously, using the whole sample. This study applies a full-information nested logit model. Though requiring a more complex computer application, this model yields efficient and consistent estimates (Green (2000) and Hensher (1986)). In addition, the results can meaningfully be compared with those of the conditional logit model (Hensher (1986)).

Knapp et al. (2001) were among the first to apply a full information nested logit model in a migration study. Unlike Knapp et al. (2001), whose two-branch and one-level nested logit model examines a choice among two alternatives, this study examines a choice of 48 alternatives. Given the complexity of the model, a valid empirical result of this nested logit model should by itself make a significant contribution to the migration literature.

Nested logit models only partially, but not fully, relax the IIA assumption. This essay also considers a mixed logit model, which fully relaxes IIA. Compared with a nested logit model, a mixed logit model is even less recognized in the migration literature, as well as in most other literatures. Mainly, this is because a mixed logit requires an evaluation of multiple integrals rather than a single integral. Moreover, since it does not have a closed form equation, a mixed logit model is estimated through simulation. Only with major improvements in computer speed and in the understanding of simulation methods could we fully utilize this model. Dahlberg and Eklöf (2003) were among the first to apply a mixed logit model in a migration-related study. Examining migrants' choice among municipalities within a single metropolitan area, they compared conditional logit with mixed logit and multinomial probit models. They concluded that a well-specified conditional logit model can provide results that are qualitatively similar to those that relax the IIA assumption.

Although also comparing a conditional logit with other models that relax the IIA assumption, this study differs from that of Dahlberg and Eklöf (2003) in five respects. First, it compares the conditional logit with a nested logit and a mixed logit rather than with a multinomial probit and a mixed logit model. The nested logit can yield some different perspectives of the migration process. Second, this study examines an individual's state destination choice, which includes long distance or short distance migration rather than short distance intra-metropolitan residential choice. Third, the choice set consists of 48 choices rather than only 26 choices. Fourth, the sample size includes many more observations (11,431 compared with 1,444 individuals). Since a mixed logit model allows parameters to differ among individuals, significant differences in sample size significantly affects the complexity of the model. Finally, this study examines both non-movers and movers instead of only movers. The migration literature has argued that excluding nonmovers in a migration model may cause a selection bias problem.

4.3. Discrete Choice Models

Discrete choice models are based on utility maximization. In a destination choice model, this means that the chosen destination must give the individual greater utility compared with other destinations. If the utility of individual i choosing state j is represented as U_{ij} , then location j will be chosen if and only if $U_{ij} > U_{il}$ for $j \neq l$.

Because researchers do not know U_{ij} , the individual's true utility, they cannot tell for sure which destination an individual will eventually choose. U_{ij} consists of two components, the observable and the unobservable components:

$$U_{ij} = V_{ij} + \varepsilon_{ij}. \quad (3.1)$$

U_{ij} consists of a predicted utility, V_{ij} , observable based on the choice's attributes, and an unobserved random component, ε_{ij} . If ε_{ij} were known, researchers would know U_{ij} and could tell for sure which destination would be chosen. Since researchers do not know ε_{ij} , the best they can do is predict the final outcome in terms of probability.

The probability of individual i choosing state j can be described as:

$$\begin{aligned} P_{ij} &= P(U_{ij} > U_{il}) \\ &= P((V_{ij} + \varepsilon_{ij}) > (V_{il} + \varepsilon_{il})) \\ &= P((\varepsilon_{il} - \varepsilon_{ij}) > (V_{ij} - V_{il})) \text{ for all } j \neq l. \end{aligned} \quad (3.2)$$

Researchers can solve equation (3.2) if they know ε_{ij} 's probability density function. They then impose a certain distribution for ε_{ij} . Each type of probability distribution imposed on ε_{ij} leads to a different discrete choice model. Table 4.1 shows the different discrete choice models that arise from various distributions imposed on the error term.

Conditional Logit Model

The conditional logit model assumes that ε_{ij} exhibits the extreme value distribution. The probability density function takes the following form:

$$f(\varepsilon_{ij}) = e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}} \quad (3.3)$$

and its cumulative density function is expressed as:

$$F(\varepsilon_{ij}) = e^{-e^{-\varepsilon_{ij}}} \quad (3.4)$$

More importantly, this model restricts all ε_{ij} to be independent and identically distributed (iid). The probability of individual i choosing destination j can be solved as a closed-form expression of:

$$\begin{aligned} P_{ij} &= \frac{e^{V_{ij}}}{\sum_j e^{V_{ij}}} \\ &= \frac{e^{\alpha' Z_{ij}}}{\sum_j e^{\alpha' Z_{ij}}} \end{aligned} \quad (3.5)$$

Z_{ij} represents all the observed factors or explanatory variables and α represents parameters obtained from the model.

With ε_{ij} being iid, equation (3.5) imposes the IIA assumption. Consider the probability that individual i chooses state j or state l :

$$P_{ij} = \frac{e^{\alpha' Z_{ij}}}{\sum_j e^{\alpha' Z_{ij}}} \quad ; \text{ and } P_{il} = \frac{e^{\alpha' Z_{il}}}{\sum_j e^{\alpha' Z_{ij}}}$$

The probability ratio of choosing between j and l is:

$$\begin{aligned} \frac{P_{ij}}{P_{il}} &= \frac{\sum_j e^{\alpha' Z_{ij}}}{\sum_j e^{\alpha' Z_{ij}}} / \frac{e^{\alpha' Z_{il}}}{e^{\alpha' Z_{il}}} \\ &= \frac{e^{\alpha' Z_{ij}}}{e^{\alpha' Z_{il}}} \end{aligned} \quad (3.6)$$

That is, the probability ratio depends only on the attributes of j and l , and does not depend on the attributes of other destinations.

Nested Logit Model

A nested logit model relaxes the IIA assumption by allowing the unobserved factors, ε_{ij} , to be correlated. First, a nested logit model partitions choices into different subsets.

Based on the partition, a nested logit model then allows ε_{ij} to be correlated within each nest, but maintains independence across nests. In other words, a nested logit maintains IIA for choices within the same nest and relaxes it for choices across nests.

Let the set of all alternatives j be partitioned into K subsets. Each subset is called a nest and denoted B . Thus, all the alternatives are partitioned into $B_1, B_2, B_3, \dots$, and B_K , and each j is now an element of B_k , where k goes from 1 to K .

The utility of individual i choosing destination j is in general the same as that in the regular conditional logit model:

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$

However, to incorporate the nesting, the observed utility can be represented as consisting of two components: (1) A , which is constant for all alternatives within a nest but varies across nests; and (2) B , which varies over all alternatives within a nest. That is,

$$U_{ij} = A_{ik} + B_{ij} + \varepsilon_{ij}, \text{ where } j \in B_k. \quad (3.7)$$

A_{ik} depends only on variables that describe nest k (as will be seen in the nested logit model applied in this study, this is the part that depends on January Temperature) and B_{ij} depends on variables that describe the alternative j .

The nested logit model assumes that the ε_{ij} exhibit the generalized extreme value distribution with a cumulative joint distribution function described as:

$$F(\varepsilon_{ij}) = \exp\left(- \sum_{k=1}^K \left\{ \sum_{j \in B_k} e^{-(\varepsilon_{ij}/\lambda_k)} \right\}\right) \quad (3.8)$$

Equation (3.8) shows that the choices are partitioned into K subsets of B_k . λ_k is a parameter indicating the degree of substitutability between unobserved utility among

choices in different nests. When λ_k equals one, choices across nests are statistically independent, thus nesting becomes unnecessary. In that case, the cumulative distribution of ε_{ij} (equation (3.8)) collapses to that of a conditional logit model (3.5).⁹

With ε_{ij} 's cumulative distribution function following equation (3.8), the probability of individual i choosing destination j can be solved as a closed-form expression of:

$$P_{ij} = \frac{e^{V_{ij}/\lambda_k} (\sum_{j \in B_k} e^{V_{ij}/\lambda_k})^{\lambda_k - 1}}{K \sum_{k=1}^K (\sum_{j \in B_k} e^{V_{ij}/\lambda_k})^{\lambda_k}} \quad (3.9)$$

The probability ratio of individual i choosing between choice j and l is

$$\frac{P_{ij}}{P_{il}} = \frac{e^{V_{ij}/\lambda_k} (\sum_{j \in B_k} e^{V_{ij}/\lambda_k})^{\lambda_k - 1}}{e^{V_{il}/\lambda_d} (\sum_{j \in B_d} e^{V_{ij}/\lambda_d})^{\lambda_d - 1}} \quad (3.10)$$

Equation (3.10) shows that their probability ratio depends not only on the attributes of choices j and l but also on those of the other choices. On the other hand, if both choices belong to the same nest (that is $k=d$), their probability ratio becomes:

$$\frac{P_{ij}}{P_{il}} = \frac{e^{V_{ij}}}{e^{V_{il}}} = \frac{e^{\alpha'Z_{ij}}}{e^{\alpha'Z_{il}}} \quad (3.10)$$

That is, this ratio depends only on the characteristics of choices j and l , which is the property of the IIA assumption (the same as equation (3.5)).

Nesting Pattern in a Nested Logit Model

An important element in a nested logit model application is developing its nesting pattern. When researchers successfully set an acceptable nesting pattern, they obtain

⁹ For a more complete description of this model, see Train (2003). He also argued that, citing Kling and Herriges (1995) and Herriges and Kling (1996), even when λ_k is greater than 1, the nested logit model is still consistent with utility maximization. On the other hand, a negative λ_k implies inconsistency.

more information about the individuals' choice decisions than what they get from the standard parameter estimates. As an illustration, let's look at the nesting pattern used in this study, as shown in Figure 4.1.

The model nests destinations into two groups: warm states and cold states. When I successfully run this nested logit model I obtain two sets of information. First, as typically known from regression results, I learn about how each explanatory variable affects the probability of a destination being chosen. Second, I learn something about unobserved factors that correlate with warmness of the destination. These unobserved factors are not yet captured by the variable, temperature, which is used to distinguish warm states from cold states. These unobserved factors could include items such as the love for, or the need to avoid, snow or preference for specific seasonal recreational activities.

Looking at the nesting pattern, researchers are often tempted to interpret the nesting as a representation of sequential choice decisions. Such an interpretation means that in choosing their destination individuals would first make a selection based on a key attribute, which in this case is the warmness of the state. Afterwards, they choose their ultimate state destination only from the group they have selected in the first sequence. This implies that states like Alabama or Arizona are not available once an individual selects cold states. That does not appropriately represent reality. Regardless of their preference towards either cold or warm states, individuals still have some probability of choosing any of the states. It is true, however, that a strong preference toward cold states implies a lower probability of choosing any of the warm states.

It is important to emphasize that nesting is not intended to represent sequential decisions. The purpose of nesting is simply to categorize choices. When researchers suspect that the unobserved factors of a certain group of choices are correlated, they can categorize them into the same nest. The nesting basically puts alternatives with similar attributes into the same nest. The nesting pattern employed depends on the researcher's judgment, which could be based on natural consideration, casual observation, or theory. As long as

the researcher suspects the error terms among certain choices are correlated, he can exercise the nesting accordingly (Hensher (1986), Train et al. (1987), and Train (2003)).¹⁰ Since researchers may see different ways of how unobserved factors correlate, they may come up with more than one nesting pattern for the same choice decision, which in turn can yield different results.¹¹

The data determine whether or not a nesting is appropriate. A nesting pattern is acceptable when the parameter, λ , falls between 0 and 1. The value of $1 - \lambda$ measures the correlation among the unobserved components of utility within a nest (Train (2003)). When the value of λ falls between 0 and 1, the model is consistent with utility maximization for all possible values of explanatory variables (Ortuzar and Willumsen (1994) and Train (2003)). In this case, an improvement in the destination attribute will increase the probability of that destination being chosen. On the contrary, a negative λ is not consistent with utility maximization since it implies that an improvement in the destination attribute decreases the probability of that destination being chosen. If λ is greater than 1, an increase in the utility of a destination in the nest not only increases its selection probability but also the selection probability of the rest of the states in the same nest (Ortuzar and Willumsen (1994)). In this case, the model is only consistent with utility maximization for a certain range of values of explanatory variables (for details, see Herriges and Kling (1996), Kling and Herriges (1995), Lee (1999), and Train et al. (1987)). Ortuzar and Willumsen (1994) state that such a range is rarely available.

Mixed Logit Model

Unlike a nested logit, a mixed logit model fully relaxes the IIA assumption. This model is similar to a conditional logit model except that it allows parameter estimates to vary across individuals. Consider the utility function expressed in equation (3.1):

¹⁰ Train et al. (1987) explains “As in all nested logit models, the direction of conditionality reflects the correlations among unobserved factors across alternatives; as such it arises from patterns in the researcher’s lack of information, rather than from the household’s decision process” (p113).

¹¹ The discrete choice literature has not developed a well-defined methodology to determine which of the nesting patterns best represents reality (Green (2000)).

$$\begin{aligned} U_{ij} &= V_{ij} + \varepsilon_{ij} \\ &= \alpha'Z_{ij} + \varepsilon_{ij} \end{aligned}$$

Like a conditional logit, a mixed logit assumes the error terms, ε_{ij} , are iid with extreme value distribution. A mixed logit, however, relaxes the restriction that α is the same for each individual, allowing it to be stochastic instead. In a mixed logit model, the person's utility is

$$U_{ij} = \alpha_i'Z_{ij} + \varepsilon_{ij}, \quad (3.12)$$

where α now differs across individuals. Researchers can estimate U_{ij} if they know the probability density function (pdf) for α . Similarly, researchers can impose a certain type of distribution for α (e.g. normal, lognormal, uniform, etc.).¹² When α is assumed to be the same for all individuals as in the conditional logit, the probability of individual i choosing state j (P_{ij}) can be described exactly as in equation (3.5), used for the conditional logit model. When α is not fixed, the probability of individual i choosing destination j (in this case, labeled as M_{ij}) can be estimated by estimating P_{ij} over all the possible values of α .

$$\begin{aligned} M_{ij} &= \int P_{ij}(\alpha) f(\alpha) d\alpha \\ &= \int \left(\frac{e^{\alpha'Z_{ij}}}{\sum_j e^{\alpha'Z_{ij}}} \right) f(\alpha) d\alpha \end{aligned} \quad (3.13)$$

Thus, a mixed logit probability is a weighted average of the logit formula evaluated at different values of α , with the weights given by the density $f(\alpha)$. This equation is a multi-dimensional integral so that it does not have a closed-form solution. Consequently, this equation is solved through simulation.

¹² Train (2003) explains that when α is restricted to have a normal density function, the model becomes a close approximation of the multinomial probit model.

Another way to look at a mixed logit model is by representing the utility as an error component specification. α can be perceived as consisting of its mean, a , and a deviation around the mean, ξ , which differs across individuals. That is,

$$\begin{aligned} U_{ij} &= \alpha'Z_{ij} + \varepsilon_{ij} \\ &= (a + \xi_i)Z_{ij} + \varepsilon_{ij} \\ &= a'Z_{ij} + \xi_i'Z_{ij} + \varepsilon_{ij} \end{aligned} \quad (3.14)$$

In this case, ε_{ij} are still assumed to be iid. The unobserved components of the utility are $\eta_{ij} = \xi_i'X_{ij} + \varepsilon_{ij}$. In the conditional logit model, $\xi_i'Z_{ij}$ are identically zero, implying no correlation in utility across alternatives. With nonzero error components, $\xi_i'Z_{ij}$, utility becomes correlated across alternatives, which relaxes the IIA assumption.

Now that we understand the mathematical representation of a mixed logit model, one may ask what this implies in real life. By allowing α to vary across individuals, a mixed logit model can represent variations in the individual's utility function. Each individual now has his or her own value of α , implying that each person can have different weights for each destination attribute. A mixed logit model incorporates taste variations that realistically exist across individuals.

Multinomial Probit Model

The multinomial probit model assumes that ε_{ij} follows a multivariate normal distribution with covariance matrix Ω . That is,

$$\varepsilon_{ij} \sim N(0, \Omega), \text{ with } \Omega = I_J * \Sigma. \quad (3.15)$$

In this case, I is an identity matrix, and Σ is the covariance of ε_{ij} or $E(\varepsilon_j \varepsilon_j')$. The density of ε_{ij} is

$$f(\varepsilon_{ij}) = \frac{1}{(2\pi)^{J/2} |\Omega|^{1/2}} e^{-1/2 \varepsilon_{ij}' \Omega^{-1} \varepsilon_{ij}} \quad (3.16)$$

For example, for $J = 3$, Ω is

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

If Ω is a diagonal matrix, that is σ_{jl} are zeros for all $j \neq l$, then the ε_{ij} are independent or uncorrelated. If all the nonzero components of Ω have the same value, then the ε_{ij} are identical or *homoscedastic*. A multinomial probit model does not have either of these restrictions, thus it fully relaxes IIA. Under this assumption, the probability of individual i choosing destination j can be expressed as:

$$\begin{aligned} P_{ij} &= \text{Prob}(V_{ij} + \varepsilon_{ij} > V_{il} + \varepsilon_{il}) f(\varepsilon_{ij}) d\varepsilon_{ij} \\ &= \int \mathbf{I}(V_{ij} + \varepsilon_{ij} > V_{il} + \varepsilon_{il} \text{ for all } i \neq j) f(\varepsilon_{ij}) d\varepsilon_{ij}. \end{aligned} \quad (3.17)$$

$\mathbf{I}(\cdot)$ is an indicator of whether the statement in the bracket is accepted or rejected, and the probability for an alternative being chosen is the integral of the conditions over all the values of ε_{ij} . Since the components of Ω are not independent, equation (3.17) is a J -dimensional integral.¹³ Here lies the drawback of a multinomial probit model. Relaxing IIA entails significant computational burden, especially with a large number of alternative choices. Estimating a multinomial probit model must rely on simulation.

Researchers can also allow α to vary across individuals rather than be fixed. In this case, the normal distribution is imposed on α with a mean, α , and a deviation around the mean, ξ , which differs across individuals. Like the mixed logit model, this multinomial probit model can represent taste variations in the individuals' utility function.

¹³ In the estimation, the multinomial probit model measures utility in terms of differences in utility rather than level of utility. The error term is also represented as differences in the errors. By default, the difference of two normally distributed variables also has a normal distribution. Thus, instead of a J -dimensional integral, the estimation of a multinomial probit model is a $J-1$ dimensional integral.

$$\begin{aligned}
U_{ij} &= \alpha'Z_{ij} + \varepsilon_{ij} \\
&= (a + \xi_i)Z_{ij} + \varepsilon_{ij} \\
&= a'Z_{ij} + \eta_{ij}
\end{aligned} \tag{3.18}$$

Equation (3.18) is equivalent to equation (3.14) of the mixed logit model, with $\eta_{ij} = \xi_i'Z_{ij} + \varepsilon_{ij}$. The main difference is that in a multinomial probit model, the η_{ij} are restricted to follow a joint normal distribution, while in a mixed logit, the ε_{ij} are restricted to be iid with extreme value distribution but $\xi_i'Z_{ij}$ are allowed to follow any kind of probability distribution. When researchers restrict α in a mixed logit to follow the normal distribution, the model becomes a close approximation of the multinomial probit model (Dahlberg and Eklöf (2003) and Train (2003)).

Heteroscedastic Extreme Value Model

A heteroscedastic extreme value model also restricts the residuals, ε_{ij} , to follow an extreme value distribution. Unlike the conditional logit model, it assumes that the ε_{ij} , while they are independent, are heteroscedastic (not identical). This assumption fully relaxes IIA. Bhat (1995) argued that in real life, non-identical error variance is more realistic than identical variance. He used the transportation mode model to illustrate this point. If the unobserved factor in choosing the best transportation is the individual's level of comfort, then the variance of comfort from taking the train must differ from that of taking an automobile.

More specifically, Bhat (1995) restricts the density function for ε_{ij} to take the form of:

$$f(\varepsilon_{ij}) = \frac{1}{\theta_j} e^{-\varepsilon_{ij}/\theta_j} e^{-e^{-\varepsilon_{ij}/\theta_j}} \tag{3.19}$$

The θ_j are scale parameters that allow the value of ε_{ij} to differ across alternatives. When the θ_j are the same for all alternatives then the model reverts to the conditional logit model.

Given this distribution, the probability of individual i choosing destination j is

$$\begin{aligned}
P_{ij} &= \text{Prob}(V_{ij} + \varepsilon_{ij} > V_{il} + \varepsilon_{il}) \text{ for all } j \neq l \\
&= \text{Prob}(\varepsilon_{ij} < V_{ij} - V_{il} + \varepsilon_{il}) \text{ for all } j \neq l \\
&= \int \prod_{i,j} \Lambda \left[\frac{V_{ij} - V_{il} + \varepsilon_{il}}{\theta_j} \right] \frac{1}{\theta_j} \lambda \left[\frac{\varepsilon_{ij}}{\theta_j} \right] d\varepsilon_{ij} \quad (3.20)
\end{aligned}$$

In this case, $\lambda(\cdot)$ and $\Lambda(\cdot)$ represent the density and cumulative density function of the extreme value distribution. That is,

$$\lambda(x) = e^{-x} e^{-e^{-x}}; \quad \text{and } \Lambda(x) = e^{-e^{-x}}$$

Unlike a mixed logit or a multinomial probit, a heteroscedastic extreme value model requires only a one-dimensional integral. The estimation of this model is not stable, however, and computation of its likelihood function requires numerical integration (SAS Institute (2002)).

In summary, a researcher's distributional assumptions regarding the error components determine which discrete choice model is applied. Researchers can impose two types of characteristics on the error components: (1) their iid property and (2) their distribution function. Assuming the error components to be correlated, non-identical, or both relaxes the IIA assumption. Assumptions regarding the distribution function of the error components determine which discrete choice model should be used.

These models not only differ in terms of their mathematical representations. Different sets of characteristics of the error components imply different practical meaning. A mixed logit or a multinomial probit model allows choice to reflect taste variation. A nested logit model allows choice decisions to be categorized into different nests with each nest containing choices with a similar attribute. A heteroscedastic extreme value model allows choice decisions for which the variance of the unobserved factors of one alternative is different from the other alternatives. Three of the five possible models are

applied in this study: conditional logit, nested logit, and mixed logit models. The next section discusses the econometric specification for these three models.

4.4. Model Specifications

This study examines models of individuals' destination choice among U.S. states. The models assume that an individual chooses a destination that maximizes utility. The models describe the individual's utility as a linear function of destination attributes as well as individual characteristics. In general, the models take the form:

$$\begin{aligned} U_{ij} &= V_{ij} + \varepsilon_{ij} . \\ &= \alpha_i' Z_{ij} + \varepsilon_{ij} \end{aligned}$$

where Z_{ij} represents both destination attributes and individual characteristics.

Three models are applied: a conditional logit model, a nested logit model, and a mixed logit model. The conditional logit model follows that applied by Davies et al. (2001). This study, however, excludes Hawaii and Alaska from the choice set, and combines the District of Columbia with Maryland. Thus, individuals have to choose from among 48 states available, including the state of origin.

Unlike Davies et al. (2001), this study includes not only place attributes but also individual characteristics as explanatory variables. The place attributes include distance, employment growth, employment size, temperature, a dummy indicating the destination being adjacent to the origin, a dummy indicating the destination being a non-origin state, and dummies representing three U.S. regions (with the Northeast as reference), which attempts to account for some state fixed effects.

This study includes two individual characteristics, age and education. To incorporate these variables, I create interaction terms between the individual characteristics and certain place attribute variables. More specifically, this study interacts age with the non-origin dummy and education with distance. The first interaction term measures the effect

of age on the likelihood of choosing a non-origin state (making an interstate move), and the second measures how education affects the likelihood of choosing more distant states. Appendix A provides a more complete description of all explanatory variables included in the model.

Based on the findings in the migration literature, we should expect that the older the individual, the less likely they are to move, and that the higher the education the more willing they are to move to distant states. Thus, we expect the coefficient of the first interaction term to be negative and the second to be positive. Similarly, based on the migration literature, we should expect that individuals will more likely choose destinations that are closer, have a larger labor market, experience high employment growth, and have milder winters.

The second model applied is the nested logit model. After experimenting with different nesting patterns, two were acceptable. Recall that a nesting pattern is acceptable when its inclusive value parameters, λ_k , fall between 0 and 1. Figure 4.1 shows the first nesting specification. It shows that in choosing their destination, individuals consider the mild winters to be an important destination attribute. It also shows that some unobserved factors associated with milder winters are not captured by the January Temperature variable. These factors might include preference for snow, seasonal recreational activities, or natural landscape associated with milder winters.

Those who prefer warmer temperature will have a higher probability of choosing Alabama or West Virginia over choosing Colorado, Connecticut, North Dakota, or Wyoming. The same applies to those who prefer colder states. This nesting specification neither implies a sequential decision process nor exclusion from the choice set of states not in the preferred nest, e.g., North Dakota not being an available choice for someone who prefers warmer winters.

Figure 4.2 describes the second nesting specification, showing that individuals can partition the alternatives into coastal states and noncoastal states. The nesting implies

that those alternatives nested in the same nest, e.g., coastal states, exhibit correlated unobserved factors such as high preference for coastal related amenities.¹⁴

The third model applied is a mixed logit model. In this model, the parameter estimates for all explanatory variables are allowed to vary across individuals, unlike in the conditional logit model. In other words, in this model each individual is allowed to have his/her own utility function. For the mixed logit model, I impose a normal distribution. Thus, the results should be a close approximation to those of the multinomial probit model.

4.5. Data

The individual data come from the one percent Public Use Microdata Sample (PUMS) of the 2000 Census of Population and Housing. This study extracted one subsample (one percent) from the PUMS data, but excluded individuals who were enrolled in school or were less than 20 years old in 2000. In the end, the sample included 11,431 individuals. Using the PUMS data, interstate migration is defined as residing in a different state in 2000 than in 1995. Appendix B shows the average characteristics of individuals in the sample.

Data on employment size and employment growth come from the Regional Economic Information System. January Temperature data come from National Climatic Data Center (NCDC). The distance between origin and destination is based on longitude and latitude coordinates. If the destination is the same as the origin, then distance is zero. The remaining variables, dummy of non-origin, adjacency dummy, and region dummies were created based on U.S. maps.

4.6. Empirical Results

Table 4.2 shows the regression results of the three models. It took only 1 minute and 23 seconds for the conditional logit model to complete the regression, 35 minutes and 20

¹⁴ Note that while it has two nests, the second logit model shows as having only one λ_k . There actually are two λ_k s, but they are restricted to have the same coefficient. The likelihood ratio test confirms that the two λ_k are not statistically different.

seconds for the nested logit model, and 9 hours, 52 minutes, and 20 seconds for the mixed logit model. As expected, the more complex models took longer to complete the estimation. Relaxing IIA entails a significant time/computational cost.

The results also show that the more complex (least restrictive) the model the more efficient is its estimation, yielding a higher likelihood value. The mixed logit yields the highest log likelihood value of $-5,837$, followed by the nested logit model at $-6,019$, and the conditional logit at $-6,025$. Based on the likelihood-ratio tests shown in Table 4.3, the likelihood values of both the nested logit and mixed logit models are statistically different from that of the conditional logit model. Thus, the restrictions imposed on the nested logit and mixed logit models to yield the conditional logit model are rejected.

The next step is to compare parameter estimates. Hensher (1986) and Train (2003) indicate that the results of these three models are directly comparable. We start by looking at the results for the conditional logit model. With the exception of the South Region dummy, all estimated coefficients are statistically significant with the expected sign. Individuals are more likely to choose destinations that are closer or adjacent to the state of origin, provide a large labor market, and experience high employment growth. The interaction terms suggest that older individuals are less likely to move to another state and that more highly-educated people are more willing to move to a distant state. The results for the region dummies indicate that people are more likely to choose the Midwest but less likely to choose the West than to choose the Northeast region, all else equal. People have no preference between the South and Northeast regions.

The results from the nested logit models look similar to those from the conditional logit model. In general, all estimated coefficients of the two nested logit models have the same sign and most have the same level of statistical significance as the corresponding coefficients in the conditional logit model. Compared with the first nested logit model (warm vs. cold), the second nested logit model (coastal vs. noncoastal) yields results that are closer to the standard conditional logit model. In line with that, the log likelihood

value of the second model (-6,021) is also closer to the standard conditional logit (-6,025) than is the log likelihood of the first model (-6,019).

For the most part, the magnitudes of the coefficients are very similar. The fact that the coefficients of the inclusive values (λ_k) are close to one may actually indicate the closeness of their results. The only qualitative differences are the insignificance of the Midwest Region dummy and the somewhat lower significance level for January Temperature in the first nested logit model. The reduction in the magnitude and significance of the January-Temperature coefficient is possibly due to the inclusion of the nesting pattern, which in this case is explained by January-Temperature as well. In other words, part of the January Temperature effect was picked up by the nesting pattern, with its residual effect becoming smaller.

Comparing the mixed logit model with the conditional logit model is more straightforward because the two models have exactly the same explanatory variables. Once again, all estimated coefficients have the same sign and same level of statistical significance as the corresponding coefficient in the conditional logit model, with the exception of a lower significance level for the Midwest Region dummy. The coefficients in the mixed logit model are consistently higher (in absolute terms) than those from the conditional logit model, with some quite a bit larger. Whether yielding larger coefficients is a norm or just a coincidence is a subject for further study. Dahlberg and Eklöf's study (2003), for instance, does not find such a pattern.

In summary, while statistically different, the results of the three models are qualitatively very similar. There are no conflicting signs and the magnitudes of the coefficients are very close, with just a few exceptions. These findings agree with those of Dahlberg and Eklöf (2003).

4.7. Conclusions

With improvements in computer speed and better understanding of simulation, researchers can now examine migration using much more complex models that allow

researchers to better represent reality. Better computer technology has allowed this study to successfully estimate complex models that would have been unfeasible to estimate just a few years back. In addition to estimating the simpler conditional logit model, this study estimated nested logit and mixed logit models with 11,431 individuals in the sample and 48 alternatives in the choice set. The two more complex models required significant time and computational costs.

The three models differ in terms of their treatment of the IIA assumption. The conditional logit model carries the IIA assumption, the nested logit model partially relaxes it, and the mixed logit model fully relaxes it. This study has compared their results, then assessed whether the need to relax the IIA assumption warrants application of the more complex nested logit or mixed logit models.

The results of these three models, while statistically different, are qualitatively very similar. The parameter estimates of the three models are of the same sign, generally of comparable statistical significance, and, with few exceptions, of comparable magnitude. Train (2003) suggested that the results of a conditional logit can often be used as a general approximation of models that relax IIA. He further suggested that which model researchers should use depends on the goals of their research. When researchers are more concerned with knowing the individuals' average preferences, violating IIA may not be much of an issue and the relatively simple conditional logit model should suffice. IIA becomes a serious issue when researchers attempt to forecast the substitution patterns among the alternatives, e.g., if researchers need to forecast how much the demand for alternative A would change due to changes in its characteristics or the characteristics of other choices.

The more thoroughly a researcher specifies a conditional logit model, the more likely that it will serve as a good approximation, regardless of the intended use. One way to improve a conditional logit model is by incorporating more individual characteristics into the model. This would let the model capture some effects of taste variations that a mixed logit or a multinomial probit model usually captures. Note that the IIA assumption, now

perceived as a restrictive assumption, was originally perceived as a natural outcome of a well-specified conditional logit model that captures all sources of correlation over alternatives. That is, a well specified conditional logit model would yield results where the residuals are independent and identical (Train (2003)). With computer technology continuing to advance, perhaps one day, researchers will not need to consider the problem of trading off between realism and computational cost.

Table 4.1

The Relationship between Probability Distribution Imposed on the Random Components
and Type of Discrete Choice Model Applied

Model Type	Utility Function	Assumption on the random components
Conditional Logit (CL)	$U_{ij} = \alpha'Z_{ij} + \varepsilon_{ij}$	ε_{ij} are iid & follow extreme value distribution function.
Nested Logit (NL)	$U_{ij} = \alpha'Z_{ij} + \varepsilon_{ij}$ Since $Z_{ij} = A_{ij} + B_{ij} \rightarrow$ $U_{ij} = \alpha'A_{ij} + \alpha'B_{ij} + \varepsilon_{ij}$ A are variables that vary across nest; B vary within nest.	ε_{ij} are correlated among choices within the same nests, but are iid among choices across nest. ε_{ij} follow generalized extreme value distribution function.
Mixed Logit (ML)	$U_{ij} = \alpha_i'Z_{ij} + \varepsilon_{ij}$ Since α vary across individuals $\rightarrow \alpha = a + \zeta_i$, where a = mean of α . $\rightarrow U_{ij} = a'Z_{ij} + \zeta_i'Z_{ij} + \varepsilon_{ij}$.	ε_{ij} are iid & follow extreme value distribution. $\zeta_i'Z_{ij}$, however, are correlated & can follow different kinds of distributions.
Multinomial Probit (MNP)	$U_{ij} = \alpha'Z_{ij} + \varepsilon_{ij}$ Like mixed logit, this model can also allows α to vary across individuals $\rightarrow$ $U_{ij} = a'Z_{ij} + \zeta_i'Z_{ij} + \varepsilon_{ij}$. $= \alpha'Z_{ij} + \eta_{ij}$	ε_{ij} are correlated & follow normal distribution. In this second case, η_{ij} are correlated & follow joint normal distribution. If the $\zeta_i'Z_{ij}$ are correlated and follow normal distribution, iid ε_{ij} still give rise to a MNP model.
Heteroscedastic Extreme Value (HEV)	$U_{ij} = \alpha'Z_{ij} + \varepsilon_{ij}$	ε_{ij} are independent but not identical (not <i>homoscedastic</i>) and follow the heteroscedastic extreme value distribution.

Source: Summarized from Dahlberg and Eklöf (2003), SAS Institute (2002), and Train (2003).

Table 4.2
Regression Results:
Comparison among Conditional Logit, Nested Logit, and Mixed Logit Models

Explanatory Variables	Conditional Logit	Nested Logit-1 (Warm vs Cold)	Nested Logit-2 (Coastal vs Non-Coastal)	Mixed Logit
Age*Non-Origin	-0.0922 *** (-36.94)	-0.0948 *** (-36.04)	-0.0937 *** (-36.06)	-0.15 *** (-13.24)
Education*Distance	0.0119 *** (12.18)	0.0123 *** (11.96)	0.0121 *** (12.09)	0.0136 *** (10.44)
South Region Dummy	-0.1260 (-1.25)	-0.0895 (-0.75)	-0.116 (-1.13)	-0.1464 (-0.99)
West Region Dummy	-0.2948 ** (-2.40)	-0.2845 ** (-2.17)	-0.2704 ** (-2.17)	-0.487 ** (-2.57)
Midwest Region Dummy	0.1709 ** (2.01)	0.1379 (1.61)	0.1968 ** (2.27)	0.2287 * (1.79)
Distance	-0.2386 *** (-18.5)	-0.2431 *** (-17.91)	-0.2417 *** (-18.29)	-0.2956 *** (-10.54)
Employment Growth	0.1362 *** (3.40)	0.1362 *** (4.83)	0.1357 *** (3.36)	0.1483 *** (2.76)
State's Employment Share	0.1240 *** (10.47)	0.1546 *** (14.87)	0.1279 *** (10.67)	0.1494 *** (10.53)
Adjacent Dummy	0.8967 *** (9.86)	0.9830 *** (10.39)	0.9332 *** (10.04)	0.94 *** (8.40)
January Temperature	0.0175 *** (3.98)	0.0139 ** (2.33)	0.018 *** (4.03)	0.0249 *** (4.35)
λ_k Cold States	-----	0.9279 * (1.41)	-----	-----
λ_k Warm States	-----	0.8669 *** (2.60)	-----	-----
λ_k Coastal/Non-Coastal	-----	-----	0.9382 *** (2.90)	-----
Log Likelihood	-6,025	-6,019	-6,021	-5,837
Covergence Time	1:23	35:20	31:07	9:52:29
Sample Size	11,433	11,433	11,433	11,433

Note: *** indicates significance at the one-percent level; ** indicates significance at the five-percent level; and * indicates significance at the ten percent level.
T-statistics are in parentheses. The t-statistics for λ_k are for testing $H_0: \lambda_k \geq 1$; and $H_1: \lambda_k < 1$.

Table 4.3
Likelihood Ratio Tests

	Nested Logit	Mixed Logit
Log Likelihood	-6,019	-5,837
Likelihood Ratio Test Statistics	12	376
Degree of Freedom	2	10
Critical Value (5% level)	5.99	20.48
Decision	Reject Ho	Reject Ho

Note: Ho = that the nested logit or the mixed logit produce the same results as the conditional logit model (for a more complete description of this test, see Green (2000) and Econometrics Laboratory (2000)).

Figure 4.1
Nesting Specification of the First Nested Logit Model

Figure 4.2
Nesting Specification of the Second Nested Logit Model

Appendix 4.1
Explanatory Variables

Name of Variable	Explanation
Distance	Distance, in thousands of miles, between the center of the origin and destination states.
Employment Share	Average of state's total employment divided by the U.S. total employment, 1995-1999.
Employment Growth	State's average annual employment growth, 1995-1999.
January Temperature	Area-Weighted State Average January Temperature, 1971-2000 (National Climatic Data Center).
Adjacent Dummy	A dummy taking a value of 1 if a destination state is adjacent to the origin state, and 0 otherwise.
Dummy of Non-Origin	A dummy taking a value of 1 if destination is not the same as the origin, and 0 otherwise.
Age*Non-Origin	An interaction term between Age and the dummy of non-origin.
Education*Distance	An interaction term between educational attainment and distance.
Dummies of Region	Dummies for each of the four U.S. regions. Three dummies are included in the model, with Northeast as the reference region.

Appendix 4.2
Characteristics of the Individuals in the Sample

Variable	Mean	Standard Deviation	Minimum Value	Maximum Value
Age	42.22	12.43	20	93
Years of Education	10.32	2.66	1	16

References

- Bhat, Chandra R. 1985. "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice." *Transportation Research B* 29: 471-483.
- Cushing, Brian and Jacques Poot. 2004. "Crossing Boundaries and Borders: Regional Science Advances in Migration Modelling." *Papers in Regional Science* 83: 317-338.
- Dahlberg, Matz and Matias Eklöf. 2003. "Relaxing the IIA Assumption in Locational Choice Models: A Comparison between Conditional Logit Model, Mixed Logit, and Multinomial Probit Models," *Working Paper 2003: 9, Department of Economics Uppsala University*.
- Davies, Paul, Michael J. Greenwood, and Haizheng Li. 2001. "A Conditional Logit Approach to U.S. State-to-State Migration." *Journal of Regional Science* 41: 337-360.
- Econometrics Laboratory. 2000. *Qualitative Choice Analysis Short Course*. California: University of California, Berkeley (available at http://elsa.berkeley.edu/eml/qca1_archive.shtml).
- Frey William H., K. Liaw, Y. Xie, and M. Carlson. 1996. "Interstate Migration of the U.S. Poverty Population: Immigration 'Pushes' and Welfare Magnet 'Pulls'." *Population and Environment* 17: 491-536
- Greene, William H. 2003. *Econometric Analysis, Forth Edition*. New Jersey, USA: Prentice-Hall.
- Hensher, David A. 1986. "Sequential and Full Information Maximum Likelihood Estimation of a Nested Logit Model." *Review of Economics and Statistics* 68: 657-667.
- Herriges, J. and C. Kling. 1996. "Testing the Consistency of Nested Logit Models with Utility Maximization." *Economics Letters* 50: 33-39.
- Kling, C. and J. Herriges. 1995. "An Empirical Investigation of the Consistency of Nested Logit Models with Utility Maximization." *American Journal of Agricultural Economics* 77: 875-884.
- Knapp, Thomas A., Nancy E. White, and David E. Clark. 2001. "A Nested Logit Approach to Household Mobility." *Journal of Regional Science* 41: 1-22.
- Lee, Bosang. 1999. "Calling Patterns and Usage of Residential Toll Service under Self Selecting Tariffs." *Journal of Regulatory Economics* 16: 45-81.

- Manski, C. and D. McFadden, eds. 1981. *Structural Analysis of Discrete Choice Data with Econometric Application*. Cambridge, USA : MIT Press.
- Mueller, Charles F. 1985. *The Economics of Labor Migration, A Behavioral Analysis*. New York: Academic Press.
- Newbold K. B. 1997. "Primary, Return and Onward Migration in the U.S. and Canada: Is There a Difference?" *Papers in Regional Science* 76: 175-198.
- Ortuzar, J. and L.G. Willumsen. 1994. *Modelling Transport, Second Edition*. New York: Wiley.
- SAS Institute. 2002. "The MDC Procedure," *SAS Online Documentation 9.1.3*. (available at <https://bearhaus.baylor.edu/SASdoc/docMainpage.jsp>).
- Train, K. 1986. *Qualitative Choice Analysis*. Cambridge, MA: MIT Press.
- Train, K. 2003. *Discrete Choice Method with Simulation*. Cambridge, MA: Cambridge University Press.
- Train, K., D. McFadden, and M. Ben-Akiva. 1987. "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices." *Rand Journal of Economics* 18: 109-123.
- White, M. and Z. Liang. 1998. "The Effect of Immigration on the Internal Migration of the Native-born Population, 1981–1990." *Population Research and Policy Review* 17: 141–166.

Chapter 5

Conclusions

The three essays of this dissertation further our understanding of migration by looking into two of its most important roles: (1) promoting regional wage convergence and (2) increasing economic efficiency through the sorting of workers and, indirectly, businesses as they search for the best location. The first essay examines why migration and regional wage differentials can co-exist over time while classical economic theory states that migration should eliminate wage differentials. The direct reason for the persistent differentials is that the classical assumptions imposed are rarely fully applicable in the real world.

In this first essay I examine various real world conditions that violate the classical assumptions and explain how each affects the interaction between migration and regional wage convergence. Each condition can contribute to persistent regional wage differentials in one of two ways: widening or creating wage differentials. Two conditions may widen wage differentials. If destination regions benefit from increasing returns to scale operation or agglomeration, or if migration selects more highly-skilled individuals, then migration would further reinforce economic activity in the destination, widening the wage gap between origins and destinations. The other conditions may create wage differentials. They include significant migration costs, differences in productivity factors, compensating differentials, differences in the extent of wage rigidity, aggregation bias, sex and racial discrimination, and slow market adjustments. Everything else considered, regional wage differentials can exist to the extent that they represent migration costs, wage rigidity, differences in productivity factors, and so on across regions.

That various components can make up the observed regional wage differentials raises a measurement issue. Researchers cannot identify exactly which components constitute the observed wage differentials. In order to empirically test the role of migration, researchers must identify if a true wage disequilibrium exists within the observed wage differentials

and, if so, how large. This is necessary because migration's equilibrating role is expected to work on wage disequilibrium but not necessarily on wage differentials. If researchers fail to control for the other components, then test results might falsely conclude that migration fails to promote regional wage convergence. These observed wage differentials may reflect factors such as migration costs or differences in productivity rather than true regional wage disequilibrium.

Further complication arises when considering that migration is an individual decision. Individuals maximize utility rather than wages. While utility includes command over goods and services, it also can include factors such as job security, job prestige, opportunity for future career advancement, place amenities, and proximity to relatives and favorite places. In other words, utility disequilibrium, not necessarily wage disequilibrium, drives individuals to migrate. Unfortunately, it is difficult for researchers to distinguish one from the other, empirically.

If all the measurement issues were resolved, would we be able to prove that migration equilibrates regional wage disequilibrium? Two opposing forces are at work. First, through its impact on supply and demand in the labor market, migration should reduce wage disequilibrium, as predicted by Neoclassical theory. Second, considering its impact on agglomeration and workers selectivity, migration could widen wage disequilibrium. In sum, migration's net effect on wage disequilibrium depends on the relative strength of these two forces.

To improve its ability to test migration's effect on wage convergence and wage disequilibrium, a study should limit its scope to a specific type of job or detailed industry. Such studies could also compare whether migration's role differs in different industries. For instance, they can test whether wages converge more readily in more competitive labor markets (such as in retail trade industries) than in less competitive ones (such as in manufacturing sectors).

I also suggest that future studies may need to view migration as a dynamic rather than a one-point-in-time event. Empirical facts show that migration does not necessarily go one way, but often works as two-way flows between regions that take time to play out completely. Considering the dynamic nature of migration, it might be more interesting to investigate the impact of migration in a bigger picture. For instance, it would be interesting to see whether migration occurs more between similar rather than between contrasting regions and whether wage convergence occurs more quickly and completely between regions long involved in back-and-forth migration.

In the second essay I examine the role of migration in allowing workers and businesses to search for the best location. In the process of searching for the best location, an individual's occupation plays an important role. Firms look for workers whose occupational skills best fit their operation and individuals look to find firms that maximize the returns to their occupational skills. Thus, in this essay I examine how individuals' choice of occupation affects their choice of destination. Unlike previous studies, this research recognizes that some people choose their occupation and destination simultaneously. Therefore, the analysis applies a two-stage maximum likelihood estimation, which has not been applied in migration studies.

As expected, I find that individuals are more likely to choose destinations where they can maximize the returns to their occupational skills. After adjusting the variance, I find that the statistical results do not vary qualitatively, compared with the unadjusted results. This, however, does not imply that the adjustment is unnecessary. In some other situations, the adjustment may significantly alter statistical conclusions. Moreover, since individuals frequently face simultaneous choices, this methodology would also be useful for other individual-level studies.

In the second essay I apply a conditional logit model to examine the destination choice among U.S. states. This model, however, violates the IIA property it imposes, which can cause the model to incorrectly predict the probability of a destination being chosen. In the third essay, I apply two other models: a nested logit, which partially relaxes IIA, and

a mixed logit, which fully relaxes IIA. Unfortunately, the models that relax the IIA assumption are much more complex. Estimating the nested logit and mixed logit models incurs significant time and computational costs. I then examine how much the outcomes of these two models differ from those of the standard conditional logit model. Based on the comparison, I assess whether relaxing the IIA assumption warrants application of the more complex nested logit or mixed logit models.

The results of these three models, while statistically different, are qualitatively very similar. The parameter estimates of the three models are of the same sign, of comparable statistical significance, as well as of comparable magnitudes. Almost all variables maintain their statistical significance, in terms of not changing from being statistically significant to insignificant or otherwise. The results suggest that if researchers are more concerned with knowing average preferences, the standard conditional logit model should provide a good approximation. On the other hand, if they plan on using the results to forecast substitution patterns among alternatives, a model that relaxes IIA would be a safer choice.

When choosing to use the standard conditional logit model, researchers can always improve the model specification. For example, researchers can incorporate more individual characteristics into the model as individual data become increasingly available. This would enable the model to capture some of the taste variations that a mixed logit model captures. Note that the IIA assumption, now perceived as a restrictive assumption, was originally perceived as a natural outcome of a well-specified conditional logit model that captures all sources of correlation over alternatives. That is, a well-specified conditional logit model would yield results where the residuals are independent and identical.

The two empirical essays suggest some obvious extensions. First, it would be interesting to merge the themes of these two essays by developing an appropriate simultaneous equations methodology within a model that relaxes the IIA assumption. Second, it would be helpful to check for robustness of the findings from the two empirical studies by

retesting them with different data. Of particular interest would be confirmation that a well-specified conditional logit model of migration would yield a good approximation to results from more complex models that relax the IIA assumption. Third, as Chapter Four indicates, the nested logit and mixed logit models are not the only alternatives that relax the IIA assumption. The multinomial probit and heteroscedastic extreme value models are other well-known choices. Each of these models comes with somewhat different distributional assumptions. For completeness, it would be useful to compare results for all of these choices. The last two models, however, may prove exceedingly difficult to solve with a dependent variable that allows 48 possible destination choices.