DISSERTATION

THREE ESSAYS IN REGIONAL ECONOMICS: MIGRATION, REGIONAL PORTFOLIO THEORY, RESILIENCE, AND AGGLOMERATION ECONOMIES

Submitted by

Jonathan Charles Care

Department of Economics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2023

Doctoral Committee:

Advisor: Stephan Weiler

Elissa Braunstein
Anita Alves Pena
Dawn Thilmany

ABSTRACT

THREE ESSAYS IN REGIONAL ECONOMICS: MIGRATION, REGIONAL PORTFOLIO THEORY, RESILIENCE, AND AGGLOMERATION ECONOMIES

Cities and counties are dynamic entities that experience constant change, generated from both local and external forces. Some locations are rich in natural amenities, others are powerhouses of manufacturing or provide a rich level of services and quality of life to their citizens. Each location maintains a unique set of characteristics that makes it appealing to a given slice of the population and set of business enterprises. Understanding these characteristics and patterns of migration is a substantial focus in the field of regional economics.

Researchers attempt to enhance our understanding by examining this phenomenon through a number of different lenses. Some have examined flows to the largest cities in the country and tried to uncover the underlying reasons for the unique advantages these metropolitan areas possess. Others have examined various measures of risk and reward to see which cities or counties outshine their competitors. Still others attempt to measure the appeal of regions by quantifying their natural amenities or investigating their resilience to negative economic events. Recent global events have brought understanding a number of these regional performance topics to the forefront of both academic and mainstream interest.

This dissertation examines several aspects of regional economics with an aim to move the conversation forward along several tracks. The first chapter explores the contribution of regional employment portfolio risk and return measures in a case study of county level migration into Colorado. The level of employment data used in the construction of the employment portfolio measures is varied to see which level of aggregation best contributes to the understanding of migration flows. The results show

that the employment portfolio composition of a county does play a role in attracting migrants and highlight interesting findings on economy-wide risk versus individual potential returns. Additionally, we find evidence of labor pooling and agglomeration effects for Colorado's largest counties. A lack of cohesion and consistency across sector-level measures of risk and return suggests that local governments should focus on creating a stable overall business environment, rather than attempting to focus on specific sectors.

The second chapter discusses the concept of economic resilience and how it complements discussions relating to regional economic growth. A total of seven models are tested, split between two different formulations of measuring resilience. Testing is performed to identify a set of independent variables that robustly contribute as explanatory determinants of resilience. The results identify several determinants of resilience that are robust across different definitions of economic resilience and provide insights that can be used by local policy makers when considering the tradeoffs between balancing growth and resilience. The chapter ends with a discussion of the strengths and weaknesses of existing measures of resilience and the advantages of future work in this area.

The final chapter of the dissertation examines the dual, decades-long decline in both migration rates and the level of economic dynamism within the United States. Specifically, the role of information generated by the churn of resources through the economy is explored within the context of county and metropolitan statistical area (MSA) in-migration rates. The difference in average annual in-migration rates is also examined using a three-fold Blinder-Oaxaca decomposition. This study finds that locally generated information on dynamism does contribute to the decision of whether to migrate. In particular, the findings show a unique role for information gained from regional dynamism when considering migration to smaller metropolitan areas, likely resulting from the more homogeneous identity these regions maintain, in comparison to larger, more multi-faceted, metropolitan areas.

The overarching goal of this work is to contribute to the literature on why individuals choose to live where they do. The topics examined over the course of this dissertation permeate several veins of the regional economic literature. However, they all work together in the service of the question "what makes a place attractive to in-migrants?" This is accomplished by looking at the risks and returns to regional employment portfolios, the degree to and speed with which regions rebound from recessions, and how information generated by the churn of resources through the economy helps in the decision to migrate. These topics represent three of the drivers among the broad portfolio of factors regional economics utilizes to try and understand behavior within a country.

ACKNOWLEDGEMENTS

I am extremely grateful to all the members of my dissertation committee for their support and guidance over the course of this endeavor. Without your assistance this project would not have been possible.

I would like to express my sincere appreciation to my advisor, Stephan Weiler for his years of support and innumerable hours spent on the phone. Our weekly chats have been a mainstay on my calendar for the past half decade. Thank you for sharing your knowledge freely, about both academia and life. I would also like to thank Elissa Braunstein, who has supported me from the very beginning, even when I completely changed directions. Your support has been instrumental to finishing this project. I would like to thank Anita Alves Pena, who was always available to provide a second set of eyes and to challenge me to consider all sides of my argument. Your feedback has made me a more rigorous researcher. I would also like to thank Dawn Thilmany for her support and the advice to workshop ideas with a larger audience to benefit from diverse viewpoints.

I would like to thank the entire Economics Department staff and faculty for their enduring support of the course of my studies. An additional thank you to my peers for the laughs and lessons learned over countless hours in the office.

Lastly, this endeavor would not have been possible without the endless support of my family. I am extremely grateful to my parents, Brenda and Charles, for raising me with a value on education and supporting me over the years. Thank you to my children, Emalyn and Declan, for understanding why their father was doing schoolwork on the weekends. Finally, thank you to my wife, Carrie, for going on this adventure with me, packing up our lives, and moving to Colorado. This dissertation would not exist without your boundless support, and I am infinitely grateful for all you have done.

TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEDGEMENTS v
LIST OF TABLES viii
LIST OF FIGURES ix
Introduction 1
Chapter 1 - Examining Colorado county level in-migration rates: an employment portfolio approach 5
1.1 Introduction 5
1.2 Migration and Regional Portfolio Theory Literature 5
1.3 Methodology 13
1.4 Data Description 18
1.5 Results 21
1.6 Conclusions 36
Chapter 2 - A Review and Comparison of Measures of Economic Resilience 39
2.1 Introduction 39
2.2 Economic Resilience Literature 40
2.3 Data Description 45
2.4 Results 52
2.5 Discussion 61
2.6 Conclusions 64
Chapter 3 - Examining how information generated by economic dynamism influences in-migration decisions 66
3.1 Introduction 66
3.2 Migration and Economic Dynamism Literature 68
3.3 Methodology 72
3.4 Data Description 74
3.5 Empirical Framework 76
3.6 Results 78
3.7 Blinder-Oaxaca Decomposition 82
3.8 Single-County and Multi-County MSA Comparison 88
3.9 Conclusions 92

Conclusion 95
Appendix A 106
Appendix B 108

LIST OF TABLES

Table 1-1: Colorado county level descriptive statistics 20
Table 1-2: Aggregate portfolio model results 22
Table 1-3: 2003 and 2013 rural-urban continuum code (RUCC) criteria 23
Table 1-4: Crosstab of migration total by county origin and destination type 24
Table 1-5: Population scale cuts 24
Table 1-6: Aggregate model results - population cut scale version 2 25
Table 1-7: Marginal effects for all counties model 28
Table 1-8: Percentage of moves by distance between counties across receiving county size 29
Table 1-9: BLS Supersector names and corresponding component 2-digit NAICS 30
Table 1-10: Supersector model results 31
Table 1-11: Full CO Supersector panel - additional independent variables, population cut scale version 2 32
Table 1-12: Expanded model marginal effects 33
Table 1-13: NAICS 51 component industries 34
Table 2-1: Descriptive statistics for dependent variables 47
Table 2-2: Independent variables and economic themes 48
Table 2-3: BLS Supersector composition by 2-digit NAICS code 49
Table 2-4: Descriptive statistics for independent variables 50
Table 2-5: U.S. Census divisions and component states 51
Table 2-6: Variance Inflation Factors (VIF) test results 51
Table 2-7: Variance-Mean Ratio model results 55
Table 2-8: Impact model results 58
Table 2-9: Multivariate regression for independent variable cross model significance 60
Table 3-1: Descriptive statistics for metro county level of analysis 77
Table 3-2: Descriptive statistics for MSA level analysis 78
Table 3-3: Metro county model results 78
Table 3-4: MSA model results 80
Table 3-5: F-tests for joint significance 81
Table 3-6: T-test for difference in means between MSA type in-migration rates 83
Table 3-7: Blinder-Oaxaca Test Results 85
Table 3-8: Single-county MSA results 89
Table 3-9: Multi-county MSA results 91
Table A-1: RUCC model results 106
Table A-2: Full Supersector model results 107

LIST OF FIGURES

Figure 1-1: Heatmap of Colorado county total in-migration counts (2001-2016) 23
Figure 1-2: Scatterplot of aggregate measures of risk and return, grouped by population cut version 2 .. 26
Figure 1-3: County-level employment portfolio returns 26
Figure 1-4: County-level employment portfolio risk 26
Figure 1-5: Annual percentage change in employment - Information Supersector 35
Figure 2-1: Correlation matrix for independent variables 53
Figure 2-2: Three example VMRs with a value of seven 62
Figure 3-1: U.S. annual mobility rates (1947-2016) 68
Figure 3-2: U.S. firm entry and exit rates (1978-2016) 82
Figure 3-3: Average annual in-migration rate by MSA type 82
Figure 3-4: Endowment results graph 86
Figure 3-5: Coefficients results graph 87
Figure 3-6: Interaction results graph 88

Introduction

The COVID-19 pandemic catalyzed a cultural shift unequaled in recent memory. Over the course of 2020, the United States experienced stay at home orders, a significant move to remote work and schooling, and a decrease in economic activity that disproportionately affected small and medium sized businesses (Fairlie, 2020). The pandemic coincided with record lows for mortgage rates (Freddie Mac, n.d.) and the monthly supply of houses in the U.S. fell to the lowest value on record (U.S. Census Bureau & U.S. Department of Housing and Urban Development, n.d.). All these occurrences raised new discussions about where people would prefer to live and what makes a given city attractive. While substantial noise was made in the news about large outflows from cities into more micropolitan areas, recent publications from both the Census Bureau (Toukabri & Delbé, 2022) and the Chicago Federal Reserve Bank (Lavelle & Kepner, 2022) show that pandemic era migration within the U.S. was not substantially different from pre-pandemic patterns.

Data from the U.S. Census Bureau shows that migration rates within the U.S. have been steadily decreasing since the 1980s. Additionally, the level of churn experienced with respect to the creation and destruction of businesses has followed a parallel decades-long decline. U.S. communities continue to experience a wide range of outcomes related to the vitality of their economy, including the degree of economic setbacks and the speed with which they recover. In some cases, this pandemic has simply laid bare underlying concerns that have existed among economists for several decades.

In the following chapters, we examine several of the intertwined trends that surround the regional economic discussion of what makes a location an attractive place to live. The first chapter of the dissertation takes a detailed view by examining county to county migration flows into Colorado counties. We add to the employment portfolio literature by following the approach used in Low & Weiler (2012) and Gulati & Weiler (2021) to extend the use of employment portfolios as explanatory variables. The empirical investigation utilizes detailed employment data from the Quarterly Census of Employment and

Wages (QCEW)[1] to construct measures of risk and reward using the Supersector groupings provided by the Bureau of Labor Statistics (BLS). These measures are then tested in a Pairwise Poisson regression with fixed effects for the originating counties and control variables for the Colorado destination counties in the spirit of Marr et al. (2019). The results show that the employment portfolio composition of a county does play a role in attracting migrants and that the impact of individual sectors varies widely. Additionally, a lack of consistent statistical significance on the Supersector groupings suggests that these groupings may be best saved for industry specific case studies, while the economy-wide aggregate measures of employment portfolio risk and return are more appropriate for analysis at the aggregate county level.

The employment portfolio composition of a region impacts aspects of the economy beyond those related directly to migration. The correlation of returns across firms in the portfolio also impacts the overall economic performance of a city or county. Much like the investments in a financial portfolio, diversity of returns across industries can help to lessen the overall variability experienced in a local economy. The second chapter of the dissertation explores the concept of economic resilience and its evolution through the regional economics literature. This concept has emerged as a popular complement to the regional growth path literature in discussions of what makes an area attractive to the people who live there, and to those people who may potentially move there. A thorough understanding of the mechanics of resilience will aid in the recovery from future recessions, by informing policymakers on choices that might dampen the severity of the impact and accelerating the recovery that occurs afterward. Regions that can learn and implement strategies based on these lessons will benefit from living in a stable economic environment, resulting in fewer deviations from their long-term growth path, and achieving closer to optimal levels of economic growth. The empirical strategy follows in the vein of Deller, Conroy, & Watson (2017) and Kitsos & Bishop (2018). We employ two measures of resilience across a set of

[1] We thank the Colorado Department of Labor and Employment and Ryan Gedney for the use of this non-public data.

seven different dependent variables to test for a cohesive set of determinants for resilience. After testing each measure separately and then testing the variable set across models, we find a number of independent variables that appear important to include when testing for resilience. We then go on to discuss what makes an impactful measure of resilience, the strengths and weaknesses of existing measures, and provide a note of caution for employing a unidimensional measure of resilience in any empirical framework. These results generate insights on how policy makers can balance the tradeoffs between a short-term focus on growth and a long-term focus on resilience.

The final chapter examines the dual declines in U.S. migration rates and economic dynamism that have occurred over the past several decades. It explores the role of information on dynamism as an element of the psychic costs that results in a sub-optimal amount of migration. Drawing on the theoretical work of Damm (2009), we develop a model of migration that highlights the role of psychic costs in the migration decision. This hypothesis is tested using a panel dataset covering the years 1998-2014. The empirical section utilizes a first difference estimator to test the significance of information variables representing dynamism at the metropolitan statistical area (MSA) and metro county levels of analysis. The results show that the information variables from dynamism are statistically significant in the regression of inward migration rates. We then employ a three-fold Blinder-Oaxaca decomposition to test for the significance and drivers of the persistent divergence in single-county and multi-county MSA average in-migration rates. We conclude that the main contributors to the difference in rates can be attributed to the effects of differing population sizes and the first lag of the contraction rate. The empirical results also show a unique role for information generated from dynamism in the migration discussion for smaller, single-county Metropolitan Statistical Areas (MSA). We theorize that the stronger role for local information at this level is related to the more homogenous identity present in smaller MSAs. This stands in direct contrast to the larger, multi-county MSAs where an individual may more easily locate in an area that aligns strongly with their tastes and preferences, due to the greater diversity of neighborhoods available in larger MSAs. This chapter takes a macro level approach to complement existing works that focus on micro level data related to individual job turn-over rates.

The three chapters of this dissertation work together to expand on the idea of why individuals decide to live where they do, what influences their decision to move, and why some locations seem to experience both a more vibrant economy and a stronger defense against recessions. This resilience, in turn, helps make an area more attractive to potential migrants by avoiding long, persistent economic downturns. Chapter 1 highlights selecting the appropriate level of aggregation for measures of employment portfolio returns at the county level. Chapter 2 provides insight on selecting a meaningful measure of economic resilience and identifying key determinants of resilience. The last main theme comes in Chapter 3, where the role of uncertainty reducing information generated by economic dynamism helps tie together the concepts of migration and dynamism. Following Chapter 3, a short section highlighting findings, policy implications, and future research completes the dissertation including a brief summary that revisits the main themes.

Chapter 1 - Examining Colorado county level in-migration rates: an employment portfolio approach

1.1 Introduction

When discussing migration, it is essential to consider the role of local labor markets in determining attractive potential destinations. Much like money flows towards investments, migration may be more likely to flow towards areas that have higher returns and lower risks. The regional employment portfolio literature has largely concerned itself with finding the right balance between risk and reward to provide a stable growth environment. Low & Weiler (2012) expand the use of employment portfolio measures as explanatory variables in a model of regional entrepreneurship. This chapter seeks to follow in that vein by extending the use of portfolio measures as explanatory variables to a model of migration. We propose that counties containing employment portfolios providing higher risk-adjusted returns will experience larger migration inflows in subsequent periods than counties with lower return or higher risk portfolios. By decomposing the overall employment portfolio return and risk measures we can better isolate the most appropriate level of employment data aggregation to use and we can examine the impact of employment growth on migration flows. This research provides contributions to both the employment portfolio and regional migration literature. This chapter provides an in-depth case study on migration into Colorado using detailed industry level employment data.

1.2 Migration and Regional Portfolio Theory Literature

Migration Literature

Within the regional economics literature there is a long litany of works which aim to explain the reasons for migration. Labor market dynamics are one reason that repeats throughout this research. The literature often discusses the role of migration as an adjustment mechanism for reducing regional economic shocks (Blanchard & Katz, 1992). The relatively high internal mobility in the U.S. has also often been posited as an explanation for why the U.S. has lower average unemployment compared to European countries (Bentivogli & Pagano, 1999; Magrini, 2004). This drives a labor perspective to migration that results from labor responding to disequilibrium in the market, in the form of higher real wages. The U.S. Census Bureau relays survey findings that show 34.8% of intercounty moves were for job related reasons, the most frequently selected reason for that type of move (Ihrke, 2014). Todaro (1969) shows that economic theory identifies wage differences as the main reason behind the decision to migrate. Similarly, Molloy, Smith, & Wozniak (2017) discuss how many benefits and costs are related to employment opportunities or are otherwise financial in nature. Benefits include a higher wage or better job match, improved job search prospects, and less-expensive housing. Costs include the literal cost of moving one's household, time costs associated with finding new housing and employment, and the loss of local networks or location-specific human capital.

However, there is mixed empirical evidence on whether higher wages are the primary reason for migration within the U.S. (Bentivogli & Pagano, 1999). Utility theory shows that individuals aim to maximize utility net of costs across locations. Saks & Wozniak (2011) show that there are cyclical fluctuations in the net cost of migration. This manifests in the costs of buying and selling a home, fluctuations in wages, and changes in the employment outlook. Bauernschuster et al. (2014) discuss that the negative impact of distance on migration can reflect not only the physical costs of moving, but the psychic costs that occur when leaving familiar surroundings.

After discussing the full extent of benefits and costs to migration, the literature moves into an investigation of cross-national differences and what type of persons are engaging in migration. Bentivogli & Pagano (1999) compare states in the U.S. to 44 regions from the Euro-11 group of countries and find that the response of migration to wage and unemployment differentials is much greater in the United States than in the Euro-11 regions. Unemployment differentials spur population flows in the U.S. that are 10 times as large as those in the Euro-11, while wage differentials give rise to flows that are double the size. The development literature has a depth of publications relating to the allocation of family members across locations and employment sectors. This portfolio model of labor allocation is evident in works like Anam & Chiang (2007) which discusses the allocation choices families make to maximize their aggregate labor portfolio return. Additionally, Anam, Chiang, & Hua (2008) present a two-period model where the portfolio motive interacts with the potential benefits of delaying migration in determining international migration of the extended family unit. Papers at both the international and regional levels display evidence that education and risk appetite play roles in the propensity to engage in long distance migration. Grogger & Hanson (2011) find that a striking feature of international labor flows is that more educated individuals are the most likely to move abroad. They develop and estimate a simple model of migration based on the Roy (1951) income maximization framework and show that migration is increasing in the level of earnings difference between the destination and the source countries, with countries that provide high rewards to skill attracting a disproportionate share of more-educated emigrants. Likewise, Bauernschuster et al. (2014) find that highly educated individuals are more mobile in general, and less sensitive to distance when they migrate. With a positive correlation between education and willingness to migrate, the next vein of the literature examines the destinations of migrants, the response, and incentives local governments may provide to try and capture these highly educated workers.

The impacts of migration are often felt disproportionately across metro, micro, and rural counties. These population dynamics can heavily impact rural counties, who may run into issues maintaining enough of a tax base to support local services such as schools, emergency responders, and infrastructure (Kilkenny, 2010; Chen et al., 2013). It is often also the case that rural economies may be more specialized than metro counties, leaving them susceptible to greater economic volatility over the course of a business cycle (Kort, 1981). Several papers discuss the impact on regional growth resulting from attracting migrants with high levels of human capital and if local governments should actively engage in this practice. McGranahan & Wojan (2007) re-examines the lessons of Florida (2002) to test their applicability to rural development, an extension from the metro-based premise of the original article. They also improve upon the data provision of the "creative class" and overcome one of the major criticisms of Florida's work by testing the impact of growth in a multivariate regression, including economic and creative class growth. They find that not all rural areas are likely to benefit from a strategy of attracting creative workers. Rural areas most attractive to creative workers tend to have a sufficient density to provide a reasonable level of services, appealing landscapes and other natural amenities, and growth in surrounding areas. In a 2005 work, McGranahan also shows that natural amenities are associated with employment growth indirectly, through their effects on net migration. Falck, Fritsch, & Heblich (2011) stress that it only makes sense for governments to subsidize cultural amenities to attract high human capital individuals if the presence of those individuals results in local knowledge spillovers. Otherwise, there is no solid justification for spending tax dollars in this manner. Their empirical findings show that increasing the local share of employees with a tertiary degree by one standard deviation increases average annual growth of regional Gross Domestic Product (GDP) per capita by about 1.0 to 2.1 percentage points. The positive effect of the level of human capital on a region's growth path is a clear indication of local knowledge spillovers induced by the presence of high-human-capital employees.

The presence of knowledge spillovers in a local area is one of the three types of agglomeration effects described by Marshall (1920), the other two benefits being reduced shipping costs from locating close to a large pool of consumers and labor market pooling – the idea that in thick markets workers can change jobs without changing residences and can bounce back more quickly following shocks by moving to responsive employers (Glaeser, 2010; Krugman, 1992). Agglomeration economies and migration to metro areas compose the final branch of migration literature covered in this paper. Andersson, Burgess, & Lane (2007) state that cities are home to 75% of Americans yet occupy less than 2% of the land area of the lower 48 states. Wheaton & Lewis (2002) find that equivalent workers in manufacturing earn higher wages when they are in urban markets that have a larger employment share in their industry. Their work complements the third benefit of agglomeration economies, suggesting that assortative matching in thick urban labor markets plus complementarities in production play an important role in generating high productivity in cities.

One of the many explanations for agglomeration economies is the New Economic Geography (NEG) introduced by Krugman, which lays out a model that endogenizes backward and forward linkages of agglomeration economies while allowing for interregional migration. However, several decades after his initial work on the subject Krugman acknowledges that the NEG more aptly describes emerging economies and that moving forward, U.S. cities will be less defined by a single industry, compared to the past. While specialized suppliers and thick labor markets for special skills are still factors, we will see information spillovers and entrepreneurial chains of influence take on a larger role (Krugman, 2010).

Indeed, recent studies show that a tightrope act is needed for cities wishing to find the best way forward. Delgado, Porter, & Stern (2010) show that agglomeration forces occur due to the presence of clusters of related and supporting businesses as opposed to diversity of economic activity. While a level of specialization appears to be needed for Marshall, Arrow, Romer (MAR) externalities, cities must take care to balance industries to reduce their overall volatility. Glaeser (2010) speaks to this balance by finding that agglomeration benefits are largest when the sectors have shocks that are heterogeneous so that their shocks are partially uncorrelated. This result, of course, requires that the sectors are still similar enough so that workers can move across them. These issues are further investigated in the branch of regional literature discussing portfolio theory, which is explored below.

Regional Portfolio Theory Literature

The application of financial portfolio theory to regional employment structures focuses on finding the optimal mix of industries for a region or county to maximize returns while minimizing risk. Financial portfolio theory focuses on the benefits of increased returns and reduced risk that diversification brings, while regional portfolio theory discusses the use of economic development strategies to increase the appeal of a region to investors and potential citizens, while balancing growth with volatility. Translating financial measures to job markets lets us define risk as the variance of employment, whereas returns are the growth in employment (Lande, 1994). This vein of literature began in the mid-1970s with Conroy (1974) which first laid out the idea of a community industrial portfolio as the collection of industries that use a community's resources to produce a stochastic stream of returns to labor and capital. He notes that the choice of diversifying industries in a portfolio is impacted by both the existing structure of the portfolio and by the effect of interindustry interdependence, measured by the covariance with other industries. Barth, Kraft, & Wiest (1975) aims to lay out the conditions by which a new or expanding industry will reduce the fluctuations in regional employment. Barth et al. (1975) recognizes that the best additional industry to produce the greatest stability in total employment may not be the least volatile industry, but rather one that is uncorrelated with the existing portfolio components.

Following these initial explorations, Kort (1981) fleshes out the theoretical model to account for the fact that there is a demonstrative positive relationship between industrial diversification and city size. Therefore, larger cities are relatively more stable compared to smaller cities. This was the first paper to test and model the relationship between the three variables: size, diversification, and stability. Kort (1981) finds that large cities tend to emulate the national economy, while smaller cities display a much wider range of cyclical instability. This means heteroskedasticity is present and that Ordinary Least Squares (OLS) is not the appropriate empirical framework.

With the number of early studies applying financial portfolio theory to regional employment questions, several risk and reward measures were created. Jackson (1984) presents an empirical analysis of the relationship between employment stability and four different measures of industrial diversity in multicounty regions of Illinois. Here the level of analysis is moved to the multicounty regional grouping instead of the MSA level. Jackson does not find support for a linear relationship between diversification and employment stability. Further, he questions if employment growth and stability are even the correct measures of return and risk and states that industry composition is too simplistic a metric to drive policy prescriptions. Jackson's conclusion is directly challenged by Wundt (1992) who responds that "while empirical results have not been without qualification, there is enough evidence to suggest that industrial composition is one important element influencing economic fluctuations in regions." Wundt (1992) produces an empirical analysis using the manufacturing employment share between 1964 and 1983 as a measure of cyclical sensitivity.

As the track record of empirical support for the application of portfolio theory to regional issues grew, several papers began to explore the appropriate level and limits of this approach. Lande (1994) uses a state level approach instead of MSAs to avoid masked data at the 2-digit Standard Industrial Classification (SIC) level. This well-known tension between unit size and data availability continued to be expressed in the next several decades of papers in this branch of the literature. Lande (1994) examines data for Tennessee at the two, three, and four-digit industry code levels and finds evidence of a U-shaped curve between employment growth and employment instability. Additionally, the evidence shows that "the actual allocation of employment is associated with much greater volatility than is optimal." The author goes on to note that employment volatility could be mitigated without harm to employment growth by implementing alternative industrial structures. More recently Deller & Watson (2016a) use the Great Recession to confirm the relationship between a county's economic diversity and employment stability. They find that higher levels of economic diversity are associated with higher levels of stability in the population-employment ratio, the unemployment rate, and the concentration of establishments. At the same time, Hong & Xiao (2016) show that regions with diversified specializations were more likely to experience both higher employment growth rates and less turbulence in employment growth compared to those with fewer specializations. Lastly, Kluge (2018) applies portfolio theory to German districts to show empirically that diversification allows regions to achieve more efficient locations on the growth-instability frontier, meaning they can achieve greater stability at given levels of economic growth.

Chandra (2002, 2003) conducts in-depth investigations on the unit size limitations for valid policy prescriptions. These papers work to bring regional portfolio theory into line with financial portfolio theory, especially the Markowitz framework, by testing for an empirically validated direct link between growth and instability resulting from the sectoral structure of the economy, particularly a non-linear frontier growth-instability relationship. Chandra (2002) shows that a convex frontier between growth and instability exists for economic growth at the state level in the U.S. The idea of a convex frontier is further examined in Chandra (2003) which extends the 2002 work to the European markets to see if a convex frontier also exists in these economies. The European system of classifying regions for analytical purposes is codified in the Classification of Territorial Units for Statistics[2]. Using the NUTS classifications, Chandra tests different levels of regional aggregation, using both stochastic frontier estimation and data envelopment analysis. The findings show that the growth-instability frontier exists for the more aggregated levels of the European economy. At the most disaggregated level of regional data (NUTS3) there is little to no evidence of the convex frontier. It also shows that the more aggregated regions are more economically diverse compared to the less aggregated regions. These findings suggest that there is a level of aggregation beyond which the applicability of regional portfolio theory ceases to be effective or meaningful due to structural limitations. Spelman (2006) corroborates the existence of a convex frontier between growth and instability using three 10-year periods spanning 1970 – 2002 for MSAs in the U.S., concluding that the metro level of analysis is appropriate and that policy implications exist. Of note is that most metro area economies fall well inside the efficient frontier, meaning they are taking on unnecessary levels of risk relative to their level of return.

Low & Weiler (2012) take the employment portfolio investigation to yet a different unit of analysis, examining the incentives for self-employment and entrepreneurship that are generated by varying levels of risk and return within commuting zones (CZ) and counties. This is the first successful extension of portfolio theory to non-metro areas using commuting zones. The empirical application embeds the risk and reward variables as explanatory measures of the rate of self-employment at the county and CZ levels. In this paper we aim to extend the approach of Low & Weiler (2012) by examining the pull factor of employment portfolio measures on inward migration at the county level. We also add to the discussion by examining different levels of employment classification for the construction of our risk and reward measures. Using the county level of observation, but varying the level of employment detail,

[2] This is abbreviated as NUTS, for the French *nomenclature d'unités territoriales statistiques*.

provides a new perspective on the applicability of portfolio risk and reward measures as explanatory variables.

While many contributions to this literature focus on the potential for prescriptive analyses, showing how regions can shape their portfolio to “increase returns”, the investigation this work will undertake accepts the portfolio composition as given and looks at how the divergence in portfolio composition across counties impacts migration flows into Colorado. A similar approach of utilizing employment portfolio measures as independent variables is taken in Gulati & Weiler (2021), where they examine the impact of employment portfolio risk and return on the survival of new businesses during various time frames. The portfolio variables are used in an explanatory fashion, rather than being the focus of policy prescriptions. This chapter adds to the literature of both regional employment portfolio theory and the “pull” determinants of regional in-migration.

1.3 Methodology

For the empirical investigation, this chapter uses a Poisson fixed effects model. This section will detail why this model was selected over other options. We show that the fixed effect version of the Poisson model has significant strengths when compared to a standard Poisson (Hausman, Hall, & Griliches, 1984; Greene, 1994; Wooldridge, 1999), that the existing literature has a history of using a Poisson fixed effects model in location studies (Papke, 1991; List, 2001), and that the model provides a link between theory and empirics (Guimarães et al., 2004).

When modeling count data, the Poisson regression is the most widely used estimator (Cameron & Trivedi, 1998). There have been numerous extensions to adapt the standard Poisson model to accommodate a variety of situations (Hausman, Hall, & Griliches, 1984; Greene, 1994). Two of the most notable works on these extensions are the model by Hausman, Hall, & Griliches (1984) to include fixed effects, and Greene (1994) which provides guidance on model selection in the face of excess zeros. However, it is the work of Wooldridge (1999) that shows that the Poisson fixed effects estimator is very robust and therefore, tends to be more reliable, in comparison to alternatives like the negative binomial estimator. For the Poisson fixed effects model to be consistent, the only requirement is that the conditional mean assumption holds. This provides flexibility to implement the Poisson fixed effects model even in the face of overdispersion. Consistent estimators will still be provided, and the larger standard errors will provide a conservative approach to determining significance.

The Poisson fixed effects model is particularly simple and is one of the few known models in which the incidental parameters problem (IPP) is, in fact, not a problem. The same is not true of the negative binomial (NB) model (Greene, 2007). The IPP was first articulated by Neyman & Scott (1948) and discusses the difficulty in utilizing time-invariant fixed effects in non-linear regression models dealing with panel data. In many instances of non-linear models with fixed effects, since the number of fixed effects increases in line with the number of observations, bias is introduced into the equation and the estimator fails to converge. There are several ways to avoid this problem. One way is to use the conditional logit specification from Chamberlain (1980) where the issue of panel fixed effects is mitigated via the transformation to the likelihood function. The method employed here is to use the Poisson fixed effects method which does not suffer from this bias.

List (1991) and Greene (2007) discuss the use of a zero-inflated Poisson model when excess zeros arise in the data, but stress that the model is appropriate when there is a belief that there are two distinct processes at play: one where the outcome will always be zero, and one where the zeros occur through the standard Poisson process. In our current study, we are discussing migration from one county to another. We find no a priori theory that would contain an "always zero" process for county-to-county migration, so a zero inflated model does not appear to be appropriate, given that the standard Poisson fixed effects model can accommodate the overdispersion of zeros while still providing consistent estimators.

The Poisson fixed effects model has been used throughout the economics literature, particularly in location studies. It has been used by Papke (1991) to model the number of firm births while explicitly accounting for unobserved location or state heterogeneity. Additionally, it was used by Henderson (1996) in modeling the impact of air quality regulations on location choices. Finally, List (1999) enumerates a handful of other studies to use the Poisson fixed effects model including: Kogut & Chang (1991), Blonigen & Feenstra (1997), and Blonigen (1996).

A final strength of using the Poisson model is that it more closely links the theoretical and empirical approaches used in this investigation. Our model follows the approach of Marré & Rupasingha (2020), Rupasingha et al. (2015), and Goetz (1999). Migration decisions are based on a utility maximization approach where potential migrants look to maximize utility across several locations, considering the cost of moving and potential earnings. Migrants maximize utility across a set of alternative locations, $j = 1, \dots, J$ compared with the utility derived at their current location, i. Utility is defined as a function of potential earnings in each location, w_i or w_j, the costs associated with moving from i to j, c_{ij}, and amenities in each location, a_i or a_j.

Potential migrants evaluate d_{ij}, the difference in utility between moving to location j and staying in the current location i, as shown in equation (1.1), and move if $d_{ij} > 0$ for some j:

$$d_{ij} = U\left(w_j - c_{ij}, a_j\right) - U(w_i, a_i) \ . \tag{1.1}$$

The random utility model presented in McFadden (1974) illustrates that the choice between two locations can be presented as a vector containing the characteristics in both places (X_{ij}) as shown in equation (1.2).

$$U_{ij} = \ \beta' X_{ij} \ + \ \varepsilon_{ij} \ . \tag{1.2}$$

In maximizing utility, the prospective migrant aims to find the location that maximizes the difference between the new location's utility and the utility of their current residence. We can express the probability that an individual moves from i to j as:

$$P_{ij} = P\left(U_{ij} > U_{ik}\right) \text{ for all } k \neq j. \tag{1.3}$$

McFadden (1974) goes on to state that if the ε_{ij} are independently and identically distributed (i.i.d.) following an extreme value Type-I distribution, equation (1.3) can be re-written as the following:

$$P_{ij} = \frac{\exp(\beta' X_{ij})}{\sum_j \exp(\beta' X_{ij})}\,. \tag{1.4}$$

Equation (1.4) provides the structure typically used to estimate a conditional logit model that has been seen often in firm location and migration studies.

By assuming independent observations, we can write the log likelihood function ($\log L$) for all migrants from any county i to a specific Colorado county j. Here $d_{ij} = 1$ if migration occurs for the individual and $d_{ij} = 0$ otherwise. Analogously, $n_{ij}$is equal to the total number of migrants moving from county i to Colorado county j.

$$\log L = \sum_i d_{ij} \log P_{ij} = n_{ij} \log P_{ij} \tag{1.5}$$

Since we are looking at migration location decisions from all counties to all Colorado counties, we can express equation (1.5) as:

$$\log L_{cl} = \sum_i \sum_j n_{ij} \log P_{ij} \tag{1.6}$$

The empirical approach considered in the migration literature has evolved over time, as access to more granular data has become available. The earlier studies such as Greenwood (1975) use aggregate data to discuss migration flows. In fact, a separate chapter in this dissertation uses aggregate county level in-migration rates to explore the relationship between migration and economic dynamism. As data improved, economists were able to analyze pairwise data, such as state-to-state or county-to-county level migration flows. This required models that were able to account for the shift from one possible destination to a multitude of possible migration destinations, including remaining in the current location. The move to conditional logit models alleviated many of these issues. This model draws strength from its direct ties to the theoretical construct of an individual utility maximization framework that flows from a random utility model (Davies, Greenwood, & Li, 2001). The move to a conditional logit model also allows for the inclusion of distance as an independent variable, a shortcoming of the earlier logit models where the decision is simply whether to migrate, as opposed to considering a specific destination.

The main difficulty of the conditional logit empirical approach, as discussed in detail by Guimarães et al. (2003, 2004), is that the conditional logit model requires assuming the independence of irrelevant alternatives (IIA). This means that locations are equivalent to the location decision-maker after the explanatory variables in the model are controlled for. If there are unobserved location-specific characteristics influencing each firm's location decision, then estimated coefficients will be biased. The nature of migration questions opens the possibility for this type of bias. The benefits of using a standard Poisson regression model in place of a conditional logit model is that it allows for the modeling of many alternative locations and avoids the potential pitfall of a violation of the Irrelevant Alternatives Assumption. Fortunately, Guimarães et al. (2003, 2004) provide an alternative estimation strategy that avoids the potential pitfalls of a conditional logit model for firm location or relocation studies with large numbers of alternative locations. They showed that given certain conditions, a Poisson regression model may be used as an alternative to a conditional logit model. These conditions result in the log-likelihood functions of both models being equivalent. Namely, they show that the model in equation (1.6) can be estimated using n_{ij} as the dependent variable.

Following Guimarães et al. (2004), fixed effects are used to control for characteristics of the origin counties and explanatory variables are included for the destination counties. Such an approach has the benefit of controlling for the push and pull factors determining the migration decisions of people. Push factors are controlled for by county-level fixed effects for origin counties and pull factors are controlled for by measured county-level characteristics.

In the following empirical model, we follow a utility maximization approach to migration decisions. The dependent variable in our model is the county-to-county total migration count occurring between 2001 to 2016. This timeframe was selected partially due to data availability, but also by considering that it covers a complete business cycle. Our empirical model is constructed as a Pairwise

Poisson regression with fixed effects for the originating counties and control variables for the Colorado destination counties in the spirit of Marré & Rupasingha (2020). This creates an $n \times m$ matrix of potential migration flows where n equals the 63 counties in Colorado and m equals the 3,142 counties and county-equivalents in the 50 states and District of Columbia. Colorado currently has 64 counties. However, the newest county, Broomfield, did not come into being until November 15, 2001. We have omitted Broomfield to avoid any data issues. To control for differences in the size of the destination counties we designate population in the destination county as the offset or exposure variable. When the exposure variable is placed into the regression, it is done so as the natural log of the variable and its coefficient is constrained to be 1. The exposure variable allows us to account for the area of space where the count of migrants took place. The exposure variable essentially modifies each observation so that the count outcome is weighted based on a measure of area. It would be similar to designating the number of days in a month as the exposure variable when explaining the number of births in a month, to be able to properly compare February to August.

1.4 Data Description

Migration Flows: Internal Revenue Service Migration Data

There are several data sources providing migration data within the United States. For this investigation, we are utilizing the Internal Revenue Service (IRS) migration data set. This data is provided on a yearly basis through a collaboration between the IRS Statistics of Income Division (SOI) and the U.S. Census Bureau. Data is currently available as pairwise county level data flows for the period 2001-2002 to 2015-2016. Migration is identified through a change in address on filed returns across consecutive years. A richness in the data is that the IRS provides separate datasets for inflows and outflows. Three measures are provided for each county pair: total number of tax returns, total number of exemptions, and aggregate adjusted gross income. Any form of data that relies on voluntary response will never be 100 percent accurate, and that is true in this case as well. Additionally, not everyone is required to file taxes. The main exemptions to the tax filing requirement are low-income earners (Internal Revenue

Service, 2022). This would result in a data set that underrepresents these lower income earners. We do not implement any adjustment for this potential bias but did want to highlight it in the data discussion.

Employment Portfolios: Colorado Level

The data used to construct the measures of employment portfolio risk and return come from the Quarterly Census of Employment and Wages (QCEW) data set. This data set provides employment counts and wage data at the firm level on a quarterly basis. Firms are classified using the North American Industry Classification System (NAICS) and are available from 2001 – 2016. Through an agreement with Colorado Department of Labor and Employment we are permitted to access the firm and unit level data for the state of Colorado, providing a richer data set to mine for insights, all while maintaining establishment anonymity. Employment portfolios are constructed using employment count data at the two-digit NAICS code level. The empirical models will include industry level measures of risk and return which are measured by the mean annual employment growth rate and the standard deviation of annual employment growth rates. These measures are constructed at the county level and cover the period from 2001-2002 to 2015-2016. The measure of portfolio return is constructed for each county as the mean of the annual employment growth rates from 2002-2016. This approach is consistent when considering the overall employment growth rate for a county, as well as later in the investigation when we examine more granular industry level employment growth rates. Similarly, the measure of portfolio risk for a county is measured as the standard deviation of the annual employment growth rates for the timeframe of 2002-2016. This same approach is carried down to the industry level to construct industry specific measures of employment risk as the standard deviation of the annual employment growth rates for that industry over the specified timeframe.

Additional Independent Variables

To account for differences between counties, we include several additional independent variables. One of the most important variables when discussing migration is the distance between county pairs. Migration tends to correlate negatively with distance (Schwartz, 1973). In this work we utilize the distances provided by the National Bureau of Economic Research. They measure distance in miles between county internal points (roughly the geographical center of a county) using the 2010 U.S. Census county list. County distances are great-circle distances calculated using the Haversine formula (County Distance Database). To account for the possibility that the relationship between distance and migration is not purely linear we include both the linear distance as well as distance squared. Additional controls include the U.S. Department of Agriculture (USDA) natural amenity scale, a housing price index, median household income, and variables to designate the degree of urbanization or rurality. The descriptive statistics for the Colorado counties comprising the data set are found below in Table 1-1.

Table 1-1: Colorado county level descriptive statistics

Variable	Mean	Std.Dev.	Min	Max
Overall Mean Employment Growth Rate	.681	1.127	-2.113	5.486
Overall Std. Deviation of Employment Growth Rate	2.976	1.694	1.165	9.983
Mean Employment Growth Rate Construction	2.465	8.896	-4.401	50.102
Mean Employment Growth Rate Education and Health	1.381	2.037	-4.745	7.352
Mean Employment Growth Rate Financial Activities	.348	2.845	-7.725	12.108
Mean Employment Growth Rate Government	.883	1.084	-1.55	4.033
Mean Employment Growth Rate Information	.41	6.55	-8.835	35.284
Mean Employment Growth Rate Leisure and Hospitality	3.1	16.284	-4.008	129.493
Mean Employment Growth Rate Manufacturing	3.318	11.362	-15.798	48.915
Mean Employment Growth Rate Nat Res and Mining	14.733	75.159	-4.635	596.29
Mean Employment Growth Rate Other Services	3.663	8.252	-3.844	42.265
Mean Employment Growth Rate Prof and Business Services	4.444	7.068	-8.561	38.202
Mean Employment Growth Rate Trade, Transportation, and Utilities	.265	1.419	-2.86	5.545
Std. Deviation of Employment Growth Rate Construction	24.946	33.416	5.724	227.917
Std. Deviation of Employment Growth Rate Education and Health	5.075	2.97	1.676	15.389
Std. Deviation of Employment Growth Rate Financial Activities	11.646	10.254	2.644	52.75
Std. Deviation of Employment Growth Rate Government	4.187	2.799	1.356	20.194
Std. Deviation of Employment Growth Rate Information	16.705	17.282	3.839	127.002
Std. Deviation of Employment Growth Rate Leisure and Hospitality	18.206	65.978	2.305	530.045
Std. Deviation of Employment Growth Rate Manufacturing	27.929	38.303	4.036	206.383
Std. Deviation of Employment Growth Rate Nat Res and Mining	60.449	272.516	5.113	2177.562
Std. Deviation of Employment Growth Rate Other Services	17.332	19.969	2.728	105.244

Std. Deviation of Employment Growth Rate Prof and Business Services	18.86	20.627	3.164	127.239
Std. Deviation of Employment Growth Rate Trade, Transportation, and Utilities	6.353	4.151	1.588	21.129
Median Household Income ($1,000s)	40.304	12.932	20.403	93.316
RUCC 2013	5.873	2.888	1	9
USDA Amenity Scale	4.026	2.317	-.71	8.52
Population Density	111.472	475.337	.731	3673.046
Population (1,000s)	70.249	142.198	.581	561.976

(n = 63)

Addressing endogeneity concerns:

Several steps are taken to minimize the effect of endogeneity on the results. First, with the exception of the portfolio measure variables, all our independent variables come from the start of our investigation window, the year 2001. In addition, county fixed effects for the origin locations are used to reduce omitted variable bias. Perhaps the most important defense against endogeneity, as mentioned in Marré & Rupasingha (2020), is taken from the estimation approach we employ. Rather than the standard cross-sectional approach used in most location and relocation studies, which inevitably raises the prospect of variables being jointly determined, our approach is to model origin-destination county pairs. Each individual county pair of in-migration usually makes a small contribution to the total inflow of migrants to a particular county, leading to a small effect on the destination county's overall economy, minimizing endogeneity concerns.

1.5 Results

Aggregate Measures of Risk and Reward

We first examine the impact of the portfolio measures on migration when the portfolio risk and return measures are calculated at the aggregate level for each county. These values come from the Bureau of Economic Analysis (BEA) employment numbers. For all the following regressions the dependent variable is the sum of claimed exemptions from the IRS migration data from 2001-2016 and the county-to-county level. This will proxy for the number of individuals moving between any particular county pair over this time period.

Table 1-2: Aggregate portfolio model results

County to County In-Migration Count	Coef.	St.Err.	t-value	p-value	95% Conf	Interval	Sig
Overall Mean Employment Growth Rate	0.492	0.091	5.42	0.000	0.314	0.670	***
Overall Std. Deviation of Employment Growth Rate	-1.003	0.066	-15.12	0.000	-1.133	-0.873	***
Distance between Counties (100s mi.)	-1.965	0.240	-8.19	0.000	-2.435	-1.494	***
Distance between Counties (100s mi.) Squared	0.001	0.000	7.34	0.000	0.001	0.001	***
Median Household Income ($1,000s)	0.064	0.006	10.45	0.000	0.052	0.076	***
USDA Natural Amenity Scale	-0.184	0.036	-5.08	0.000	-0.255	-0.113	***
Population Density (Persons/square mile)	0.001	0.000	16.90	0.000	0.001	0.001	***
Mean dependent var	88.562		SD dependent var		2,052.423		
Number of obs	40,387		Chi-square		1,885.427		
Prob > chi2	0.000		Akaike crit. (AIC)		6,009,669.082		

**** p<0.01, ** p<0.05, * p<0.1*

From these results we see that all the included variables show up as significant at the 1% level. Furthermore, except for the natural amenity scale, all variables have intuitive coefficient signs. Notably, distance between counties negatively impacts migration. The portfolio measures display the expected results, there is a positive coefficient on the return variable and a negative coefficient on the risk measure. This coincides with the intuitive understanding of being attracted to returns and offput by risk. One puzzle is the negative coefficient on the USDA natural amenity rank variable. One possible explanation could be that the results are driven by migration to metro counties, which are typically lower in natural amenities than rural counties. The bias toward metro/urban migration can be seen by visualizing the data. The following heat map (Figure 1-1) shows migration to Colorado counties. The counties receiving the largest number of migrants are colored dark blue. These counties are clustered in the Colorado Front Range area, which include the metro areas of Denver, Fort Collins, and Boulder.

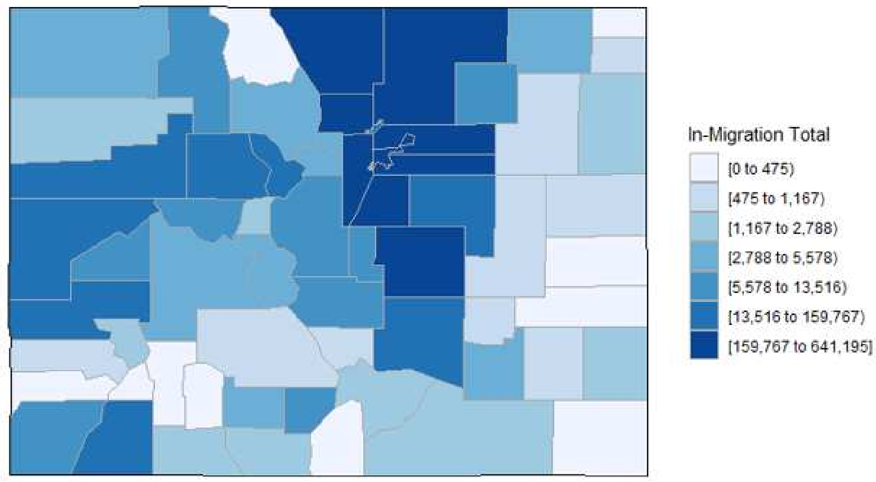

Figure 1-1: Heatmap of Colorado county total in-migration counts (2001-2016)

To try and account for metro preference, the model was re-run, including the 2003 version of the Rural Urban Continuum Codes (RUCC) to assign county types. Counties are assigned a value from 1 to 9 based on their population size and their geographic adjacency to a metro area. The 2003 vintage of the RUCC values cover most of our study period, as codes are updated every 10 years. The 2003 and 2013 criteria are the same and are presented in Table 1-3 below.

Table 1-3: 2003 and 2013 rural-urban continuum code (RUCC) criteria

Code	Description
Metro Counties:	
1	Counties in metro areas of 1 million population or more
2	Counties in metro areas of 250,000 to 1 million population
3	Counties in metro areas of fewer than 250,000 population
Nonmetro Counties:	
4	Urban population of 20,000 or more, adjacent to a metro area
5	Urban population of 20,000 or more, not adjacent to a metro area
6	Urban population of 2,500 to 19,999, adjacent to a metro area
7	Urban population of 2,500 to 19,999, not adjacent to a metro area
8	Completely rural or less than 2,500 urban population, adjacent to a metro area
9	Completely rural or less than 2,500 urban population, not adjacent to a metro area

Source: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx

The following table shows that 90% of migration into Colorado counties occurs between metro areas, with the remaining 10% occurring across other county-type combinations.

Table 1-4: Crosstab of migration total by county origin and destination type

		Origination County Type			
		Metro	**Micro**	**Rural**	**Total**
Destination County Type	**Metro**	90.2%	3.0%	1.1%	94.4%
	Micro	2.6%	0.8%	0.4%	3.8%
	Rural	1.0%	0.4%	0.5%	1.9%
	Total	93.8%	4.2%	2.0%	100.0%

This version of the model produced results similar to the first model. However, further investigation revealed there was a tension between the county level employment portfolio variables and the RUCCs, which are assigned based on metro area adjacency. This meant that some of Colorado's smallest counties were being considered metro areas (RUCC of 1 to 3), even though their employment portfolio measures were subject to greater volatility due to having a smaller county level employment base. Because of this inconsistency we moved to a population-based classification. The RUCC model results are available as Table A.1 in Appendix A.

To try and account for rural-urban preference, the regression was rerun using a categorical variable that grouped counties as either small, medium, or large based on their population size. This was done across three different sets of population cuts as a sensitivity test. The three scales are as follows:

Table 1-5: Population scale cuts

County Type	Version 1	Version 2	Version 3
Small	Population < 10,000	Population < 15,000	Population < 20,000
Medium	10,000 to 50,000	15,000 to 75,000	20,000 to 100,000
Large	Population > 50,000	Population > 75,000	Population > 100,000

The results were consistent across the three models. The table below uses the results from the Version 2 population buckets to present a midpoint.

Table 1-6: Aggregate model results - population cut scale version 2

County to County In-Migration Count	Coef.	St.Err.	t-value	p-value	95% Conf	Interval	Sig
Overall Mean Employment Growth Rate	-0.114	0.072	-1.60	0.111	-0.255	0.026	
Overall Std. Deviation of Employment Growth Rate	-0.282	0.031	-9.13	0.000	-0.343	-0.222	***
Distance between Counties (100s mi.)	-1.995	0.171	-11.70	0.000	-2.329	-1.661	***
Distance between Counties (100s mi.) Squared	0.001	0.000	9.97	0.000	0.001	0.001	***
Median Household Income ($1,000s)	0.036	0.009	4.14	0.000	0.019	0.052	***
USDA Natural Amenity Scale	-0.127	0.062	-2.05	0.041	-0.248	-0.005	**
Population Density (Persons/square mile)	0.000	0.000	6.07	0.000	0.000	0.000	***
Population Cuts							
< 15,000	0.000	.	.	.	.	.	
15,000 to 75,000	1.563	0.156	10.00	0.000	1.256	1.869	***
> 75,000	4.355	0.160	27.21	0.000	4.041	4.668	***

Mean dependent var	88.562	SD dependent var	2,052.423
Number of obs	40,387	Chi-square	1,343.135
Prob > chi2	0.000	Akaike crit. (AIC)	3,245,057.573

**** p<0.01, ** p<0.05, * p<0.1*

The largest concern that emerges from this model is that the measure of overall portfolio return no longer shows as significant. The grouping variable, based on population, does show as significant and confirms the urban migration bias shown in the literature. Additionally, the scatterplot in Figure 1-2 below shows that the risk-reward trade off clearly differs across the three population groups. The group of small (often rural) counties shows no relationship between risk and reward, while the more urban counties show a positive association between risk and reward, as portfolio theory would expect. Interestingly, by comparing the blue (largest counties) and green (medium counties) lines in the plot below, it appears that the most-urban counties can achieve higher levels of return at a similar risk profile compared to the medium-sized counties. This result supports Marshall (1920) and Krugman (1992) in the assertion that labor market pooling can decrease employment risk in larger metro areas. This adds to the argument that migrants consider not only the returns of potential locations, but the risk adjusted return, and more often select a larger metro area, plausibly due to its higher risk adjusted return to employment.

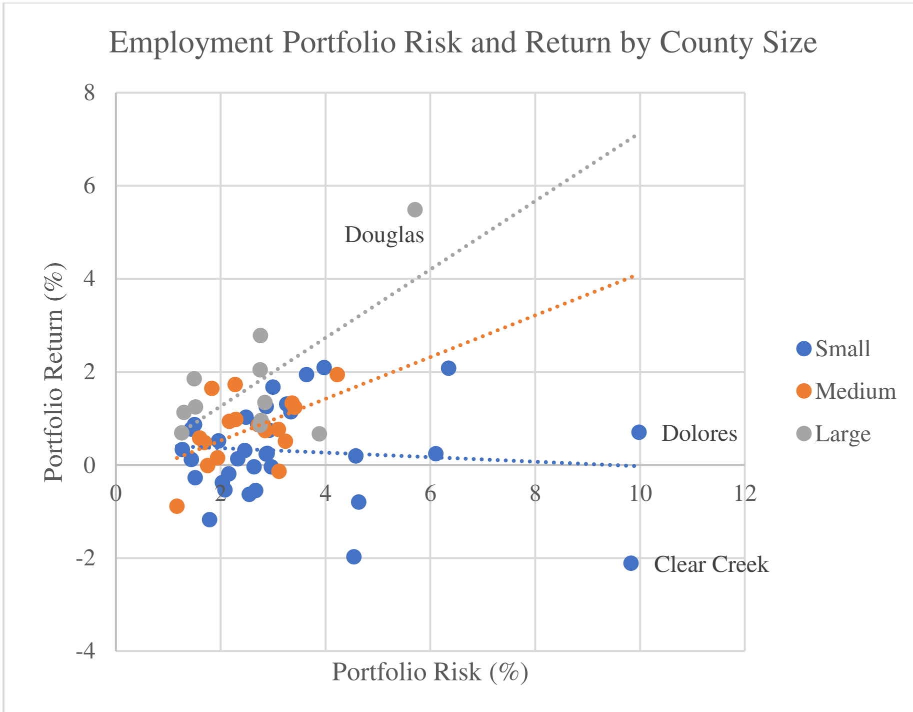

Figure 1-2: Scatterplot of aggregate measures of risk and return, grouped by population cut version 2

It is worth discussing the main outliers in Figure 1-2 to get a sense of what industries may be beneficial or concerning in the portfolio model. Looking at the two main outliers of the small county group identifies Clear Creek County and Delores County as having high degrees of risk paired with low returns. Clear Creek County is best characterized as a mainly residential location with a poorly diversified employment structure highly concentrated in the mining sector. Similarly, Dolores County has a high population of commuters and the in-county employment is based heavily in the agricultural sector. Among the set of high population counties, Douglas County stands out as having the highest portfolio returns. This is likely driven by the high percentage of employment in this county relating to the science, engineering, and technology sectors.

Additionally, the spatial distribution of the aggregate portfolio return and risk measures provide insight into the conditions unique to Colorado, which may not be present if this case study was replicated in other states. Figure 1-3 shows the aggregate employment portfolio returns at the county level, while Figure 1-4 shows the corresponding aggregate employment portfolio risk measurements.

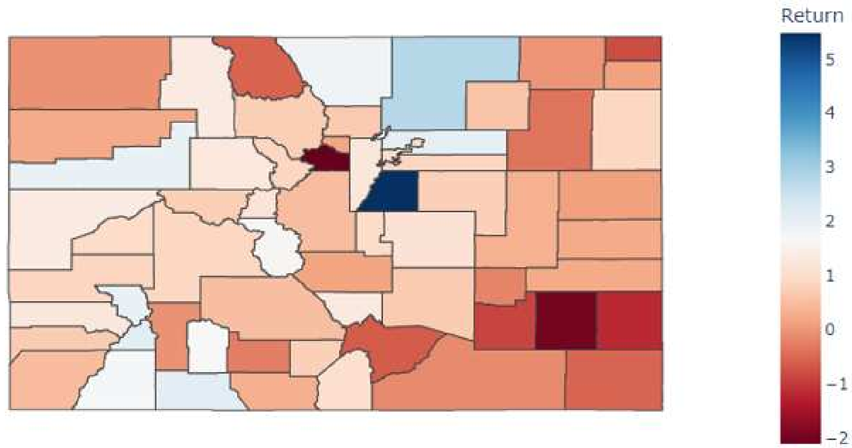

Figure 1-3: County-level employment portfolio returns

Figure 1-3 shows that employment portfolio returns are highly concentrated along the northern Colorado Front Range counties with some additional bright spots coming from Garfield County in the north-west and the counties of Ouray, San Juan, and Archuleta in the south-west corner of the state. The counties of along the Eastern Plains show persistently low returns, driven largely by a combination of sparse population and poorly diversified employment structures.

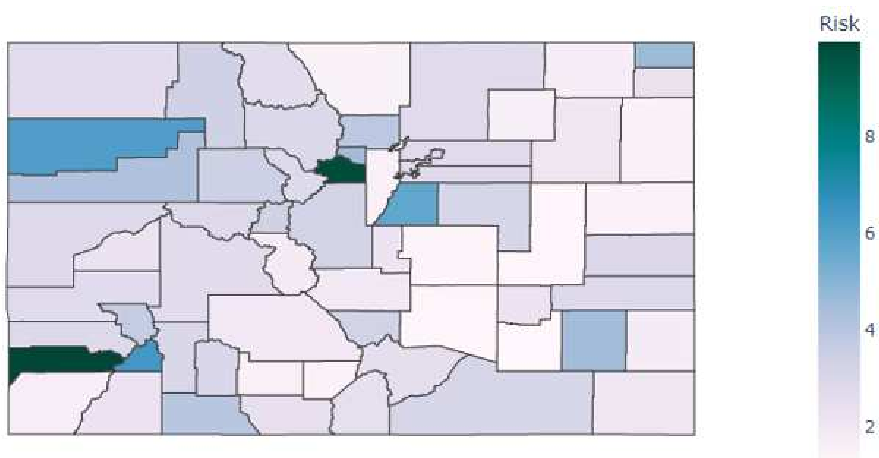

Figure 1-4: County-level employment portfolio risk

Figure 1-4 analogously plots the County-level employment portfolio risk measure for Colorado's counties. As mentioned previously Clear Creek and Dolores Counties experience the largest swings in employment, due mainly to having concentrated employment structures in mining and agriculture, respectively. The industrial structures of these two counties are small, relative to their population, due to the large number of commuters living in these locations. In the next tier of the risk scale, Douglas County pairs its high return with a moderate degree of risk, due to exposure to the technology sector. To the east of Dolores County in the south-west corner of Colorado, San Juan County also experiences a higher level of risk, though this is likely due to the small size of the county. The US Census lists 2021 total employment as 244 in a county with a population of 803. Similar reasons likely explain the variability in employment growth to the north in Rio Blanco County, which had 2022 employment ranging between 2,635 and 2,855.

Due to the nature of the Poisson model, it can be helpful to transform the coefficients into marginal effects to permit an easier interpretation of the associated impact of the independent variables on the dependent variable. For the Poisson model, the marginal effect for any variable k is calculated as (exp (β_k)-1)*100. The resulting values are interpreted as the percentage change in the expected county pair migration count for a unit change in the explanatory variable, all else constant. The calculated marginal effects for the full aggregate employment model are presented below.

Table 1-7: Marginal effects for all counties model

Variable	**Marginal Effect**
Overall Mean Employment Growth Rate	-10.77
Overall Std. Deviation of Employment Growth Rate	-24.57
Distance between Counties (100s mi.)	-86.40
Distance between Counties (100s mi.) Squared	0.10
Median Household Income ($1,000s)	3.67
USDA Natural Amenity Scale	-11.93
Population Density (Persons/square mile)	0.00
Population Cuts	
Small: < 15,000	
Medium: 15,000 to 75,000	377.31
Large: > 75,000	7686.68

The table of marginal effects show a strong preference for both local and metro migration. The marginal effect of distance (-86.40) shows that increasing the distance between county pairs by 100 miles would be associated with an 86.40% decrease in the expected migration count, all else equal. The chart below shows the percentage of moves stratified by the distance between the sending and receiving counties. Most moves (approximately 65%) occur between counties that are less than 100 miles apart. This is consistent across county types.

Table 1-8: Percentage of moves by distance between counties across receiving county size

Distance (mi)	Metro	Micro	Rural	Total
Up to 100	65.69%	61.90%	73.64%	65.70%
101 to 200	3.33%	17.72%	20.10%	4.19%
201 to 500	4.22%	9.76%	4.99%	4.45%
501 to 1000	17.59%	8.87%	1.25%	16.96%
Greater than 1000	9.16%	1.76%	0.02%	8.71%

Changing the size of the receiving county has the largest impact on the expected migration count. Moving from a small county (population < 15,000) to a large county (population > 75,000) results in an associated increase in expected migration of 7,686%. Certainly, changing a county's population by that amount is implausible, but it does illustrate the bias in migration towards urban areas. Remembering that the measure of reward was not significant in the model, we can examine the measure of risk. Increasing the aggregate measure of risk by one percentage point is associated with an expected 24.57% decline in anticipated in-migration. We next turn our attention to the Supersector model to see what insights are available by drilling down from the aggregate measures of risk and return to the industry level.

BLS Supersector Model Results

The QCEW data provides a richness that permits us to dig deeper to try and find the subtleties masked by a single measure of risk and return. To do so, we deepen our analysis by aggregating the employment variables at the Supersector level constructed by the Bureau of Labor Statistics (BLS). This maps the 2-digit NAICS codes into 11 categories, divided between Goods-Producing Industries and Service-Providing Industries. The breakdown can be found in Table 1-9 below.

Table 1-9: BLS Supersector names and corresponding component 2-digit NAICS

Industry Type	**Supersector**	**Component NAICS**
Goods-Producing	Natural Resources and Mining	11, 21
	Construction	23
	Manufacturing	31, 32, 33
Service-Providing	Trade, Transportation, and Utilities	42, 44, 45, 48, 49, 22
	Information	51
	Financial Activities	52, 53
	Professional and Business Services	54, 55, 56
	Education and Health Services	61, 62
	Leisure and Hospitality	71, 72
	Other Services[3]	81
	Government	91, 92, 93

Source: https://www.bls.gov/sae/additional-resources/naics-supersectors-for-ces-program.htm

Using these expanded employment categories, we calculate the mean employment growth rate (return) and the standard deviation of employment growth rate (risk) for each Supersector and utilize these variables in a more extensive version of the aggregate portfolio regression. For ease of digestion, the portfolio variables are displayed in a table showing the direction of the coefficient with significance stars, where appropriate. The additional independent variables are then found in a traditional regression table below the Supersector table. Full results are reported in Table A.2 of Appendix A. Again, the model was run across the three different population scales as a robustness check. The results are reported below:

[3] NAICS 81 includes Repair and Maintenance, Personal and Laundry Services, Religious, Grantmaking, Civic, Professional, and Similar Organizations, and Private Households as subsectors.

Table 1-10: Supersector model results

	Industry Type	Supersector	Population Cut 1	Population Cut 2	Population Cut 3
Return	Goods-Producing	Natural Resources and Mining	-***	-***	-***
		Construction	+***	+***	+***
		Manufacturing	-**	-**	-
	Service-Providing	Trade, Transportation, and Utilities	+***	+	+
		Information	-***	-**	-***
		Financial Activities	+	+	+
		Professional and Business Services	-**	-	-
		Education and Health Services	-	-***	-***
		Leisure and Hospitality	+	+	+
		Other Services	+	+	+
		Government	+	+	+**
Risk	Goods-Producing	Natural Resources and Mining	+***	+***	+***
		Construction	-***	-***	-***
		Manufacturing	-	-	-
	Service-Providing	Trade, Transportation, and Utilities	+**	+	+
		Information	+	+	+**
		Financial Activities	-	-	-*
		Professional and Business Services	-*	-**	-
		Education and Health Services	+**	+*	+***
		Leisure and Hospitality	-***	-***	-***
		Other Services	+	+	+
		Government	-	-**	-**

Table 1-11: Full CO Supersector panel - additional independent variables, population cut scale version 2

County to County In-Migration Count	Coef.	St.Err.	t-value	p-value	95% Conf	Interval	Sig
Distance between Counties (100s mi.)	-2.125	0.158	-13.48	0.000	-2.433	-1.816	***
Distance between Counties Squared (100s mi.)	0.001	0.000	11.12	0.000	0.001	0.001	***
Median Household Income ($1,000s)	0.041	0.008	5.12	0.000	0.025	0.057	***
Population Density	0.000	0.000	3.37	0.001	0.000	0.001	***
USDA Amenity Scale	-0.058	0.051	-1.14	0.255	-0.159	0.042	
Population Cuts							
< 15,000	0.000						
15,000 to 75,000	1.070	0.223	4.80	0.000	0.633	1.507	***
> 75,000	2.492	0.221	11.30	0.000	2.060	2.925	***
Mean dependent var		88.562	SD dependent var			2,052.423	
Number of obs		40,387	Chi-square			7,869.594	
Prob > chi2		0.000	Akaike crit. (AIC)			2,756,162.293	

**** p<0.01, ** p<0.05, * p<0.1*

In examining the results, it is easiest to begin by looking at Table 1-11, the additional independent variables. As the additional independent variables were again consistent across population cuts, the coefficients for Population Cut Scale 2 are presented as a midpoint version. The expanded model shows that distance, density, and household income continue to show up with significance and largely intuitive signs. The only discrepancy with the prior model is that the natural amenity index no longer shows as significant, though it continues to show a negative coefficient. Additionally, the population cuts continue to show a bias toward urban migration. Translating the coefficients into marginal effects shows similar, though slightly diminished, magnitudes compared to the aggregate risk/return model. Table 1-12 presents the marginal effects for all significant variables from the expanded model. The table of marginal effects shows a strong preference for both local and metro migration. The marginal effect of distance (-88.06) shows that increasing the distance between county pairs by 100 miles would be associated with an 88.06% decrease in the expected migration count, all else equal. Similarly, changing a destination county from a small-population county to a large-population county has an associated increase in expected migration of 1,108%.

Transitioning to an examination of the Supersector portfolio variables shows significance for several of the industry-level measures of return and risk, as well as an encouraging level of consistency across the three population cut scales. Amongst the return variables, the mean employment growth rate for the Construction Supersector has a positive coefficient, while the Natural Resources and Mining Supersector and the Information Supersector show negative coefficients. The risk measures for the Supersectors showing significance and consistency across models are split equally between positive and negative coefficients. The standard deviation of employment growth rates for the Natural Resources and Mining and the Education and Health Services Supersectors have positive coefficients, while the Construction and the Leisure and Hospitality Supersectors display negative coefficients. Utilizing the marginal effects in Table 1-12 shows that a one percentage point increase in the mean employment growth rate for a Supersector is associated with impacts on expected migration ranging from -13.24% to 32.45%. The comparative range of marginal effects for the risk variables is -12.72% to 7.25%.

Table 1-12: Expanded model marginal effects

Variable	**Marginal Effect**	**Significance**
Mean Employment Growth Rate		
Natural Resources and Mining	-13.238	***
Construction	32.445	***
Manufacturing	-6.293	**
Education and Health Services	-7.040	***
Information	-9.877	**
Std. Deviation of Employment Growth Rate		
Natural Resources and Mining	4.081	***
Construction	-6.480	***
Education and Health Services	7.251	*
Government	-12.716	**
Leisure and Hospitality	-10.327	***

Professional and Business Services	-2.858	**
Additional Independent Variables		
Distance between Counties (100s mi.)	-88.057	***
Distance between Counties (100s mi.) Squared	0.100	***
Median Household Income ($1,000s)	4.185	***
Population Density (Persons/square mile)	0.000	***
Population Cuts		
Small: < 15,000		
Medium: 15,000 to 75,000	191.538	***
Large: > 75,000	1108.542	***

Additional Supersector Investigation

The richness of the Supersector dataset permits deeper explorations into coefficients that are puzzling at first glance. The negative coefficient on the Information Supersector return variable is one such instance. The negative coefficient is initially puzzling, given the name of the sector and the economy-wide transformation of the past two decades, but makes more sense with an understanding of the industries that comprise the group. This Supersector is made up entirely of NAICS 51, which contains the following industries:

Table 1-13: NAICS 51 component industries

511	Publishing Industries (except Internet)
512	Motion Picture and Sound Recording Industries
515	Broadcasting (except Internet)
517	Telecommunications
518	Data Processing, Hosting, and Related Services

As we can see from the above table, many of the component industries in this Supersector are ones that have experienced significant decline or consolidation over the past few decades. It is, therefore, not surprising that the data show a persistent contraction in this Supersector over the study period.

Additionally, these industries may be less likely to require migration, as they may be more conducive to remote employment. When we calculate the population weighted average annual return for this Supersector we get -1.93%. So, we have a case where migration is occurring, but Information sector employment is becoming a less important sector for migration decisions. Figure 1-3 below shows the annual employment counts for the five Colorado counties with the highest Information Supersector employment.

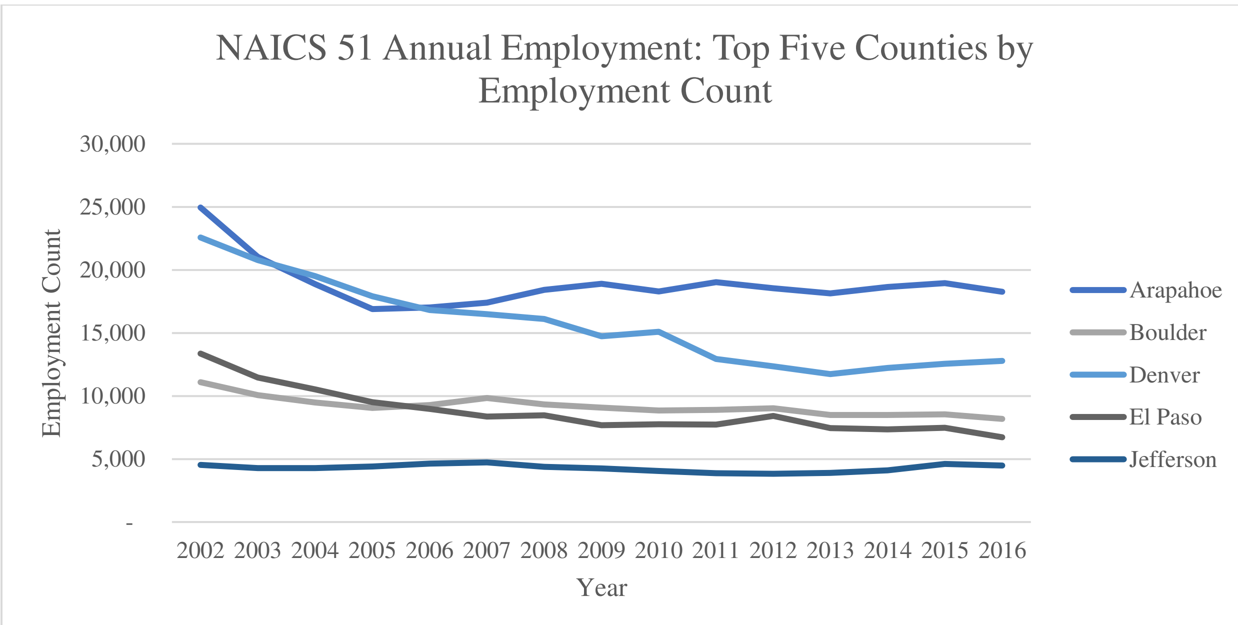

Figure 1-5: Annual Employment Count NAICS 51 (Information Supersector)

This time series graph shows a persistent loss of jobs in the Information Supersector over the duration of the study period. The main contraction appears to occur during the first half of the 2000s, likely in the wake of the Dot Com Bubble crash and the September 11th attacks. Most of the counties shown in Figure 1-5 then appear to stabilize, with Denver County being the exception that continues a substantial decline through 2013. Employment in this Supersector for Denver County went from 22,568 jobs in 2002 to 12,787 in 2016. Those losses equate to a 43% contraction in NAICS 51 for Denver County and a 34% contraction for the top five counties. The top five counties represented here comprised 49% of the statewide total of NAICS 51 employment in 2002, with Denver County alone accounting for 14%. By the end of our time window, these counties still accounted for 43% of statewide employment in this Supersector, while Denver County's share of overall employment fell to 11%. These numbers demonstrate that this was not an isolated incident, but representative of a statewide contraction in this industry. We next look at a comparison of the various models explored in this paper to identify the level of data that best balances insight with accuracy.

Model Comparison

Increasing the level of detail with respect to the employment data yielded returns in the form of model accuracy. Moving from the baseline model with aggregate risk and return measures to a model including population cuts resulted in a substantial decrease of the Akaike information criterion (AIC) from 6,009,669 to 3,245,057. Additionally, moving from the aggregate measures of risk and return to the Supersector portfolio approach further decreased the AIC to 2,756,162. This moderate improvement is still impressive when accounting for the penalty AIC places on models with additional variables and recognizing that this expanded model resulted in a net increase of 20 independent variables. However, this decrease in AIC should be tempered with the realization that the significance of our independent variables was not entirely consistent across the three population cut schemes. This inconsistency, coupled with the increased difficulty of explaining the counter-intuitive negative coefficients on some return variables and the positive coefficients on some of the risk variables, points towards the aggregate portfolio measure model with population cuts as providing the best application of our original research aim of embedding measures of portfolio return and risk into a model of migration, with the more detailed data being saved to provide context and support for specific case studies.

1.6 Conclusions

This work aims to further the regional portfolio field by introducing employment portfolio measures into a model of migration. Our secondary aim was to embark on an investigation of the best level of employment information to utilize in a model that was defensible, while remaining explainable. This was done with the aim of trying to explain why individuals are attracted to a particular place. The model of migration containing aggregate portfolio measures of risk and return best accomplishes these dual aims. The largest puzzle revealed by our analysis was the lack of significance on the measure of portfolio return. One potential explanation for this could be that potential in-migrants consider economy wide risk, but only concern themselves with their personal expected return and that of their family. This is in line with the self-selection models of migration put forth in works such as Borjas, Bronars, and Trejo (1992), Dostie & Léger (2009), and Kennan & Walker (2011). Additionally, Marshall (1920) and Krugman (1992) both discuss the idea of labor pooling as a risk reduction technique that also provides agglomeration effects. Our investigation shows evidence of this effect by revealing most in-migrants to Colorado select higher population areas. Additionally, Figure 1-2 shows that workers in the largest counties are achieving higher risk-adjusted returns compared to other areas. The lack of consistency across the Supersector portfolio measures illustrates that the best way local and regional governments can attract migrants is by focusing on providing an overall stable environment, rather than by trying to focus on any one specific sector.

This chapter furthers the discussion on regional portfolio theory by showing that accepted measures of regional risk and return can provide explanatory power in a model of migration. Additionally, it shows that county-level analysis with these measures provides worthwhile insight for academics and policy makers. With regards to the utilization of employment data in migration models, the realization that deep, industry specific employment data may not be necessary, may provide a degree of relief to those wishing to continue exploring the convergence of migration and regional portfolio theory.

Within the context of the empirical framework explored here, an interesting next step would be to utilize the measure of distance more fully by segmenting the sample and re-running the analysis. It is likely that the motivations for moving differ substantially by segment, and therefore the appeal and strength of the portfolio measures likely also vary between groups. It is easy to imagine a band occurring somewhere between zero and 100 miles where the psychic costs of moving are minimal, and the decision is driven mainly by considering the economic benefits. In the case of Colorado, 70 percent of the migration was internal to the state, with the remaining 30 percent being out of state migrants. In fact, movement between the counties of Adams, Arapahoe, Denver, and Jefferson totaled approximately one-third of the total in-migration count. While the top 10 originating counties in Colorado accounted for 62 percent of total migrants, the top 10 originating counties outside of Colorado only made up 8 percent of total in-migrants. It would be interesting to compute these percentages for other states to get a sense of how Colorado compares. This would allow us to see if Coloradoans have a greater tendency to remain in the state compared to neighboring states, giving insight to how attitudes vary. Additional work in this realm of the literature would provide more insight and stronger evidence on best practices regions can employ to foster stable and competitive environments that will increase their likelihood of being selected by potential migrants.

Chapter 2 - A Review and Comparison of Measures of Economic Resilience

2.1 Introduction

How a region responds to, or weathers, an economic downturn plays a substantial role in a region's appeal to potential in-migrants. As the occurrence of recessions have become less frequent in the past half century because of increased hands-on guidance from the Federal Reserve, the impact of a negative economic event is now perceived as more significant. The Great Recession brought about a new wave of publications in the literature on the concept of economic resilience. The study of how regions handle negative shocks is an important complement to the more pervasively studied topic of the determinants of specific regional growth paths. The work on resilience following the Great Recession brought about solid advances in the theoretical formulation of economic resilience – progressing from engineering resilience through ecological resilience to adaptive resilience: a recognition that the economy that entered the recessions differs fundamentally from the one that emerged from the event (Martin, 2012).

In this chapter, we test seven models of economic resilience to identify the determinants of resilience. We identify a set of independent variables that presents as significant across a range of formulations of resilience. These results persist across two types of dependent estimators, as well as the economic indicators used as inputs to these representations of resilience. We find that the empirical representation of resilience is an area in need of greater investigation, owing to the lack of correlation across our seven proxies of resilience. While different aspects could one day be combined into a multi-dimensional index, doing so now, without a full understanding of the subtleties each indicator contains would risk dampening our understanding of the net effects by not developing an appreciation of the full effect of each component piece. This approach risks leaving our understanding of resilience incomplete. This chapter augments the empirical approach of Deller, Conroy, & Watson (2017) by changing the unit of analysis from the county level to metropolitan statistical areas (MSAs). It also extends the work of Kitsos & Bishop (2018) by both applying their model to a U.S. data set and extending the approach to additional dependent variables. Our results show that the impact measure prescribed by Kitsos & Bishop (2018) is widely applicable to additional dependent variables and translates well to data from different countries. For Deller et al. (2017) we find that their county level results remain largely consistent when considering the MSA level, while our discussion section highlights some potential reservations with their dependent variable formulation as a measure of economic resilience.

2.2 Economic Resilience Literature

The first decade of the 21st century contained several economic recessions that significantly impacted people around the world. In the U.S. the most substantial of these were the recession that followed the dot com bubble and the September 11th attacks in 2001, and the Great Recession which lasted from 2007 to 2009. In both cases, the impact of these events at the local level varied greatly and often lasted well beyond the official recession end date, sometimes altering the trajectory of local economies for years to come (Han & Goetz, 2015). These differing growth patterns resulted in an increased interest by regional economists and economic geographers in what has been termed economic resilience (Foster, 2007; Martin, 2012). The study of resilience at the local level has been focused as much on defining what resilience should mean as it has been on the measurement of resilience (Foster, 2007; Ormerod, 2010). While concrete definitions of resilience have long existed in other disciplines, including ecology, biology, and psychology, its definition within the regional economic context remains murky. As stated in Martin & Sunley (2015), there is no universally agreed upon definition, no generally accepted methodology on how resilience should be measured, and currently no cohesive theory to relate it to other regional concepts. However, we can see in the literature how the concept, as applied to regional economics, has developed over time. Regardless of the definition, any examination of the empirical data quickly shows differences in resilience across regions. Di Caro (2017) reinforces that explaining this "why" is an open question, and one worthy of additional empirical investigation.

Resilience can be thought of as the main visible outcome variable measuring the level of risk adjustment. The diversification of industries within the employment base of a city, county, or region ties in directly to the level of resilience it experiences. Kort (1981) first laid out the theoretical model that showed the relationship between industrial diversification, stability, and city size. The previous chapter of this dissertation provided a deep dive into regional portfolio theory and how it can be utilized to help a location achieve the best risk adjusted return by implementing policies to encourage the development of an array of industries possessing imperfectly correlated returns. Because of the extensive literature review on this topic in the prior chapter, only an abbreviated overview is included throughout this chapter, but the tie-in to this paper is important enough to merit the inclusion of these ideas for a second time.

The first iteration of resilience as an economic concept is pinned to the belief that each regional economy moves along its own predetermined growth path. Any shocks that occur will temporarily move the economy off that trajectory, however, it will return to that prior path if given enough time (Holling, 1973; Pimm, 1984). Therefore, in this "engineering resilience" the items of interest are the size of the dip and the speed with which an economy returns to its prior growth path. Engineering resilience carries an assumption that the system is at or near equilibrium prior to experiencing the shock. This type of resilience draws a strong parallel to the Plucking model of Friedman (1993) and is akin to an economy operating at or near its steady state growth path (Simmie and Martin, 2010).

The second type of resilience draws on the ecological sciences and recognizes that multiple equilibria can exist within a given system. Here, the main point of interest is the magnitude of shock that a regional economy can absorb before being thrown out of its current growth path and into a new state of being (McGlade et al., 2006). The larger the shock that a system can absorb without moving past this threshold into a new state of being, the more resilient the system is said to be. Additionally, the quality of the new state must be considered. In a system with multiple equilibria, the new state could be of a higher quality, in which case the system would be seen as more resilient than one in which the new equilibrium is of a lower quality than the initial state. This new state can include one where the effects of the recessionary event are felt long after the factors that led to the event have been resolved (Simmie and Martin, 2010). This state of being has been termed hysteresis, also known as path dependence, a term which initially found its way into the economic domain from its origins in the natural sciences (Martin, 2012). The idea of hysteresis does not require the concept of an equilibrium, but rather can refer to an event that permanently impacts the growth path of economy, such as in Romer (2001).

In both prior discussions of resilience, one item that has not been addressed is the formation of the underlying economy. When a negative shock occurs, the economy that comes out of the recession is often fundamentally different from the economy that went into the recession (Martin, 2012). These differences are a result of firms going out of business where those assets are then employed in new pursuits, changes in the employment structure and education level resulting from labor migration, and new investments resulting from decreased costs as the recessionary event drove down labor costs (Han and Goetz, 2015). Martin has termed this version of resilience as adaptive resilience (Simmie and Martin, 2010; Martin, 2012). This perspective incorporates Schumpeter's idea of creative destruction, that the economy responds to shocks by rearranging its component pieces, and lands on a new growth path as a result. Ramlogan and Metcalfe (2006) make the point that economies are naturally restless, since the main driver of capitalism is actually knowledge, and knowledge is always increasing and evolving. The main theme from Simmie and Martin (2010) is that if regional economies are never at rest in an equilibrium state, and components are always reorganizing through churn and creative destruction, then resilience is not a response to a shock but an ongoing process and capacity that is built over time by the actions of actors and implementation of policies. Similarly, Evenhuis (2020) asserts that resilience is an underlying capacity to cope with disturbances, while the process that we actually can see is the adaptation of a region.

The level of resilience may also be impacted by the choice of industries in which a location specializes. Ray, MaLachlan, Lamarche, & Srinath (2017) argue that goods producing industries are less resilient to shocks compared to service producing industries. Therefore, it is not simply the degree of diversification that is important, but the specific nature of the underlying industries. This could depend, to a large extent, on the educational attainment of workers, which varies by industry. The empirics of Kitsos

& Bishop (2018) suggest that areas with a higher percentage of individuals holding bachelor's degrees were more resilient, and colloquial evidence from the pattern of recovery following the Great Recession and housing market collapse in the United States would concur with that assertion.

Dialing in the optimal degree of industrial specialization for a location is an important balancing act to maximize economic growth, economic resilience, and the level of cross- and inter-industry spillovers that occur. DiCaro (2017) emphasizes the role of diversity in fostering an environment that encourages Jacobs externalities, which are those that occur in an urban setting due to diversity of industry. Jacobs (1969) argues that the most important knowledge spillovers tend to be external to a firm's industry, and therefore diversity of industry is important. She continues, saying that since this diversity is greatest in cities, it is cities that are the source of innovation. Harrison et al. (1996) echoes this when concluding that a more diverse economy provides the environment for the exchange of skills requisite for new industries and fields to emerge. Deller & Watson (2016b) stress that walking this razor's edge is important to balance maximizing resilience (which occurs through greater diversification and minimizes the impact of negative shocks while leading to a quicker recovery) while also maximizing economic returns (which development theory says results from the competitive advantages that are presented through specialization). Deller & Watson (2016b) maintain that a healthy balance can be found by distributing the goals temporally, with growth as a short-term goal (developing the competitive advantage of a single industry) and resilience and stability occupying a long-term goal by balancing the range of industries in a geographic location.

It is important for the practitioners of regional economics to work toward an agreed upon definition and empirical structure of economic resilience, as not doing so risks introducing confusion into the literature. Watson & Deller (2022) makes this point when comparing the concepts of economic stability, resilience, and robustness. They echo the idea that clarity between these ideas and an extensive mapping of these concepts is required. Martin's numerous works on the topic, both alone (Martin, 2012) and with co-authors (Simmie & Martin, 2010; Fingleton et al., 2012; Martin & Sunley, 2015) makes this same point while compiling and delineating existing works according to the engineering, ecological, and evolutionary/adaptive nomenclature. The diversity of definitions persists across disciplines as well. Tiernan et al. (2019) finds a similar lack of a cohesive definition in the disaster resilience literature in their work that reviews papers published after 2012. This literature also utilizes the term adaptive resilience, but does so in reference to “a community’s behavior after the disaster.” Additionally, agreement is needed on if resilience refers only to the ability to resist economy-wide impacts, such as recessions, or if it includes “man-made” impacts, such as the city-county consolidations discussed in Matti & Neto (2022). Their work finds that after controlling for demographic, economic, and geographic factors that consolidation does not promote stability across several measures, such as employment, the unemployment rate, per capita income, or count of establishments.

In addition to the disagreement surrounding the colloquial definition of resilience, the concept of resilience lacks a cohesive empirical definition since it comprises multiple effects that are themselves still frequently studied components of the regional economic literature. Ray et al. (2017) discusses that defining and quantifying resilience requires splitting out the industry-mix effects from the regional effects. This makes problematic the fact that most studies of economic resilience rely heavily on changes in the employment or unemployment rate as their dependent variable (Di Caro, 2015; Fingleton & Palombi, 2013; Lee, 2014). A notable exception is Ray et al. (2017) which has the distinct aim of partitioning regional employment growth rates into region effects, industry-mix effects, gender-mix effects, and interaction effects. Similarly, Dormady et al. (2022) contributes a fully formulated theoretical and empirical structure for resilience, however, they are operating from the viewpoint of businesses employing production functions, as opposed to local governments. Fortunately, more recent studies such as Deller & Watson (2016a) have taken to testing several resilience measures, opening the door to a larger discussion about the strengths and weaknesses of the various dependent variables. For example, Deller & Watson (2016a) examines four measures of stability: the unemployment rate, the employment/population ratio, the concentration of establishments, and average weekly wages. In their extensive literature review covering spatial economic resilience, Modica & Reggiani (2015) detail the range of variables that authors have used to measure resilience, with the number of variables used ranging from 1, in Martin (2012), to an index comprised of 29 different variables in Cutter et al. (2008). They include a table detailing the utilized variables for a selection of ten papers, which include everything from the impartiality of courts to disaster prevention plans to voter participation.

Complementing the need to move beyond one dimensional explanations of resilience is a parallel push by the literature to narrow in on the determinants of resilience. The impetus for this vein of the literature can be seen as early as Fingleton, Garretsen, & Martin (2012) who touch upon several potential determinants in Section 2 of their work, but ultimately leave this exploration for future research. Other authors continued this search for regional determinants, such as Fingleton & Palombi (2013) and Diodato & Weterings (2014). Additionally, after constructing and testing their measure of resilience, Han & Goetz (2015) profess that "an exploration of the determinants of resilience based on the internal properties of individual counties would be an interesting next step". They do however offer a menu of potential areas in which to find determinants such as: "income, population density, land area, population age, educational attainment, social capital stocks, and industrial structure". The literature takes up the hunt for these independent variables with two solid investigations coming from the main inspirations for this paper: Deller et al. (2017) and Kitsos & Bishop (2018). Both look to uncover which independent variables contribute positively, or negatively, to the resilience of a locality. Extrapolating on these works, we now move on to the description of the variables used in our investigation.

2.3 Data Description

To explore the multi-faceted nature of resilience, we examine several dependent variables. The variables used in this investigation are total employment, the employment – population ratio, the unemployment rate, per capita income, average annual wages and salary, and real GDP per capita. This study uses data from 2004 through 2014. The data from 2004 to 2007 are used to create a baseline of conditions prior to the recession, while the inclusion of data up to 2014 ensures that our variables capture the lowest points of the recession in each metropolitan area. This timeline was chosen because it lines up with the periods studied in Deller et al. (2017) and Kitsos & Bishop (2018). This will allow us to compare our results more easily to these prior works. Metropolitan statistical areas (MSAs) were selected as the unit of analysis for this study. This selection allows for an extension of the work from Deller et al. (2017), whose investigation took place at the county level. Complete data for our time frame was available for 360 of the 384 metropolitan statistical areas in the United States, defined according to the Office of Management and Budget (OMB). The OMB's 2010 standards state that each metropolitan statistical area must contain a minimum of one urbanized area which contains a minimum population of 50,000 inhabitants (US Office of Management and Budget, 2010).

A main goal of this chapter is to compare multiple definitions of economic resilience and their corresponding empirical constructs. This will provide insight on the need to further specify what a particular definition of resilience is representing in future empirical investigations. To accomplish this, we draw on two main sources of inspiration. The first is an application of Kitsos & Bishop (2018). We extend their work by testing their Impact variable on a U.S. dataset to draw comparisons with their work across Great Britain's Local Authority Districts (LADs). Additionally, we extend the use of this measure beyond their sole investigation of the employment – population ratio to also look at real GDP per capita and unemployment rates.

To smooth out the data and obtain an accurate picture of the impact the recession had on the various dependent variables, Impact is measured as X_j , the average from 2004 – 2007 for an MSA minus X_i , the average of the four lowest values occurring between 2008 – 2014.

$$Impact = X_j - X_i \tag{2.1}$$

The only exception to this rule is for the Impact variable for unemployment. In this case it is the average of the four highest values occurring between 2008 – 2014 that is subtracted from the initial conditions average. This difference is then multiplied by -1 for the interpretation of coefficients to be consistent with the other impact models. The resulting dependent variables can be interpreted as the gap between the pre-recession conditions and the smoothed trough point of the recession. The descriptive statistics for these measures are found below in Table 2-1.

The second main source of inspiration for this investigation comes from Deller et al. (2017). In this work, the authors calculate the variance-mean ratio (VMR) as their measure of stability, for four separate stability indicators. They examine total employment, unemployment rates, per capita income, and average annual wages and salary. Their investigation takes place at the county level and adds a novel construction to account for spatial dependence and heterogeneity. The variance-mean ratio, also known as the index of dispersion, is a normalized measure that is used to indicate if observed occurrences are clustered in time or space or more evenly distributed. The formula for the construction of the ratio is as follows:

$$\sum_{t=2007}^{2013} \frac{\left(\mu_t - \underline{\mu}\right)^2}{\underline{\mu}} \tag{2.2}$$

In the above equation μ_t represents the value of a particular stability indicator in year *t*. The distance between the single year observation and the mean is calculated and then normalized by dividing by the mean. This value is calculated for each year from 2007 to 2013 and summed to produce the VMR for each unit of observation. The VMR inherits properties from the assumption of its underlying Poisson distribution that the variance is equal to the mean. Because of this assumption, a VMR of 1 indicates a Poisson process where the occurrence of the observed variable is a random walk over time. Values greater than 1 indicate that the data is more variable or volatile, while values less than 1 are said to be under-dispersed and occur more uniformly across time than the random walk indicated by a Poisson process.

Table 2-1: Descriptive statistics for dependent variables

VARIABLES	mean	sd	min	max
Income VMR	0.752	0.634	0.0269	5.779
Wages VMR	0.509	0.376	0.0489	3.686
Employment VMR	1.759	4.626	0.00442	46.16
Unemployment VMR	2.932	1.503	0.317	11.46

IMPACT_GDP	1.464	3.967	-14.78	45.97
IMPACT_EMP_POP	0.0264	0.0215	-0.0322	0.133
IMPACT_UNEMPLOYMENT	3.849	1.786	0.500	12.13

(n = 360)

The independent variables selected ensure coverage of the same themes utilized in both Deller et al. (2017) and Kitsos & Bishop (2018). To do this, the independent variables from both papers were categorized and then independent variables were selected with an aim of consistency, based on data availability. Table 2-2 below shows the included variables along with their thematic category.

Table 2-2: Independent variables and economic themes

Theme	**Variable**
Initial economic conditions	Employment-Population Ratio (2007) Gini coefficient of income equality
Industrial Diversity and Employment Composition	Percentage of Employment in Manufacturing Percentage of Employment in Finance, Insurance, Banking Percentage of Employment in Construction Percentage of Employment in Services Production Herfindahl-Hirschman Index
Entrepreneurship	Average Establishment births per 1000 population Percentage of Households with Self-employment income
Demographics	Percentage of Population 18-35 Percentage of Population 36-55 Percentage of Population 65+ Percentage of Population Black or African American
Pop Density	Population Density
Geography	New England Census Division Middle Atlantic Census Division East North Central Census Division West North Central Census Division South Atlantic Census Division East South Central Census Division West South Central Census Division Mountain Census Division Pacific Census Division

The U.S. Census County Business Patterns survey provides annual MSA level data on employment down to the 6-digit NAICS code level. The 2007 data was used in the construction of the variables representing industrial diversity and employment composition. For the employment composition the 2-digit NAICS code aggregates were used to calculate the percentage of employment in manufacturing; the percentage of employment in finance, insurance, and banking; the percentage of employment in construction; and the percentage of employment in services production. The 2-digit NAICS codes were consolidated according to the Bureau of Labor Statistics Supersectors presented in Table 2-3.

Table 2-3: BLS Supersector composition by 2-digit NAICS code

Industry Type	Supersector	Component NAICS
Goods-Producing	Natural Resources and Mining	11, 21
	Construction	23
	Manufacturing	31, 32, 33
Service-Providing	Trade, Transportation, and Utilities	42, 44, 45, 48, 49, 22
	Information	51
	Financial Activities	52, 53
	Professional and Business Services	54, 55, 56
	Education and Health Services	61, 62
	Leisure and Hospitality	71, 72
	Other Services*	81
	Government	91, 92, 93

Source: https://www.bls.gov/sae/additional-resources/naics-supersectors-for-ces-program.htm

The 3-digit NAICS code data were used in the construction of the Herfindahl-Hirschman Index (HHI). While the HHI is more commonly used to identify the share of a company within an industry, it can also be used in a location specific manner to measure the level of specialization by comparing employment shares across industries (Rhoades, 1993). The HHI is frequently employed by the U.S. Department of Justice when considering the effects of firm mergers in an industry. As a measure of industrial diversity, the HHI is calculated by summing the squares of the employment shares of each 3-digit NAICS code sector. Using the 3-digit NAICS code data results in an HHI constructed from 87 different sectors, with values that range from 0.0272 to 0.164. A higher HHI reflects a higher level of specialization in an MSA's economy, while a lower value reflects greater diversification.

The demographic population variables based on age, race, and employment status come from the American Community Survey (ACS) published by the U.S. Census. Additionally, the ACS also provided the data for the Gini coefficients of income equality. The data for the average establishment births per 1,000 population is a three-year average covering 2004 to 2007. This helps to smooth the data on what can often be a more volatile variable. The establishment births data is provided by the U.S. Census' Statistics of U.S. Businesses (SUSB) survey. This survey provides establishment birth and death data on an annual basis for multiple geographic levels.

Table 2-4: Descriptive statistics for independent variables

VARIABLES	mean	sd	min	max
Percentage of Population 18-35	0.241	0.0415	0.154	0.439
Percentage of Population 36-55	0.278	0.0227	0.188	0.331
Population Density	259.4	280.5	6.845	2,269
Employment-Population Ratio	0.598	0.0519	0.394	0.736
Percentage of Households with Self-employment income	0.116	0.0261	0.0571	0.227
Percentage of Population 65+	0.130	0.0331	0.0530	0.312
Percentage of Population Black or African American	0.104	0.108	0	0.507
Gini coefficient of income equality	0.443	0.0276	0.370	0.534
Herfindahl-Hirschman Index	0.0403	0.0141	0.0272	0.164
Percentage of Employment in Manufacturing	0.129	0.0708	0	0.534
Percentage of Employment in Finance, Insurance, Banking	0.0608	0.0208	0.00915	0.194
Percentage of Employment in Construction	0.0652	0.0239	0.0200	0.179
Percentage of Employment in Services Production	0.798	0.0653	0.430	0.944
Average Establishment births per 1000 population	2.612	0.786	0.916	6.719
New England Census Division	0.0417		0	1
Middle Atlantic Census Division	0.0917		0	1
East North Central Census Division	0.144		0	1
West North Central Census Division	0.0944		0	1
South Atlantic Census Division	0.206		0	1
East South Central Census Division	0.0750		0	1
West South Central Census Division	0.122		0	1
Mountain Census Division	0.0917		0	1
Pacific Census Division	0.133		0	1

(n = 360)

To provide a level of regional control without sacrificing a large number of degrees of freedom, regional fixed effects are included based on an MSA's Census division. There are nine Census divisions, shown in Table 2-5 below.

Table 2-5: U.S. Census divisions and component states

<u>New England</u> Connecticut Maine Massachusetts New Hampshire Rhode Island Vermont	<u>Mountain</u> Arizona Colorado Idaho New Mexico Montana Utah Nevada Wyoming	<u>East North Central</u> Indiana Illinois Michigan Ohio Wisconsin	<u>West North Central</u> Iowa Kansas Minnesota Missouri Nebraska North Dakota South Dakota	<u>South Atlantic</u> Delaware District of Columbia Florida Georgia Maryland North Carolina South Carolina Virginia West Virginia
<u>East South Central</u> Alabama Kentucky Mississippi Tennessee	<u>West South Central</u> Arkansas Louisiana Oklahoma Texas		<u>Pacific</u> Alaska California Hawaii Oregon Washington	<u>Middle Atlantic</u> New Jersey New York Pennsylvania

Source: https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html

To finalize the variable set for the regression models, the full set of variables in Table 2-4 were included, the model was run, and then the variance inflation factors (VIF) were calculated for each independent variable and the overall model. The results of the VIF test are presented below in Table 2-6. The VIF statistic is a measure of the multicollinearity present between the independent variables of a model. Our results show two variables with VIFs exceeding a value of 10, the general limit of concern. These variables are the percentage of employment in manufacturing and the percentage of employment in services production. To resolve this issue the model was run with each variable omitted in turn and the VIFs were recalculated. The results showed that omitting either variable was sufficient to resolve the VIF concern, with no remaining variable's VIF exceeding a value of 5. The model omitting the percentage of employment in services production was selected, as the average adjusted R-squared across the different dependent variables was greater compared to removing the percentage of employment in manufacturing.

Table 2-6: Variance Inflation Factors (VIF) test results

Variable	VIF – Full Model	VIF – Percentage Employment in Services removed
Percentage of Population 18-35	3.77	3.74
Percentage of Population 36-55	3.3	3.3

Population Density	1.47	1.4
Employment-Population Ratio	2.84	2.81
Percentage of Households with Self-employment income	1.95	1.91
Percentage of Population 65+	2.91	2.77
Percentage of Population Black or African American	1.78	1.66
Gini coefficient of income equality	1.59	1.52
Herfindahl-Hirschman Index	1.12	1.1
Percentage of Employment in Manufacturing	27.13	1.72
Percentage of Employment in Finance, Insurance, Banking	1.62	1.61
Percentage of Employment in Construction	5.86	1.96
Percentage of Employment in Services Production	23.91	
Average Establishment births per 1000 population	2.45	2.44
Middle Atlantic Census Division	3.07	3.06
East North Central Census Division	3.91	3.9
West North Central Census Division	3.08	3.08
South Atlantic Census Division	4.95	4.93
East South Central Census Division	2.7	2.68
West South Central Census Division	3.65	3.63
Mountain Census Division	3.04	3.03
Pacific Census Division	3.78	3.77
Mean VIF	4.99	2.67

2.4 Results

The empirical portion of this work tests seven different models, as well as testing the significance of the independent variables across models. While it is difficult to draw a single conclusion from this many models, what stands out is that no two definitions of economic resilience behave identically with respect to finding significant determinants. All models contain determinants that are statistically significant and are intuitive within the context of economic theory. The fact that most independent variables are found to be significant across the multiple models shows that the literature has made substantial progress in identifying the underlying factors that impact resilience. The main takeaway these results present is that different dimensions of resilience are impacted by different independent variables and the regional economic literature needs to develop a framework that considers these differences when selecting a variable, or ideally multiple variables, to represent economic resilience.

We first turn our attention to the results of the variance-mean ratio (VMR) models. We test four dependent variables to show the divergence of results depending on how resilience is defined. The four models can be grouped into two income-based models and two employment-based models. When interpreting the coefficients, it is useful to consider the income- and employment-based models separately because, as the correlation matrix in Figure 2.1 shows, these measures are not strongly correlated. While the wage VMR and income VRM have a correlation of 0.58, both display essentially no correlation with the employment VMR (0.12 and 0.14 respectively). Additionally, both income-based measures display a slight negative correlation with the unemployment VMR (-0.22 and -0.18 respectively). This means that as unemployment becomes less stable, income and wages become more stable. Deller et al. (2017) argue that this could be the result of an income floor provided by government income stabilization programs, such as unemployment insurance.

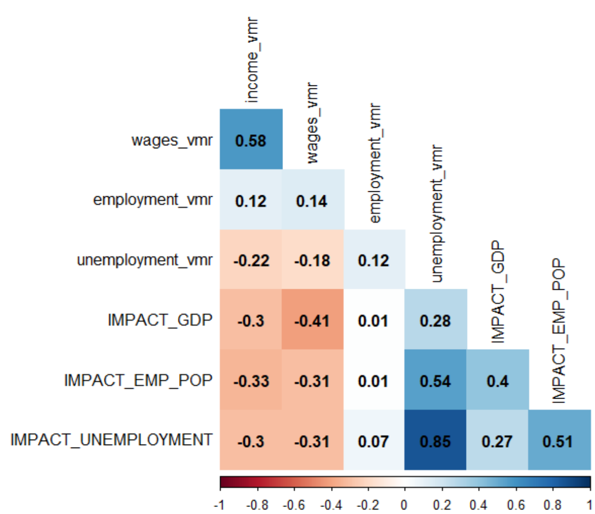

Figure 2-1: Correlation matrix for independent variables

This inverse relationship also presents itself in the coefficient of the employment-population ratio in 2007, which controls for economic conditions at the start of the recession. The coefficient is positive and significant for the income model, but negative and significant for the unemployment model. In interpreting the coefficients of the independent variables, one should keep in mind that higher levels of the VMR represent greater instability in the distribution over time, and values below 1 indicate a more stable distribution of values across the recession and recovery timeframe. This means that a higher employment-population ratio is associated with a less stable distribution of income variables, while it is also associated with a more stable distribution of unemployment rates. This could reflect that areas closer to full employment going into the recession experienced better outcomes after all other variables were controlled for, but still experienced variation in their incomes. The results suggest that the income and unemployment models cannot function as interchangeable measures of resilience but are both important pieces of a more complex characterization of resiliency.

In the employment and unemployment models, age demographics have the anticipated impact on stability. In the employment model, the higher the population over the age of 65, the more stable the distribution of employment is over time. This makes sense as these individuals have often removed themselves from the labor pool. On the opposite side of the coin, the unemployment model experiences a stabilizing force as the percentage of the population 18-35 increases. Additionally, the unemployment rate distribution is made less stable by higher levels of construction and manufacturing employment and the higher concentration of industries as expressed through the Herfindahl-Hirschman Index. Greater variability in employment rates through time were associated with higher levels of employment in construction and the financial, banking, and insurance sectors.

For the income VMR model, the percentage of the population who are Black or African American and the percentage of the population who are employed in finance, insurance, or banking had negative coefficients and therefore work to reduce the variability in income over time, while the greater the employment-population ratio and a higher level of income inequality, as measured by the Gini coefficient, are associated with less stable income distributions over time, as shown by their positive coefficients.

The wage VMR model contains the fewest number of significant determinants, with only the percentage of the population who are 65 and older and the Gini coefficient of income equality displaying statistical significance. The percentage of the population who are 65 and over has a positive coefficient, indicating the larger this group the more of a stabilizing effect it has on wage distributions over time. This is likely due to most individuals in this group being retired, so the larger the percentage of this group, the fewer wage earners there are to compete for jobs, reducing the supply and demand dynamics of the labor market. The Gini coefficient of income equality has a positive coefficient, paralleling the Income VMR model.

If we look at the remaining variables in our regressions, there are a few notable observations. The two variables that did not show up as significant in any of the VMR models are the percentage of the population 36 - 55 and the percentage of households with self-employment income. The binary census division variables were significant in most occurrences. In every case where one of these variables was significant, its coefficient was positive. The reference group is the New England census division, indicating that other regions experienced, on average, less stable distributions of their resilience variables in comparison. Three of these divisions were the only variables to show up as significant across all four models: the Middle Atlantic division, the East North Central division, and the East South Central division.

Table 2-7: Variance-Mean Ratio model results

	(1)	(2)	(3)	(4)
VARIABLES	Income VMR	Wages VMR	Employment VMR	Unemployment VMR
Percentage of Population 18-35	-1.558	0.0300	-13.87	-6.465**
	(1.040)	(1.062)	(10.84)	(2.520)
Percentage of Population 36-55	-0.446	2.101	-6.534	-5.965
	(1.908)	(2.234)	(14.28)	(3.447)
Population Density	-0.000145	3.67e-05	0.00830**	0.00107***

	(0.000110)	(7.44e-05)	(0.00268)	(0.000154)
Employment-Population Ratio	3.850**	0.772	-7.091	-9.143***
	(1.159)	(0.474)	(3.930)	(2.432)
Percentage of Households with Self-employment income	0.0528	-1.011	1.740	-6.682
	(1.885)	(1.281)	(7.543)	(3.778)
Percentage of Population 65+	-0.122	-1.488*	-38.19**	-5.354
	(1.191)	(0.781)	(13.42)	(3.057)
Percentage of Population Black or African American	-1.563***	-0.594	-2.980	-0.320
	(0.402)	(0.335)	(1.817)	(0.692)
Gini coefficient of income equality	7.038***	1.959*	14.45	-2.444
	(1.607)	(0.856)	(7.963)	(3.400)
Herfindahl-Hirschman Index	-1.994	0.0567	1.070	18.86*
	(1.165)	(0.692)	(8.493)	(8.601)
Percentage of Employment in Manufacturing	-0.954	-0.565	1.795	7.953***
	(0.840)	(0.328)	(2.022)	(1.278)
Percentage of Employment in Finance, Insurance, Banking	-3.935*	-0.995	22.03**	-0.0691
	(1.742)	(1.102)	(8.096)	(5.158)
Percentage of Employment in Construction	0.322	1.530	25.36**	22.00***
	(2.167)	(1.673)	(10.19)	(4.085)
Average Establishment births per 1000 population	-0.0691	-0.0721	0.511**	0.700***
	(0.0464)	(0.0476)	(0.195)	(0.124)
Middle Atlantic Census Division	0.331***	0.173***	1.091***	0.0845
	(0.0188)	(0.0217)	(0.305)	(0.0484)
East North Central Census Division	0.257***	0.134***	0.473***	0.238***
	(0.00955)	(0.0177)	(0.119)	(0.0350)
West North Central Census Division	0.420***	0.130***	1.422***	-0.0563
	(0.0286)	(0.0290)	(0.170)	(0.0370)
South Atlantic Census Division	0.347***	0.158***	0.368*	0.299***
	(0.0145)	(0.0212)	(0.195)	(0.0327)
East South Central Census	0.216***	0.108***	3.982***	0.261***

Division				
	(0.0147)	(0.0170)	(0.235)	(0.0348)
West South Central Census Division	0.507***	0.130***	1.023***	0.0734
	(0.0247)	(0.0237)	(0.160)	(0.0749)
Mountain Census Division	0.142***	0.0349	1.340***	0.374***
	(0.0227)	(0.0260)	(0.167)	(0.0536)
Pacific Census Division	0.267***	0.0226	1.376***	0.475***
	(0.0234)	(0.0188)	(0.128)	(0.0334)
Constant	-3.662**	-0.940	1.964	8.656***
	(1.398)	(0.592)	(4.224)	(2.321)
Observations	360	360	360	360
Adjusted R-squared	0.128	0.036	0.359	0.410

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Next, we consider the results of the Impact-based models in the vein of Kitsos & Bishop (2018). In the model showing the impact on the real GDP per capita, all the census division regional variables are significant, with negative coefficients. This shows that being anywhere outside of the New England area is associated with a reduction in the impact to real GDP per capita. The only other independent variable associated with a reduction in GDP impact comes from having a larger percentage of the population in the 18 to 35-year-old demographic. This again speaks to the benefits of having a large percentage of the working population in the earlier portion of their careers. This could be associated with decreased switching costs for younger workers or could point to age discrimination present in hiring processes. For the GDP Impact model, the percentage of the population who are Black or African American, employment in manufacturing, and a higher Herfindahl-Hirschman Index are all associated with a greater impact to real GDP per capita levels during the recession. Greater employment in manufacturing and a higher Herfindahl-Hirschman Index, which is indicative of a more specialized economy, result in a larger gap between the pre-recession average and the worst performing years during the recession across all three impact models. Both the Employment-Population ratio and Unemployment Impact models show that metro areas with a higher average number of establishment births per 1000 population are associated with worse performance. Without a way to separate purposeful (pull-based) entrepreneurship from necessary (push-based) entrepreneurship, in the context of this model, it is difficult to speculate on the dynamics underlying this variable.

For the Employment-Population ratio Impact model, the Gini coefficient is the only variable, outside of the regional fixed effects, that has an impact reducing effect on the dependent variable. In this model greater income inequality is associated with a smaller gap between the pre- and peak-recession time periods. Finally, in the Unemployment model, the 2007 employment-population ratio has an impact reducing effect. This variable serves as a proxy for economic conditions prior to the Great Recession and can be interpreted as metros that had a strong economy going into the recession fared better over the duration of the recession than those metros who were already experiencing economic difficulty at the start of the recession.

Table 2-8: Impact model results

VARIABLES	(1) GDP Impact	(2) Employment-Population Impact	(3) Unemployment Impact
Percentage of Population 18-35	-12.82*	0.0329	-1.961
	(6.134)	(0.0350)	(2.165)
Percentage of Population 36-55	-10.89	0.0129	1.755
	(7.374)	(0.0463)	(2.601)
Population Density	9.62e-05	3.82e-07	0.00147***
	(0.000933)	(3.03e-06)	(0.000322)
Employment-Population Ratio	-6.989	-0.0584	-17.64***
	(7.336)	(0.0348)	(2.435)
Percentage of Households with Self-employment income	-10.70	-0.0494	-1.998
	(7.947)	(0.0558)	(6.283)
Percentage of Population 65+	-5.039	-0.0570	-2.344
	(8.514)	(0.0484)	(5.636)
Percentage of Population Black or African American	5.326**	0.0178	0.337
	(2.027)	(0.0118)	(1.409)
Gini coefficient of income	2.315	-0.101*	-4.897

equality			
	(11.19)	(0.0521)	(4.522)
Herfindahl-Hirschman Index	31.49**	0.375***	14.09**
	(13.12)	(0.0962)	(4.765)
Percentage of Employment in Manufacturing	9.976***	0.139***	6.260***
	(2.858)	(0.0175)	(0.948)
Percentage of Employment in Finance, Insurance, Banking	13.96	0.0874	2.138
	(12.67)	(0.0646)	(6.593)
Percentage of Employment in Construction	25.11	0.158	20.32***
	(15.04)	(0.0983)	(5.181)
Average Establishment births per 1000 population	0.954	0.0128***	0.521***
	(0.521)	(0.00208)	(0.150)
Middle Atlantic Census Division	-4.058***	-0.0104***	-0.0648
	(0.252)	(0.00122)	(0.0827)
East North Central Census Division	-2.479***	-0.00168**	0.291***
	(0.133)	(0.000569)	(0.0841)
West North Central Census Division	-2.628***	-0.00677***	0.0224
	(0.249)	(0.00125)	(0.0430)
South Atlantic Census Division	-3.003***	-0.00604***	0.148
	(0.157)	(0.000486)	(0.0847)
East South Central Census Division	-2.525***	-0.00974***	-0.0255
	(0.180)	(0.000436)	(0.0879)
West South Central Census Division	-2.950***	-0.0106***	-0.242**
	(0.422)	(0.00113)	(0.104)
Mountain Census Division	-2.187***	-0.00242**	0.282*
	(0.166)	(0.000952)	(0.133)
Pacific Census Division	-2.601***	-0.00276**	0.814***
	(0.194)	(0.000840)	(0.0771)
Constant	7.207	0.0291	12.31***
	(7.260)	(0.0311)	(1.918)
Observations	360	360	360
Adjusted R-squared	0.101	0.311	0.367

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results of the two types of models make it clear that the impact of an independent variable to a measure of resilience largely depends on how resilience is defined. Our results show that the sign of a coefficient can flip depending on if the dependent variable is employment based or income based, such as in the case of the employment-population ratio in the VMR models. This result corroborates the findings of Deller et al. (2017). While this points to a need to be specific in what a researcher's measure of resilience is intended to represent, it also presents an opportunity to test what independent variables are significant across models. To do this we run a multivariate regression and then perform an F-test on each independent variable to test their significance to the entire set of models. The results are presented below.

Table 2-9: Multivariate regression for independent variable cross model significance

	(1)	(2)
VARIABLES	F statistic (7, 337)	Prob > F
Percentage of Population 18-35	2.26	0.0292
Percentage of Population 36-55	1.81	0.0840
Population Density	19.29	0.0000
Employment-Population Ratio	12.83	0.0000
Percentage of Households with Self-employment income	1.54	0.1537
Percentage of Population 65+	3.19	0.0028
Percentage of Population Black or African American	1.47	0.1756
Gini coefficient of income equality	3.03	0.0042
Herfindahl-Hirschman Index	5.92	0.0000
Percentage of Employment in Manufacturing	10.37	0.0000
Percentage of Employment in Finance, Insurance, Banking	1.27	0.2669
Percentage of Employment in Construction	7.41	0.0000
Percentage of Employment in Services Production	8.42	0.0000
Average Establishment births per 1000 population	10.61	0.0000

*** p<0.01, ** p<0.05, * p<0.1

The results in Table 2-9 show that only three variables are not significant across the set of models. These three variables are the percentage of households with self-employment income, the percentage of the population that is Black or African American, and the percentage of employment in finance, insurance, and banking. The remaining variables tested as significant across the set of models, with only the percentage of population 35 to 55 being borderline at a p-value of 0.084.

2.5 Discussion

The lack of correlation between the different dependent variables and the differences in the empirical results between the income-based models and the employment-based models illustrate the importance of being intentional when attaching the term economic resilience to a dependent variable. This is also true when considering how the dependent variable is constructed mathematically, for example the variance-mean ratio and Impact constructions utilized in this chapter.

When considering the VMR model results, a distinction needs to be made between an economy that is resilient and an economy that is thriving. Resilience in the Deller et al. (2017) context means a metro that was insulated from the effects of the recession. This means that they were protected from the negative impacts occurring in the national economy because their stability indicators did not vary substantially during the time of the recession. However, it also means that they were not substantively impacted by the significant nation-wide recovery that resulted from a large amount of government stimulus, because they did not experience a corresponding improvement in their stability indicators after the recession. For the purposes of the VMR models, resilient may be synonymous with insular.

Stability in the Income and Wage VMR models is likely not a good thing, it can also mean that income and wages are stagnant. Anecdotally, it could be that the most stable income and wage areas are ones where there are high poverty levels, high unemployment, and most individuals are living on a stable baseline income provided through government assistance programs. Areas that are unstable (high VMR) reflect areas with moving incomes and wages. This could be either good or bad. The VMR is a good first level tool to indicate if a variable is relatively constant over time. However, the VMR itself is not sufficient to indicate the direction of the variable over time. Figure 2-2 below shows three different series that have the same VMR of seven but represent three very different economic experiences.

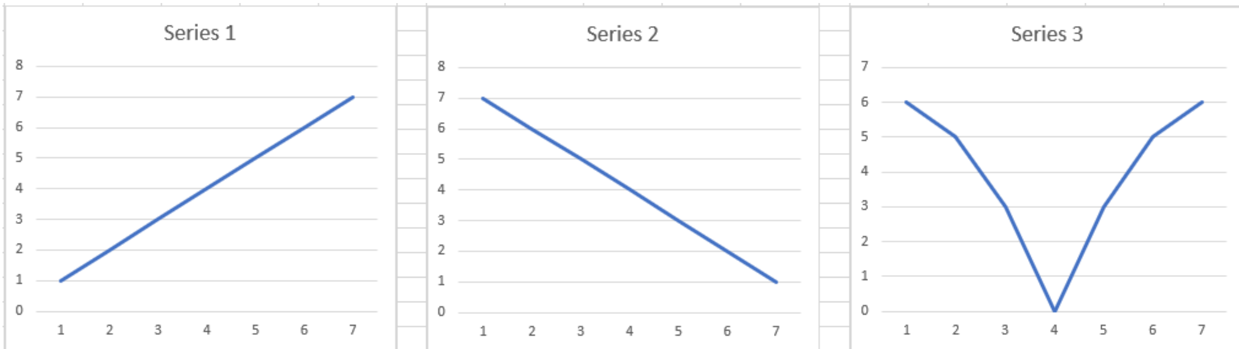

Figure 2-2: Three example VMRs with a value of seven

When interpreting the empirical results, the Kitsos & Bishop (2018) models provide intuitive coefficients. Positive coefficients mean the variable in question exacerbates the gap between the pre-recession conditions and the recessionary trough. On the other hand, negative coefficients are attached to independent variables who are impact-reducing at higher levels. By contrast, the coefficients on the VMR models indicate whether an independent variable is stability inducing if the coefficient is negative or more exacerbating if the coefficient is positive. However, as we have seen in the figure above, knowing that the VMR is higher does not provide an indication of the direction of the trend a region is experiencing. This makes it more difficult to determine the true impact an independent variable has from an economic viewpoint.

When discussing resilience, it must be acknowledged that there are two components: the ability to experience a minimal drop in the dependent variable during the impactful event, and the ability to quickly recover the ground that was lost and return to the existing growth path, or a new growth path. While the Impact measure from Kitsos & Bishop does quantify the differential from the conditions preceding the recession to the bottom of the trough and smooths the data by taking the average of the four years prior to the recession and the four most negative values during the recession, it is also only a one-dimensional view. It adequately addresses the first component of economic resilience but does not provide a window into the speed of recovery. To do so, one would need to consider a two-component measure of resilience.

Han & Goetz examine this in their 2015 work where they define resilience as the ratio of the decrease in employment to the rebound for a region. Rebound is quantified as the six-month measure of recovery. They expand on this in their subsequent 2019 paper where they also incorporate the concept of duration for both the decline and rebound into an idea they term an *impulse.* This multi-dimensional construction of resilience has an advantage in being able to showcase regions that experience shorter/smaller drops and longer/larger rebounds, though it would make an investigation into the determinants of resilience difficult because each region's dependent variable would cover a different length of time, based on their own unique resilience profile.

The need to examine resilience in multiple dimensions extends beyond the construction of a dependent variable that considers multiple aspects of the same variable, such as the discussion above referencing Han & Goetz (2015). Any work that does not consider multiple outcome variables in their investigation is neglecting the fact that any single proxy for resilience contains some weakness simply by the nature of constant change found in the component pieces of the economy. Using multiple dependent variables will present a more realistic picture of the impact recessions have on localities. For example, employment can hide the true impact of a negative shock if individuals switch from higher-income to lower-income jobs. This switching effect means using employment as the dependent variable would understate the true impact. This justifies the inclusion of dependent variables like wages and salary. Real GDP per capita is an output measure, but output may not be as impacted due to variations in recessionary impact on the markets where the output is distributed. Recessions also have a temporally uneven impact on retirements, as individuals who are downsized may decide to retire and remove themselves from the labor force. This reduces the denominator of the unemployment rate, distorting and overstating the unemployment impact.

This discussion also highlights the idea that it is preferable to examine the dynamics of several individual measures of resilience, as opposed to a composite index composed of several variables. This is driven mainly by the fact that regional economics does not yet have an agreed upon definition and measurement of economic resilience, since this area of the literature is still being explored and developed. Evenhuis (2020) argues that to get a good view into the factors that determine regional resilience multiple adaptation processes should be investigated either across time or space. The VMR model showed that determinants, such as the employment-population ratio, can have coefficients with opposing, but still significant, signs, depending on which dependent variable was being investigated. These results mirrored the empirical findings of Deller et al. (2017). This was presented in their discussion of the differences between employment-based measures of resilience and the wage and salary-based measures. The use of an index, made up of different measures like these, would only show the net impact of a determinant and would obscure the underlying dynamics. This would result in an incomplete understanding of the determinants of economic resilience and would limit the insight generated in empirical investigations.

2.6 Conclusions

Regional economic resilience has emerged as a newly popular complement to the regional growth path literature in discussions of what makes an area attractive to the people who live there, and to potential in-migrants. We started with a discussion of the evolution of the literature from the theoretical side, progressing from engineering resilience to ecological resilience to the latest concept of adaptive resilience. The last of these three makes the explicit recognition that the economy emerging from a recession is fundamentally different from the one that entered the recession because of reorganization occurring through economic forces, an argument similar to Schumpeter's creative destruction. Most of our empirical investigation focused on trying to identify variables that would proxy well for resilience, and to identify a set of independent variables that are determinants of economic resilience. Inspiration for the empirical section came from two sources, Deller et. al (2017) and Kitsos & Bishop (2018). We tested seven different dependent variables split across two models, to construct a set of dependent variables that maintained significance across regressions. We then conducted a discussion on the strengths and weaknesses of the empirical representations of resilience and noted that while the theoretical side of the work has matured substantially there is a gap on the empirical side of this research area. Resilience is a multi-faceted concept that cannot be represented with a single variable, but this work must be built up through testing many different stand-ins for resilience and understanding the different interactions that occur. Eventually, there may be an agreed upon index of resilience, however we stress that moving to the index approach too quickly would obscure the gross effects of the components while only presenting the net impact. A thorough understanding of the mechanics of resilience will aid in the recovery of future recessions, by both dampening the severity of the impact and accelerating the recovery that occurs afterward.

In this chapter we generated findings and insights that can provide guidance for local policy makers as they construct policies to balance growth with resilience. It is important to note that due to the short-term focus on growth and the long-term focus on resilience, these policies may need to be distributed intertemporally. Fortunately, it is likely that local governments will be able to identify policies that simultaneously impact resilience and growth in a positive manner. One example would be to encourage inflows in the segment of the population between 18 and 35. Additionally, the empirical findings suggest that resilience and economic growth could benefit by shifting employment away from manufacturing to more service-based sectors. Regions that can uncover and implement these lessons will benefit from living in a stable economic environment, resulting in fewer deviations from their long-term growth path and achieving closer to optimal levels of economic growth.

Chapter 3 - Examining how information generated by economic dynamism influences in-migration decisions

3.1 Introduction

The U.S. Census Bureau estimates that the average individual will move 11 times over the course of their life. In the 2013 Annual Social and Economic Supplement to the Current Population Survey, they note that of the 35.9 million individuals who moved between 2012 and 2013, approximately 20% indicated employment was the primary reason for the move (Ihrke, 2014). While this is an impressive number, the percentage of individuals moving has decreased substantially over the past three decades. A significant area of focus in the regional economics literature is investigating the forces driving internal migration. One branch of the literature explores if this decline is concerning, as high migration rates often accompany lower unemployment rates, or if it means that advances in technology have led to more efficient matching between individuals, employers, and locations with lower levels of friction. This decline in the propensity to migrate parallels a similar decline in economic dynamism from entrepreneurship – the opening and closing of business establishments - in the United States. Employment, job creation, and migration are shown to be linked, though the exact nature of this relationship remains murky from an empirical standpoint. Contributions such as Molloy, Smith, & Wozniak (2017) attempt to untangle the causal link between migration and employment, though much work remains.

One aspect that has been underexplored in linking migration to entrepreneurial dynamism is how this relationship varies based on the characteristics of the destination county. Prior work has explored both the differences in entrepreneurship by county size and the rural-urban migration decision. Bunten, Weiler, Thompson, & Zahran (2015) provides evidence that firm births and deaths, often described as economic dynamism, generate information that contributes to economic growth. This information on dynamism is used by others in deciding whether to pursue entrepreneurship. The question this paper works to answer is if the availability and quantity of information also influences the decision to migrate.

We couple the concept of economic dynamism with an extension of the theoretical model of Damm (2009) to motivate an empirical investigation into the role of information in the migration decision.

The generation and availability of information on dynamism helps to inform potential movers of the relative health of a given local economy and the path of future wages they might expect to receive. Because of this, the directional impact of these variables on migration rates will depend on when the potential migration decision occurs relative to an economic cycle. Migration is inherently riskier during an economic downturn and relatively easier during an economic boom. The empirical results of our investigation back up the claims found in our hypothesis. Spatial differences in the availability of economic information generated from the opening and closing of businesses, as proxied through the births, expansions, contractions, and deaths of local establishments and the non-employer share of total establishments, have a statistically (and economically) significant effect on inward migration rates. To test this hypothesis, we exploit a panel data set covering the years 1998-2014. We utilize a first difference estimator to test the significance of information variables at both the MSA and county levels of analysis. Our focus is on explaining the annual total in-migration rate at these two levels. This paper utilizes a higher-level view of migration at a national scale, while a previous chapter of this dissertation focuses on a more granular view of both migration and employment metrics, using Colorado as a case study.

The rest of the chapter is organized as follows. Section 3.2 provides some background on both the migration and dynamism literatures and frames the place of this paper within them. Section 3.3 lays out a theoretical model that incorporates the role of available information from dynamism as a psychic cost into the migration decision framework. Section 3.4 discusses the data that is utilized in this paper, while Section 3.5 describes the empirical model. Section 3.6 presents our results with Section 3.7 conducting a Blinder-Oaxaca decomposition between single-county and multi-county MSA migration rates. We then conduct some robustness checks in Section 3.8, while Section 3.9 concludes.

3.2 Migration and Economic Dynamism Literature

The domains of internal migration and economic dynamism have been linked in the regional economic literature for the span of several decades. One of the most noticeable trends over that time has been the decline in internal migration in the United States. This pattern has displayed itself as a consistent and continuous trend starting in the 1980s as illustrated in Figure 3-1.

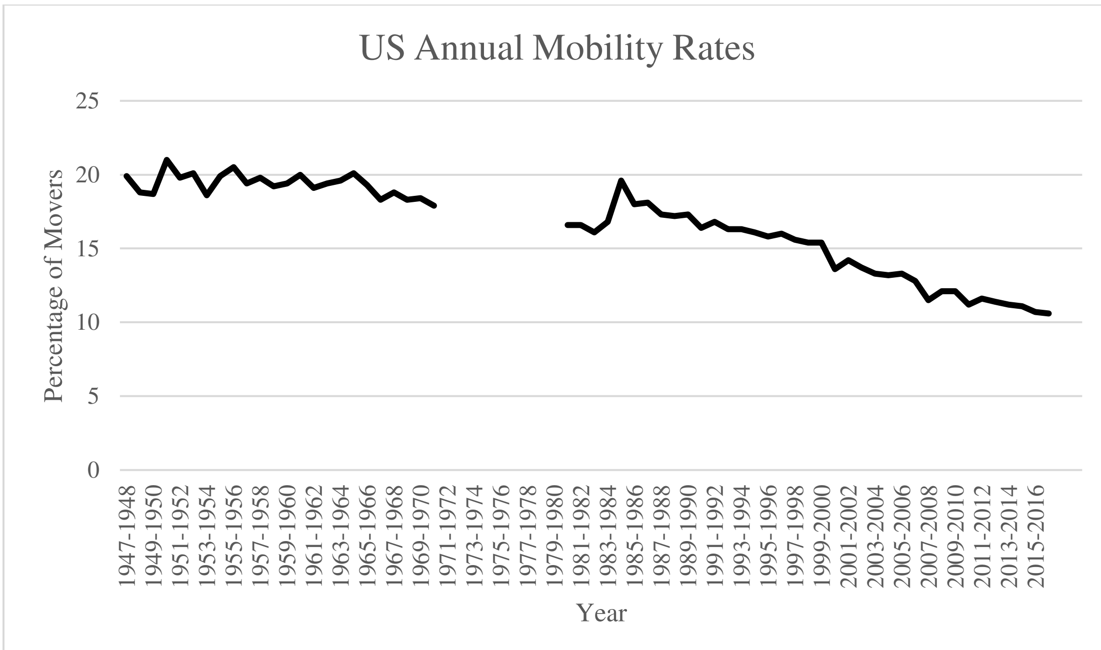

Figure 3-1: U.S. annual mobility rates (1947-2016)

A concerning feature of this downward trend has to do with migration's role as an adjustment mechanism for reducing regional economic shocks (Blanchard & Katz, 1992). Dao, Furceri, & Loungani (2014, 2017) show that in response to local demand shocks, state population adjustments, via net migration, are lower than in decades past. Also, the impact of a demand shock is likely to be uneven; harming low skill and marginal workers to a disproportionately high degree, as is shown in Bound & Holzer (2000). They go on to illustrate that declines in wages and employment are unevenly distributed across the U.S. While migration does occur to equilibrate the shocks, less-educated workers have relatively low rates of population adjustment in response to these demand changes. In a related work, Zabel (2012) discusses that the responses to labor demand shocks are intertwined with the state of the housing market. They find that homeowners are less likely to migrate due to higher costs and greater difficulty overcoming the challenges to move in response to changes in the labor market.

If labor markets become more rigid, due to declining migration, a historically important adjustment mechanism becomes less effective (Hyatt, McEntarfer, Ueda, & Zhang, 2018). The relatively high internal mobility in the U.S. has often been posited as an explanation for why the U.S. has lower average unemployment compared to European countries (Bentivogli & Pagano, 1999; Magrini, 2004). Other explanations for decreased migration elevate the roles of increased information availability and better technology in allowing individuals to migrate with less friction, greater certainty, and reduce the amount of reverse migration. Kaplan & Schulhofer-Wohl (2017) posits that improved information may make immigrants better able to choose a good initial destination. Additionally, they hypothesize that migration has fallen because job opportunities have become more similar across locations, making it easier to match an employee’s skills without needing to change locations. It is necessary for researchers to understand the reasons for this decline to determine if the decline in mobility is negatively impacting individuals and the economy as a whole. Once this determination is made, policy makers can then determine the optimal response.

The quantity and quality of available information on dynamism surrounding a potential migration destination can be thought of as a non-pecuniary cost in the migration cost function. This would classify information on dynamism as part of the family of psychic costs present in a migration decision. Sjaastad (1962) is one of the earliest papers to explicitly confront the idea of psychic costs and the impact that it has on migration. He argues that the presence of psychic costs, such as the reluctance to leave familiar surroundings and to move away from friends and family, reduces the consumer surplus of the potential migrant and reduces the overall level of migration to a sub-optimal level relative to a world without psychic costs. An early point of contention between Sjaastad (1962) and Maddox (1960) was whether psychic costs should be included in the study of the migration decision, with Maddox arguing that psychic costs were significant and could not be ignored, while Sjaastad argued that because they do not represent any real resource cost they should be ignored. Subsequent work, such as Deaton, Morgan, & Anschel (1982) catalogs and attaches dollar values to psychic costs for Appalachian migrants by comparing the differences needed to induce return migration to Eastern Kentucky between families who had moved to Lexington, Kentucky and Cincinnati, Ohio. Deaton et al. (1982) group psychic costs into three groups relating to the satisfaction a person has with their job, those aspects related to having friends and acquaintances in the local community, and their attitude towards the public and private services and facilities of the community where they live.

While the availability of information is shown to be an input to the migration decision, the generation of this information can result from the process of economic dynamism. Economic dynamism is a general name for a collection of different processes. Some examples of economic dynamism are the rate of job transitions by workers and the rates of firm births and deaths. Additionally, work by Davis, Faberman, & Haltiwanger (2012) examines the relationships between various measures of dynamism, such as worker and job flows over time. They find that while the relationships between variables vary based on aggregate economic conditions, the pace of worker flows and job churn has been declining in the U.S. economy over the past 30 years. All measures of dynamism involve the churn of resources through the economy and generate information as a result.

Information on dynamism is unique because it is often a very localized good (Weiler, 2006). The implication of this is that it can be difficult to observe information in regions with thinner markets. Lang & Nakamura (1993) discuss how markets with comparatively fewer home sales result in sub-optimal appraisals. An increase in available information helps to decrease the variance around the estimate of a home's true value. Similarly, a lack of economic dynamism leads to a lower availability of information, which can make a market appear less attractive compared to alternative markets with a higher availability of information. These geographical information asymmetries can result in a sub-optimal allocation of assets across the range of available markets, much in the spirit of Akerlof (1970).

In attempting to explain the dual decrease in mobility and economic dynamism, Molloy, Smith, & Wozniak (2017) test several demographic and socioeconomic factors. They examine trends like the aging of the population, the increase in dual income households, and even the advancement of technology that has increased the ability to engage in remote work or telecommuting. However, they are ultimately unable to link any of these explanations as the primary driver of the decline in mobility. Cooke (2011) provides additional evidence by showing that the age shift of the U.S. population can only explain a small portion of the reduction in migration. More recent work from Molloy et. al. (2019) points to the geographic concentration of metro areas and weak versus strong metro areas as potential explanations. Kaplan & Schulhofer-Wohl (2017) argue that a change in the occupational composition of many cities has reduced the need for individuals to move long distances to change occupations. Additionally, Hyatt, McEntarfer, Ueda, & Zhang (2018) examines the contribution of macroeconomic factors— particularly declining rates of job change—on migration rates.

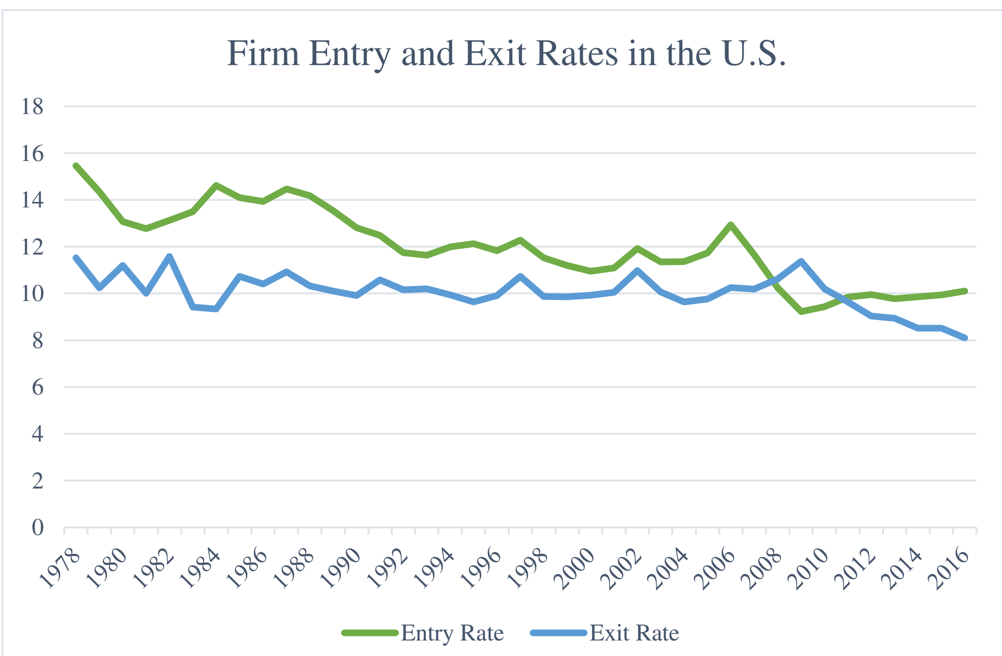

Figure 3-2: U.S. firm entry and exit rates (1978-2016)

Looking for explanations involving labor markets appears to be a promising avenue. A main reason is that the decline in migration has been accompanied by a three-decades long decline in economic dynamism, as measured by job-to-job transitions, job creation and destruction, and worker reallocation. Hyatt & Spletzer (2013) state that "between 1998 and 2010, hires and separations rates fell between 10 percent and 38 percent, depending upon the data source." Job creation and job destruction rates experienced similar declines, while job-to-job flows exhibited the largest decline, falling by 47 to 53 percent. Davis & Haltiwanger (2014) corroborate this decline in churn and job reallocation and find that it will be difficult to maintain high employment without restoring labor market fluidity. They expand on the "fluid labor market hypothesis" of Shimer (2001) to create a model that links labor mobility to job creation.

This chapter aims to help connect the parallel declines in migration and economic dynamism by arguing that the activities that make up economic dynamism generate observable information that is incorporated into the migration decision framework. Therefore, areas with higher levels of dynamism provide larger amounts of information which decreases the psychic costs associated with information availability in the migration decision and increases the amount of consumer surplus captured by the potential migrant. We illustrate this in the theoretical framework and then test if this information generating process is significant in the empirical results presented below.

3.3 Methodology

To motivate the empirical work, we discuss the theoretical contribution of Damm (2009) which constructs a model of out-migration in the presence of monetary and psychic costs and benefits. The model employed in that work is, itself, largely based on the well cited paper of Nakosteen & Zimmer (1980) which models the self-selection process of migrants based on anticipated returns. Additional vital work in the realm of self-selection migration models comes from Borjas (1987) and Borjas, Bronars, and Trejo (1992). In this formulation, U_{i1} is the expected utility of a person in location 1, the potential migration destination, U_{i0} is the expected utility of the person in the home location, and C_i is the expected cost to move, inclusive of monetary and psychic costs. We assume that moving costs do not vary by destination, but that they do differ across individuals.

An individual makes the decision to migrate if the net benefit of migration

$$M_i^* > 0 \tag{3.1}$$

and they decide to remain in the origin destination if

$$M_i^* \leq 0 \tag{3.2}$$

where

$$M_i^* = \alpha_0 + \alpha_1(U_{i1} - U_{i0}) - C_i - \varepsilon_i \tag{3.3}$$

In this formulation α_0 and α_1 are parameters to be estimated and ε_i is a stochastic error term. This formulation of the migration decision equation (3.3) allows for the propensity to migrate to increase and decrease linearly with expected gains in utility and increases in costs, respectively. Because we only witness migration or the lack thereof, it is understood that we cannot directly observe and recover the values of utility and costs, particularly as we are considering both the pecuniary and non-pecuniary benefits and costs of migration. In order to model the decision-making process in a way that is empirically tractable, assume the destination utility, origin utility, and cost functions can be modeled as follows

$$U_{i1} = \theta_{01} + X_i'\theta_{11} + Z_i'\theta_{21} + \varepsilon_{i1} \tag{3.4}$$

$$U_{i0} = \theta_{00} + X_i'\theta_{10} + Z_i'\theta_{20} + \varepsilon_{i0} \tag{3.5}$$

$$C_i = \gamma_0 + X_i'\gamma_1 + Z_i'\gamma_2 + \varepsilon_{ic} \tag{3.6}$$

In this system of equations, the X vector represents the personal attributes of individual i. In addition to the more easily quantifiable traits such as human capital and skill levels, this vector includes the profile of psychic benefits and costs, such as the preference for natural amenities, the preference for living close to family, and the level of information on dynamism needed to feel confident in the migration decision. All these tastes and preferences will vary across individuals. Z represents a vector of regional attributes at the destination location, which could include items such as demographics, labor market attributes, and housing market attributes, while θ and γ are the parameters to be estimated. Finally, the ε terms are stochastic errors. The structural form of the model is represented by equations (3.3)-(3.6). The reduced form of the migration decision equation is found by substituting equations (3.4)-(3.6) into equation (3.3). The reduced form is represented as follows

$$M_i^* = \beta_0 + X_i'\beta_1 + Z_i'\beta_2 - \varepsilon_i^* \qquad (3.7)$$

Presumably individuals select the location out of all possible locations that maximizes the value of M_i^*. However, we do not observe the value of M_i^*, only that the individual moves if the location that maximizes M_i^* is not the origination location and remains in their original location otherwise.

In an earlier version of this chapter, we constructed an income-based model inspired by the works of Lang & Nakamura (1993) and An & Becker (2013). This model incorporates an information generating process into the utility maximization process. While this model did not ultimately feel compatible with our classification of the information generated through economic dynamism as a potential psychic cost, the model does present a unique twist on Lang & Nakamura (1993) and is included in Appendix B.

3.4 Data Description

IRS Migration Data

Approximately two decades of migration data is available from a collaboration between the IRS Statistics of Income Division (SOI) and the U.S. Census Bureau. This data provides separate county to county inflow and outflow files for each year. Migration is identified through a change in address on filed returns across consecutive years. The methodology, strengths, and weaknesses of this data set are compiled in the Summer 2015 Statistics of Income Bulletin from the IRS, authored by Kevin Pierce. Strengths include a significant year to year match at the individual return level – lending confidence that the same population is being tracked over time. A significant weakness is that the available units of measure are the count of filed returns or the number of claimed exemptions, not an explicit count of individuals. Additionally, there is not 100% compliance in the filing of tax returns, and some individuals do not need to file a return based on their financial position. This could lead some demographics to be under-represented. This is a limited concern because the individuals who generally miss filing returns are often older – a population that tends to move less than average.

Firm Births, Expansions, Contractions, and Deaths

The occurrence of firm births, expansions, contractions, and deaths over time is tracked by the U.S. Census Bureau, in cooperation with the Office of Advocacy of the U.S. Small Business Administration (SBA), through their annual data series Statistics of U.S. Businesses (SUSB). This survey provides county level data on the number of firms that were created, expanded, contracted, or closed within the year of observation by U.S. businesses with paid employees. All calculated rates are expressed per thousand employees to account for population differences. The term "firm birth" refers to a firm hiring its first employee, thereby moving from being a "solopreneur" to being an employer. The Firm Churn variable is expressed as the product of the birth and death rates as opposed to the sum of the two rate. This is done following the example of Bunten et al. (2015) and signifies the role of the indirect positive information effect that accompanies the simple death rate. As a proxy for local economic dynamism, the product shows that a large amount of information can be generated, even in the presence of a low net birth rate. As opposed to an additive effect, constructing the churn variable as births*deaths creates a geographic effect of information. This effect is in keeping with the public good nature of information. Such public goods also have significant networking qualities, which makes a multiplicative, geometric relationship more appropriate, compared to an additive relationship. The overall effect is greater than the sum of its parts. When including the product in addition to the birth and death rates we can see the positive information effect generated through the, typically negative, impact of businesses closing. The *a priori* expectation would be for the positive information variables of birth rate and expansion rate to be accompanied by positive coefficients, resulting in higher migration by mitigating psychic costs, while the more negative information variables of death rate and contraction rate would reducing migration by showing evidence of a less robust economy. However, this may also be dependent on the relative position of the economy in the business cycle at a given time, potentially disallowing such a simplistic expectation.

Employment Portfolio Measures

To control for varying levels of employment opportunity and risk between MSAs, we utilize the Bureau of Economic Analysis employment data to construct rolling 10-year means and standard deviations for employment growth rates. Following the approach of Low & Weiler (2012) the 10-year mean growth rate proxies for the expected return to employment in the MSA, while the 10-year standard deviation provides a measure of the risk of seeking employment in a particular MSA. Similar measures of employment portfolio risk and return were examined deeply in Chapter 1. By including these variables, we hope to control for the relative opportunity of a migration destination, allowing the dynamism variables discussed above to proxy solely for the availability of information.

Employer share of establishments

Using the Non-Employer establishment data from the Census Bureau and combining it with the total establishments data from the SUSB series discussed previously allows us to calculate the share of total establishments that are employing institutions. As our theoretical model incorporates information on dynamism into the psychic cost vector, our assertion is that the employer share of establishments serves as a proxy for the availability of wage data.

Additional Independent Variables

In order to control for other pull factors of migration that a location can exhibit we include additional independent variables in our regression. These include the log of population, a housing price index, the growth rate of housing prices, a measure of employment specialization, the poverty rate, and median household income (expressed in $1,000s). Year level fixed effects are also included with the aim of minimizing omitted variable bias.

3.5 Empirical Framework

To test the impact of information on dynamism on in-migration rates we construct panel datasets at both the MSA and county levels. The raw data covers the years 1998-2014. However, we lose several years because we use a first-difference estimator and lagged independent variables. Based on testing, we include up to the 4th lag to balance maximizing the AIC with preserving our degrees of freedom. This reduces our timeframe to 2003-2014. The full empirical specification is as follows:

$$In - migration\ rate_t = \ln(population_{t-1}) + \sum_{n=1}^{4} firm\ birth\ rate_{t-n} + \sum_{n=1}^{4} firm\ expansion\ rate_{t-n} + \sum_{n=1}^{4} firm\ death\ rate_{t-n} + \sum_{n=1}^{4} firm\ contraction\ rate_{t-n} + \sum_{n=1}^{4} employer\ percentage\ of\ establishments_{t-n} + \sum_{n=1}^{4} firm\ churn_{t-n} + housing\ price\ index_{t-1} + housing\ price\ growth\ rate_{t-1} + employment\ specialization_{t-1} + poverty\ rate_{t-1} + median\ household\ income_{t-1} + 10\ year\ mean\ employment\ growth\ rate_{t-1} + 10\ year\ mean\ standard\ deviation\ of\ employment\ growth\ rates_{t-1} + year\ fixed\ effect_t$$

The dependent variable is the annual total in-migration rate for the unit of analysis. The independent variables of interest are the informational variables with their four lags. Control variables include the population, housing price index, housing price growth rate, poverty rate, median household income, and the employment portfolio measures of risk and return described in the data section. Because it is unlikely that contemporaneous realizations of the data are incorporated into a migration decision, we use the first lagged value of most variables and include additional lags of the information generating variables to see how, and if, older observations are utilized in the decision to migrate.

Table 3-1: Descriptive statistics for metro county level of analysis

Variable	Mean	Std.Dev.	Min	Max
Log of Population	11.905	1.139	8.662	16.125
Firm Birth rate	4.089	1.382	0	18.508
Firm Expansion rate	10.454	1.76	3.639	28.432
Firm Death rate	3.93	1.186	0	35.092
Firm Contraction rate	10.443	1.779	4.424	28.935
Firm Churn	17.158	10.73	0	161.19
Employer perc. of Establishments	24.486	4.806	9.903	48.61
Housing Price Index	365.765	201.412	91.67	1627.64
Housing Price Growth rate	1.48	6.505	-40.68	38.06
Employment Specialization	40.19	15.251	9.493	152.033
Poverty Rate	13.194	5.034	2.2	40.7
Median HH Income ($1,000s)	50.965	12.81	24.808	122.641
10-year mean employment growth rate	1.409	1.646	-7.353	13.297
10-year mean standard deviation of employment growth rates	2.409	1.495	.474	22.596

(n = 9656)

Additionally, here are the descriptive statistics for the MSA level analysis. A quick comparison shows that the descriptive statistics for the two sets are fairly similar, even though the number of metro counties outpaces the number of MSAs.

Table 3-2: Descriptive statistics for MSA level analysis

Variable	Mean	Std.Dev.	Min	Max
Log of Population	12.599	1.076	9.936	16.81
Firm Birth rate	4.1	2.321	1.547	51.271
Firm Expansion rate	10.941	5.576	4.657	130.162
Firm Death rate	3.92	2.076	1.578	48.084
Firm Contraction rate	10.781	5.554	4.493	143.701
Firm Churn	20.509	74.136	2.687	1987.593
Employer perc. of Establishments	27.236	3.89	13.304	37.849
Housing Price Index	367.43	185.915	106.68	1606.45
Housing Price Growth rate	2.401	6.656	-40.68	35.091
Employment Specialization	36.824	10.105	14.412	125.845
Poverty Rate	14.281	4.333	4.698	40.7
Median HH Income ($1,000s)	45.185	8.763	22.746	96.352
10-year mean employment growth rate	1.352	1.22	-2.682	9.485
10-year mean standard deviation of employment growth rates	1.827	.805	.463	7.258

(n = 5817)

3.6 Results

This section presents the results of the empirical strategy proposed above. Based on data availability we were able to test both metro designated counties and Metropolitan Statistical Areas. We present both sets of results as they do show variation with respect to which variables are significant.

Table 3-3: Metro county model results

In Migration Rate	Coef.	St.Err.	t-value	p-value	95% Conf	Interval	Sig
Log of Pop. Lag 1	-3.791	1.536	-2.47	0.014	-6.802	-0.781	**
Firm Birth Rate Lag 1	0.503	0.133	3.79	0.000	0.243	0.762	***
Firm Birth Rate Lag 2	0.104	0.076	1.37	0.171	-0.045	0.252	
Firm Birth Rate Lag 3	0.018	0.065	0.28	0.781	-0.109	0.144	
Firm Birth Rate Lag 4	0.007	0.044	0.17	0.867	-0.079	0.094	
Firm Expansion Rate Lag 1	0.012	0.018	0.68	0.499	-0.023	0.047	
Firm Expansion Rate Lag 2	0.024	0.015	1.62	0.106	-0.005	0.054	
Firm Expansion Rate Lag 3	0.007	0.015	0.50	0.615	-0.022	0.036	
Firm Expansion Rate Lag 4	0.032	0.013	2.44	0.015	0.006	0.057	**
Firm Death Rate Lag 1	0.678	0.182	3.72	0.000	0.320	1.036	***
Firm Death Rate Lag 2	0.271	0.118	2.30	0.022	0.040	0.503	**
Firm Death Rate Lag 3	0.130	0.078	1.67	0.096	-0.023	0.282	*
Firm Death Rate Lag 4	0.098	0.044	2.21	0.027	0.011	0.184	**
Firm Contraction Rate Lag 1	-0.048	0.024	-2.00	0.046	-0.094	-0.001	**
Firm Contraction Rate Lag 2	-0.075	0.029	-2.58	0.010	-0.132	-0.018	**
Firm Contraction Rate Lag 3	-0.043	0.027	-1.63	0.102	-0.096	0.009	
Firm Contraction Rate Lag 4	-0.003	0.017	-0.19	0.850	-0.037	0.030	

Employer Perc. of Establishments Lag 1	0.194	0.050	3.92	0.000	0.097	0.291	***
Employer Perc. of Establishments Lag 2	-0.013	0.019	-0.69	0.493	-0.050	0.024	
Employer Perc. of Establishments Lag 3	-0.050	0.014	-3.69	0.000	-0.076	-0.023	***
Employer Perc. of Establishments Lag 4	-0.030	0.013	-2.29	0.022	-0.056	-0.004	**
Firm Churn Lag 1	-0.111	0.027	-4.06	0.000	-0.164	-0.057	***
Firm Churn Lag 2	-0.037	0.016	-2.37	0.018	-0.067	-0.006	**
Firm Churn Lag 3	-0.019	0.011	-1.71	0.088	-0.041	0.003	*
Firm Churn Lag 4	-0.014	0.007	-1.96	0.050	-0.028	0.000	*
Housing Price Index Lag 1	-0.001	0.000	-5.71	0.000	-0.001	-0.001	***
Housing Price Growth Rate Lag 1	-0.005	0.002	-2.56	0.011	-0.008	-0.001	**
Specialization Lag 1	-0.002	0.002	-0.94	0.349	-0.007	0.002	
Poverty Rate Lag 1	-0.006	0.005	-1.18	0.238	-0.017	0.004	
Median Household Income ($1,000s) Lag 1	-0.008	0.003	-2.18	0.029	-0.014	-0.001	**
10-year Mean Employment Growth Rate Lag 1	-0.007	0.037	-0.19	0.848	-0.079	0.065	
10-year Mean Standard Deviation of Employment Growth Rate Lag 1	0.110	0.061	1.81	0.071	-0.009	0.229	*
Year Fixed Effects							
2004	0.263	0.072	3.65	0.000	0.122	0.404	***
2005	0.579	0.108	5.34	0.000	0.367	0.792	***
2006	0.357	0.136	2.62	0.009	0.090	0.624	***
2007	0.274	0.141	1.94	0.052	-0.003	0.551	*
2008	0.560	0.258	2.17	0.030	0.053	1.067	**
2009	0.209	0.262	0.80	0.424	-0.304	0.722	
2010	0.190	0.311	0.61	0.542	-0.420	0.799	
2011	0.824	0.354	2.33	0.020	0.129	1.518	**
2012	0.946	0.391	2.42	0.016	0.179	1.712	**
2013	0.603	0.399	1.51	0.131	-0.180	1.385	
2014	-0.701	0.409	-1.71	0.087	-1.503	0.101	*

Mean dependent var	-0.164	SD dependent var	0.838
R-squared	0.519	Number of obs	8781
F-test	150.888	Prob > F	0.000
Akaike crit. (AIC)	15810.779	Bayesian crit. (BIC)	16115.234

*** *p<0.01*, ** *p<0.05*, * *p<0.1*

The metro county results show a wide range of the variables representing information on dynamism expressing statistical significance. We see the full set of firm death rate and firm churn variables show as significant. Additionally, the firm contraction rate and the employer percentage of establishments show persistent significance through most of their lag structures. When we compare this to the MSA level regression below there is a clear divergence of which informational variables are significant. The MSA results show a larger role for the late birth rate and early expansion rate lags. Similarly, the early contraction rate lags and most of the death rate lags provide statistically significant values for variables representing information on dynamism, while the remaining dynamism information variables are less cohesive compared to the metro county results. We find no role for the firm churn variable and only the second lag of the employer percentage of establishments shows as significant. Overall, it appears that the information on dynamism variables present a less cohesive story when grouping at the MSA level compared to the metro county level. In Section 3-8, we will further examine how the grouping of counties into MSAs impacts the influence of these variables.

Table 3-4: MSA model results

In Migration Rate	Coef.	St.Err.	t-value	p-value	95% Conf	Interval	Sig
Log of Pop. Lag 1	0.004	0.017	0.21	0.830	-0.029	0.036	
Firm Birth Rate Lag 1	0.039	0.032	1.23	0.220	-0.023	0.101	
Firm Birth Rate Lag 2	0.035	0.028	1.26	0.206	-0.019	0.090	
Firm Birth Rate Lag 3	-0.067	0.026	-2.52	0.012	-0.118	-0.015	**
Firm Birth Rate Lag 4	-0.042	0.020	-2.07	0.039	-0.082	-0.002	**
Firm Expansion Rate Lag 1	0.009	0.015	0.58	0.558	-0.021	0.039	
Firm Expansion Rate Lag 2	0.037	0.017	2.13	0.033	0.003	0.071	**
Firm Expansion Rate Lag 3	-0.006	0.017	-0.34	0.733	-0.039	0.027	
Firm Expansion Rate Lag 4	-0.052	0.021	-2.46	0.014	-0.094	-0.011	**
Firm Death Rate Lag 1	0.068	0.038	1.78	0.074	-0.007	0.142	*
Firm Death Rate Lag 2	-0.001	0.028	-0.05	0.959	-0.057	0.054	
Firm Death Rate Lag 3	0.045	0.025	1.81	0.070	-0.004	0.093	*
Firm Death Rate Lag 4	0.049	0.024	2.05	0.040	0.002	0.097	**
Firm Contraction Rate Lag 1	-0.049	0.024	-2.01	0.044	-0.096	-0.001	**
Firm Contraction Rate Lag 2	-0.029	0.014	-2.13	0.033	-0.056	-0.002	**
Firm Contraction Rate Lag 3	-0.010	0.018	-0.55	0.586	-0.044	0.025	
Firm Contraction Rate Lag 4	0.030	0.020	1.48	0.138	-0.010	0.070	
Employer Perc. of Establishments Lag 1	0.005	0.006	0.88	0.378	-0.006	0.016	
Employer Perc. of Establishments Lag 2	0.008	0.004	2.09	0.037	0.000	0.015	**
Employer Perc. of Establishments Lag 3	-0.005	0.005	-1.09	0.275	-0.015	0.004	
Employer Perc. of Establishments Lag 4	0.004	0.004	0.94	0.350	-0.004	0.012	
Firm Churn Lag 1	-0.001	0.003	-0.22	0.825	-0.007	0.005	
Firm Churn Lag 2	-0.002	0.003	-0.92	0.360	-0.008	0.003	
Firm Churn Lag 3	0.003	0.003	1.10	0.272	-0.003	0.009	
Firm Churn Lag 4	0.003	0.002	1.33	0.183	-0.001	0.008	
Housing Price Index Lag 1	0.000	0.000	-4.02	0.000	-0.001	0.000	***
Housing Price Growth Rate Lag 1	-0.006	0.002	-2.36	0.018	-0.010	-0.001	**
Specialization Lag 1	-0.003	0.002	-1.95	0.051	-0.007	0.000	*
Poverty Rate Lag 1	0.013	0.004	3.06	0.002	0.005	0.021	***
Median Household Income ($1,000s) Lag 1	0.007	0.003	2.44	0.015	0.001	0.012	**
10-year Mean Employment Growth Rate Lag 1	-0.072	0.021	-3.40	0.001	-0.113	-0.030	***
10-year Mean Standard Deviation of Employment Growth Rate Lag 1	-0.018	0.030	-0.58	0.562	-0.077	0.042	
Year Fixed Effects							
2004	-0.051	0.041	-1.24	0.216	-0.132	0.030	
2005	0.100	0.050	2.01	0.044	0.003	0.197	**
2006	-0.095	0.050	-1.92	0.055	-0.192	0.002	*
2007	-0.105	0.059	-1.77	0.077	-0.222	0.012	*
2008	-0.292	0.073	-3.99	0.000	-0.435	-0.148	***

2009	-0.484	0.077	-6.31	0.000	-0.634	-0.334	***
2010	-0.590	0.095	-6.22	0.000	-0.776	-0.404	***
2011	-0.197	0.111	-1.77	0.077	-0.415	0.021	*
2012	-0.336	0.127	-2.64	0.008	-0.586	-0.087	***
2013	-0.615	0.107	-5.72	0.000	-0.826	-0.405	***
2014	-1.726	0.093	-18.54	0.000	-1.909	-1.544	***

Mean dependent var	-0.118	SD dependent var	0.612
R-squared	0.500	Number of obs	5157
F-test	79.526	Prob > F	0.000
Akaike crit. (AIC)	6267.458	Bayesian crit. (BIC)	6549.027

*** *p<0.01*, ** *p<0.05*, * *p<0.1*

From the results above we can see that the variables proxying for information on dynamism do enter the regression in a statistically significant fashion. All the informational variables show significance in some combination of the results presented. While individual variables show significance it is also important to see if the information on dynamism variables maintain significance when considered as a complete set. To assess this, we conduct F-tests on each set of dynamism variable lags and for all information on dynamism variables together. The results for both the Metro County and MSA levels are presented in Table 3-5 below.

Table 3-5: F-tests for joint significance

	Metro		**MSA**	
Variable Set	F Stat	p value	F Stat	p value
Birth Rate	8.68	0.0000	3.97	0.0032
Expansion Rate	2.12	0.0756	3.02	0.017
Death Rate	6.49	0.0000	1.57	0.1806
Contraction Rate	2.58	0.0352	2.7	0.0292
Firm Churn Rate	5.34	0.0003	0.86	0.4893
Employer Percentage of Establishments	5.44	0.0002	2.94	0.0194
All Information on Dynamism Variables	4.9	0.0000	5.43	0.0000

For the Metro County results all the information on dynamism variable sets are significant at the 10 percent level, with all but the expansion rate being significant at the five percent level. For the MSA level there is similar reassurance that the information variable sets are jointly significant. However, the firm death rate and firm churn sets are not jointly significant. It is also noteworthy that in the regression results the coefficients show a decay as we go deeper into the lag structure. It would appear that all information regarding dynamism is valuable, but more recent information is clearly more impactful than older information.

3.7 Blinder-Oaxaca Decomposition

Large geographic differences exist between the county level groupings and MSAs, with MSAs being composed of anywhere from a single county up to the case of the Atlanta-Sandy Springs-Roswell, GA MSA which contains 29 counties. Looking at Figure 3-2 below we can compare the average in-migration rate of the single county MSAs to the rate of the multi-county MSAs and find that the multi-county MSA average is consistently higher than the single-county average for the entire duration of the investigation window.

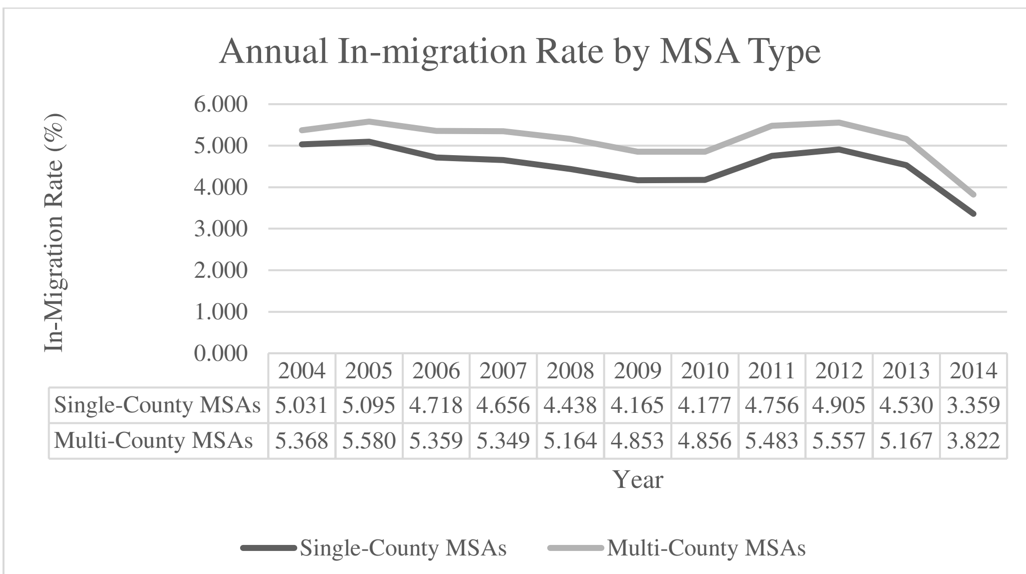

Figure 3-3: Average annual in-migration rate by MSA type

With such a persistent gap between the two groups it is worth checking for statistical differences to see if these groupings are an avenue that should be explored further in our endeavor to understand in-migration rates. To test more formally for a difference between the two groups we conduct a t-test with the assumption of unequal variances. This test, the results of which can be found below in Table 3-6, shows a statistically significant difference in the mean annual in-migration rate between the two groups.

Table 3-6: T-test for difference in means between MSA type in-migration rates

t-Test: Two-Sample Assuming Unequal Variances	*Multi-County MSAs*	*Single-County MSAs*
Mean	5.141759164	4.53016531
Variance	4.455156503	10.45074573
Observations	2519	1650
Hypothesized Mean Difference	0	
df	2567	
t Stat	6.794478081	
P(T<=t) one-tail	6.73178E-12	
t Critical one-tail	1.645447442	
P(T<=t) two-tail	1.34636E-11	
t Critical two-tail	1.960888555	

We can examine if the included independent variables are key to explaining the differences in the dependent variable between the two groups. This can be accomplished by running a decomposition between the groups. One way to conduct this decomposition is by running a Blinder-Oaxaca decomposition on the data, using a binary variable to indicate if an MSA is single- or multi-county in composition. Historically, the Blinder-Oaxaca decomposition has been used in labor economics to examine cases of bias in labor market outcomes or wage differentials (Blinder, 1973; Oaxaca, 1973). The technique was developed and published simultaneously by Oaxaca and Blinder in 1973 in the context of showing discrimination in wage differentials. Both works are based on the 1955 publication of Kitagawa, where she presents a method of summarizing and comparing the differences in two or more sets of specific rates. The decomposition is written as follows:

$$\Delta\bar{Y} = \bar{Y}_A - \bar{Y}_B$$

This first expression shows the difference in means between group A and group B. In this equation we can substitute the average characteristics for each group (the $\overline{X'}$s) and the vectors of returns to those characteristics (the $\hat{\beta}$s) in for the $\bar{Y}_A$ and $\bar{Y}_B$ terms as seen below:

$$\Delta\bar{Y} = \overline{X'}_A\hat{\beta}_A - \overline{X'}_B\hat{\beta}_B$$

The two terms on the right-hand side can be factored to express the three-fold Blinder-Oaxaca decomposition, where the first term represents the endowments, the second term represents the coefficients, and the final term represents the interactions.

$$\Delta\bar{Y} = (\bar{X}_A - \bar{X}_B)\hat{\beta}_B + \overline{X'}_B(\hat{\beta}_A - \hat{\beta}_B) + (\bar{X}_A - \bar{X}_B)'(\hat{\beta}_A - \hat{\beta}_B)$$

Here the decomposition is written with respect to group B, which in the case of this paper is the single-county MSA group. This means we take group B's mean in-migration rate and try to uncover what it would take to converge this group's outcomes to that of the multi-county group. The endowments term of the equation represents the contribution of differences in the independent variables across the two groups. The group differences in the coefficients are expressed by the second term, the coefficients term. Lastly, the interaction term expresses the portion of the difference between the groups that results from the simultaneous cross-group differences in endowments and coefficients (Hlavac, 2022). In the context of labor economics, the two-fold decomposition is often employed. This version breaks down differences between two groups to see what can be explained by the independent variables included in our model and what portion of the difference is attributed to omitted variables, often termed in the labor literature to be bias, discrimination, or omitted variables. The decomposition is really a partial equilibrium approach where each term works to answer what the difference between the two groups would have been had either the endowments or coefficients been the same between groups while the other portion of the equation was allowed to vary.

The goal of the test is to break down the 0.5487% difference in the average annual in-migration rate between single-county MSAs and multi-county MSAs. Understanding the contributing factors will help us to know what levers metro areas might employ to increase in-migration rates in single-county MSAs. Table 3-9 below shows that there are significant contributions to the difference in means in all three terms of the decomposition. The total of the three coefficients sum to the difference in means, but show different directional impacts, with the endowments effect working to close the difference, while the coefficient and interaction effects contribute positively to the difference in means. The bootstrapped standard errors included in the table are the result of 1,000 random samples with replacement from the observation set. Each of the 1,000 random samples is run through the decomposition equation. The bootstrapped standard error is then calculated as the standard deviation of these 1,000 decomposition estimates (Hlavac, 2022). In Table 3-7 both the coefficients and bootstrapped standard errors are presented. For each of the three components of the decomposition the coefficient is a minimum of 3.5 times the size of the bootstrapped standard error, giving a high degree of confidence that all three coefficients are statistically different from zero.

Table 3-7: Blinder-Oaxaca Test Results

	Endowments	Coefficients	Interaction
Coefficient	-0.65673	0.92747	0.27797
Bootstrap St. Err	0.07240	0.07796	0.07590

The significance in differences between the two groups can also be seen visually in the following charts showing the endowments, coefficients, and the interaction results. Here we can see the contributions of the individual variables to each overall effect. In each graph the individual variables are plotted with their coefficient and an error bar representing a 95% confidence interval. When conducting a Blinder-Oaxaca decomposition, the individual variables can have substantially differing scales across the three components. This is easy to see when comparing the scale of the endowment effect in Figure 3-3 to that of the coefficient effect in Figure 3-4. The individual endowment effects range from approximately -0.6 to 0.6, while the coefficient effects range from approximately -10 to 5.

First, we investigate the endowment effect. The endowment effect measures the part of the gap in the outcomes related to differences in the explanatory variables. This shows the mean change in the in-migration rate that would occur if the single-county group of MSAs had the multi-county group's independent variables, while maintaining their current coefficients (Etezady, Shaw, Mokhtarian, & Circella, 2021). This swap would reduce the gap between the two groups by 0.6567 percentage points. Figure 3-3 shows that the negative endowment effect is driven largely by the $\ln(population)$. Of note is that the same variable has a substantial offsetting positive effect in the coefficients chart where it is joined by the first lag of the contraction rate in contributing to the positive impact of the coefficients term. Other variables showing a statistically significant positive effect are the third lag of the birth rate and the first lag of the housing price index. In addition to the substantial impact of population, the employment portfolio return variable contributes to the negative side of the endowments effect.

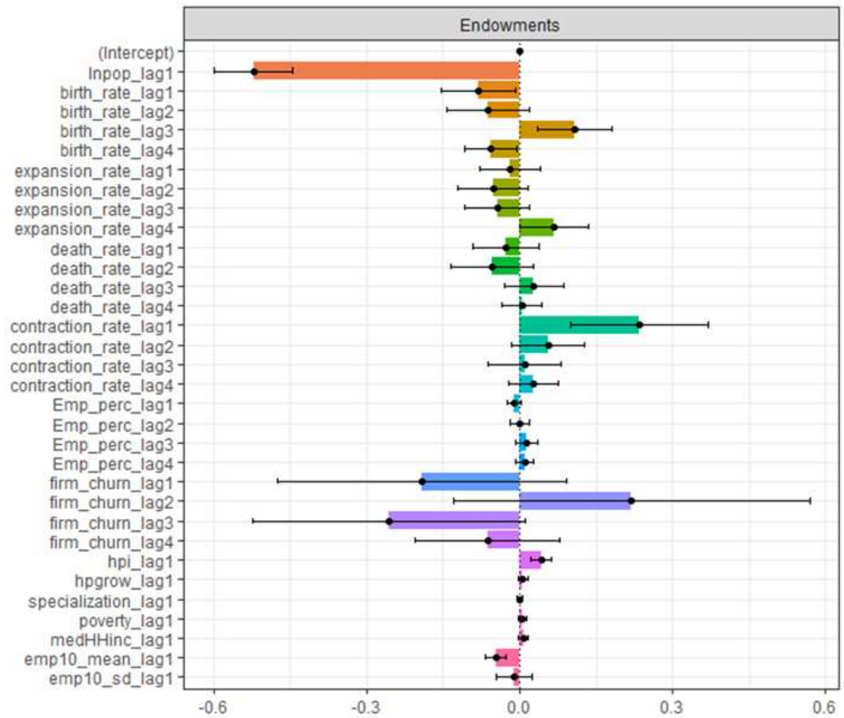

Figure 3-4: Endowment results graph

Next, we move on to the coefficient effect. To recall, the coefficient effect shows the portion of the difference in mean in-migration rates that is attributed to differences in the group coefficients, weighted by the single-county group's vector of mean explanatory variables. It can also be thought of as the rate of return to the different independent variables. This is the change in the gap that would occur if the single-county group had the coefficients of the multi-county group, but held its independent variable levels constant (Etezady, Shaw, Mokhtarian, & Circella, 2021). Doing so would increase the gap between the two groups by 0.927 percentage points. The results show that the gap between the two groups is mitigated by the intercept, the second and third lags of the expansion rate, and the first lag of the housing price index. On the other hand, the population variable, the first, third, and fourth lags of the contraction rate, the specialization index, the median household income, and the employment portfolio return measure all contribute to increasing the gap in in-migration rates between the two, although the impact of the last three variables mentioned does not appear to be very large. To summarize this component, the single-county MSA group would experience worse outcomes with respect to in-migration rates if it had the same returns to its independent variables that the multi-county MSA group experiences.

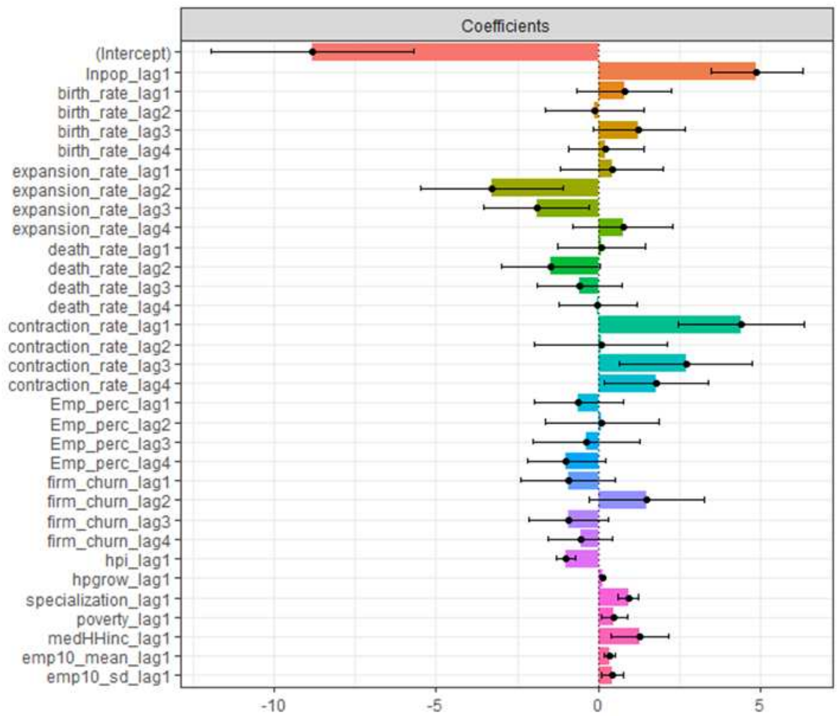

Figure 3-5: Coefficients results graph

Finally, moving on to the interaction effect, which examines the portion of the total gap that exists as a result of the interaction of differences in the endowments and coefficients of the two groups. This can be interpreted as either the portion of the gap that results from simultaneously changing the endowments and the coefficients, or the portion of the gap that remains after controlling for the endowment and coefficient effects (Etezady, Shaw, Mokhtarian, & Circella, 2021). In this model the interaction effect comes out to approximately 0.278 percentage points. The gap between the two groups is increased by the presence of the population variable, as well as the second lag of the expansion rate. Countervailing impacts are produced by the first lag of the contraction rate and the two variables measuring the risk and return to the employment portfolio. These effects mostly offset to produce a small overall interaction impact of around 0.278 percent.

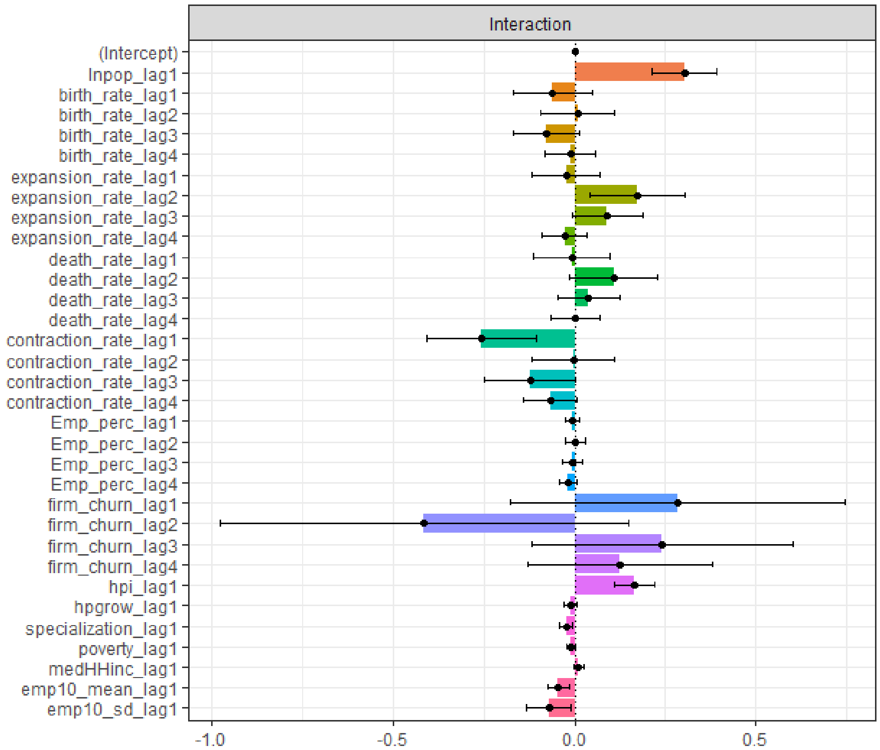

Figure 3-6: Interaction results graph

Overall, when the difference in means between the two groups of MSAs is decomposed into the component pieces the main drivers of the difference appear to be the difference in population and the first lag of the contraction rate. While it is informative to understand what is driving the difference, in this instance it does not provide much in the way of policy levers for those communities looking to increase the rate of in-migration.

3.8 Single-County and Multi-County MSA Comparison

Looking back to the initial empirical results, the difference in physical area covered likely explains most of the divergence between the MSA and county-level analyses, along with other differences between the included independent variables. However, in the interest of being thorough, we explore this divergence by running the regression independently for MSAs that coincide with a single county compared to MSAs composed of multiple counties. When looking at the data, 150 MSAs coincide with a single county, while 229 are comprised of more than one county. The results for the split MSA regressions can be found below, with the single-county results being presented first.

The results of the single-county MSA group are very encouraging with respect to the role of variables representing information on dynamism in the migration decision. There is strong evidence that multiple lags of the birth rate, the firm expansion rate, and the rate of firm churn all contribute to determining the in-migration rate for the selected MSAs. Additionally, strong effects are seen on the lagged log of population, both the housing price growth rate and housing price index, and the poverty rate measure. Interestingly, the median household income and the measure of return for the employment portfolio cease to be significant, in contrast to the full MSA sample results from Table 3-4.

In examining the coefficients, it is important to remember that the empirical approach utilizes a first-difference estimator on a series that is trending negative over time. That means the resulting dependent variable of $\Delta y_{t,t-1} = y_t - y_{t-1}$ is negative for the majority of observations. Likewise, the differenced independent variables may be negative or positive, varying by MSA. When paired with the resulting regression coefficients the impact on the independent variable could be positive or negative on a case-by-case basis. This causes difficulty in formulating broad stroke conclusions to the question of how the information on dynamism variables impact the migration rate. It feels more appropriate to simply stress the fact that they do show as significant within the regression and relegate quantifying the positive or negative impacts to future work, which could include individual MSA case studies.

Table 3-8: Single-county MSA results

In Migration Rate	Coef.	St.Err.	t-value	p-value	95% Conf	Interval	Sig
Log of Pop. Lag 1	-5.966	1.469	-4.06	0.000	-8.848	-3.084	***
Firm Birth Rate Lag 1	0.050	0.058	0.85	0.394	-0.065	0.164	
Firm Birth Rate Lag 2	0.115	0.060	1.92	0.055	-0.003	0.233	*
Firm Birth Rate Lag 3	-0.184	0.056	-3.26	0.001	-0.295	-0.073	***
Firm Birth Rate Lag 4	-0.148	0.040	-3.72	0.000	-0.226	-0.070	***
Firm Expansion Rate Lag 1	0.071	0.037	1.93	0.054	-0.001	0.143	*

Firm Expansion Rate Lag 2	0.139	0.037	3.79	0.000	0.067	0.211	***
Firm Expansion Rate Lag 3	0.056	0.037	1.53	0.126	-0.016	0.128	
Firm Expansion Rate Lag 4	-0.068	0.036	-1.89	0.059	-0.139	0.003	*
Firm Death Rate Lag 1	0.060	0.043	1.40	0.163	-0.024	0.145	
Firm Death Rate Lag 2	0.068	0.049	1.41	0.160	-0.027	0.163	
Firm Death Rate Lag 3	-0.016	0.049	-0.32	0.752	-0.112	0.081	
Firm Death Rate Lag 4	-0.011	0.036	-0.31	0.757	-0.081	0.059	
Firm Contraction Rate Lag 1	-0.069	0.043	-1.62	0.105	-0.153	0.015	
Firm Contraction Rate Lag 2	0.037	0.030	1.25	0.210	-0.021	0.095	
Firm Contraction Rate Lag 3	-0.014	0.039	-0.37	0.713	-0.090	0.062	
Firm Contraction Rate Lag 4	0.049	0.034	1.42	0.156	-0.019	0.116	
Employer Perc. of Establishments Lag 1	-0.068	0.033	-2.08	0.037	-0.133	-0.004	**
Employer Perc. of Establishments Lag 2	0.027	0.045	0.61	0.543	-0.061	0.116	
Employer Perc. of Establishments Lag 3	0.042	0.033	1.30	0.195	-0.022	0.106	
Employer Perc. of Establishments Lag 4	-0.046	0.028	-1.66	0.098	-0.100	0.008	*
Firm Churn Lag 1	0.006	0.005	1.33	0.183	-0.003	0.015	
Firm Churn Lag 2	-0.013	0.004	-3.16	0.002	-0.021	-0.005	***
Firm Churn Lag 3	0.014	0.004	3.28	0.001	0.006	0.022	***
Firm Churn Lag 4	0.010	0.002	4.53	0.000	0.006	0.015	***
Housing Price Index Lag 1	-0.001	0.000	-2.42	0.016	-0.001	0.000	**
Housing Price Growth Rate Lag 1	-0.006	0.003	-2.30	0.022	-0.011	-0.001	**
Specialization Lag 1	-0.829	0.503	-1.65	0.100	-1.816	0.158	
Poverty Rate Lag 1	0.018	0.007	2.66	0.008	0.005	0.031	***
Median Household Income ($1,000s) Lag 1	0.000	0.000	0.05	0.961	0.000	0.000	
10-year Mean Employment Growth Rate Lag 1	8.940	6.726	1.33	0.184	-4.252	22.133	
10-year Mean Standard Deviation of Employment Growth Rate Lag 1	2.740	6.728	0.41	0.684	-10.456	15.936	
Year Fixed Effects							
2004	0.019	0.077	0.25	0.804	-0.132	0.171	
2005	0.214	0.104	2.06	0.039	0.010	0.418	**
2006	0.001	0.133	0.01	0.993	-0.259	0.262	
2007	0.186	0.164	1.13	0.258	-0.136	0.507	
2008	-0.086	0.223	-0.39	0.699	-0.523	0.351	
2009	0.212	0.263	0.81	0.421	-0.304	0.728	
2010	0.196	0.287	0.68	0.496	-0.368	0.759	
2011	0.361	0.319	1.13	0.258	-0.265	0.986	
2012	0.081	0.350	0.23	0.818	-0.605	0.766	
2013	-0.159	0.368	-0.43	0.665	-0.882	0.563	
2014	-0.989	0.375	-2.64	0.008	-1.725	-0.253	***

Mean dependent var	-0.149	SD dependent var	0.667
R-squared	0.603	Number of obs	1635
F-test	38.004	Prob > F	0.000
Akaike crit. (AIC)	1967.945	Bayesian crit. (BIC)	2200.119

*** *p<0.01*, ** *p<0.05*, * *p<0.1*

In contrast to the single-county MSA results above, the multi-county MSA sample shows a substantially diminished role for the information on dynamism variables in explaining the changes in in-migration rates. Table 3-9 shows a smaller number of the dynamism information variables showing as statistically significant, and a third fewer number of significant variables. Additionally, the levels of significance in the sample are lower with only two variables showing as significant at the $p<0.01$ level. These results lend credence to the idea that even though the multi-county MSAs outnumber the single-county MSAs by almost 80 locations, the results in Section 3.6 were driven, to a large extent, by this smaller group. A possible reason for this could be that going to large metro areas is often more of a reflex response, while selecting a smaller MSA requires more information. Multi-county MSAs have highly developed brands, as well as heterogeneity across the MSA. Comparatively, a smaller, single-county MSA has more of a singular cohesive identity, making it more vital that an individual identifies with the MSA's core personality. This requires the potential migrant to uncover and understand more substantive information about these small MSAs.

Table 3-9: Multi-county MSA results

In Migration Rate	Coef.	St.Err.	t-value	p-value	95% Conf	Interval	Sig
Log of Pop. Lag 1	-1.664	1.860	-0.90	0.371	-5.311	1.984	
Firm Birth Rate Lag 1	0.043	0.046	0.93	0.354	-0.048	0.133	
Firm Birth Rate Lag 2	-0.123	0.066	-1.88	0.060	-0.252	0.005	*
Firm Birth Rate Lag 3	-0.088	0.060	-1.48	0.139	-0.205	0.029	
Firm Birth Rate Lag 4	-0.020	0.035	-0.58	0.559	-0.089	0.048	
Firm Expansion Rate Lag 1	0.114	0.036	3.18	0.001	0.044	0.184	***
Firm Expansion Rate Lag 2	-0.005	0.032	-0.16	0.872	-0.068	0.057	
Firm Expansion Rate Lag 3	-0.005	0.032	-0.16	0.873	-0.069	0.059	
Firm Expansion Rate Lag 4	0.022	0.026	0.85	0.395	-0.029	0.073	
Firm Death Rate Lag 1	0.083	0.069	1.22	0.224	-0.051	0.218	
Firm Death Rate Lag 2	-0.017	0.057	-0.30	0.766	-0.128	0.094	
Firm Death Rate Lag 3	0.056	0.043	1.31	0.189	-0.028	0.140	
Firm Death Rate Lag 4	0.080	0.041	1.95	0.051	0.000	0.161	*
Firm Contraction Rate Lag 1	0.091	0.044	2.07	0.038	0.005	0.178	**
Firm Contraction Rate Lag 2	-0.009	0.036	-0.25	0.806	-0.080	0.062	
Firm Contraction Rate Lag 3	0.034	0.039	0.88	0.379	-0.042	0.110	
Firm Contraction Rate Lag 4	0.104	0.041	2.54	0.011	0.024	0.185	**
Employer Perc. of Establishments Lag 1	0.106	0.041	2.59	0.010	0.026	0.187	**
Employer Perc. of Establishments Lag 2	-0.013	0.020	-0.67	0.500	-0.051	0.025	
Employer Perc. of Establishments Lag 3	-0.007	0.023	-0.31	0.757	-0.052	0.038	
Employer Perc. of Establishments Lag 4	-0.032	0.017	-1.90	0.058	-0.065	0.001	*

Firm Churn Lag 1	-0.006	0.006	-1.04	0.296	-0.018	0.005	
Firm Churn Lag 2	0.006	0.007	0.89	0.372	-0.007	0.019	
Firm Churn Lag 3	-0.001	0.004	-0.30	0.768	-0.009	0.006	
Firm Churn Lag 4	-0.010	0.005	-2.09	0.037	-0.020	-0.001	**
Housing Price Index Lag 1	-0.002	0.000	-4.17	0.000	-0.002	-0.001	***
Housing Price Growth Rate Lag 1	0.001	0.004	0.22	0.828	-0.007	0.008	
Specialization Lag 1	-1.527	0.840	-1.82	0.069	-3.173	0.120	*
Poverty Rate Lag 1	0.008	0.011	0.77	0.439	-0.013	0.029	
Median Household Income ($1,000s) Lag 1	0.000	0.000	-0.15	0.877	0.000	0.000	
10-year Mean Employment Growth Rate Lag 1	7.060	9.051	0.78	0.435	-10.689	24.809	
10-year Mean Standard Deviation of Employment Growth Rate Lag 1	3.680	10.242	0.36	0.719	-16.405	23.764	
Year Fixed Effects							
2004	0.124	0.067	1.84	0.066	-0.008	0.255	*
2005	0.405	0.105	3.86	0.000	0.199	0.612	***
2006	0.363	0.131	2.78	0.006	0.107	0.620	***
2007	0.367	0.152	2.42	0.016	0.069	0.664	**
2008	0.596	0.242	2.47	0.014	0.122	1.069	**
2009	0.259	0.265	0.98	0.329	-0.261	0.778	
2010	0.143	0.293	0.49	0.626	-0.432	0.718	
2011	0.561	0.331	1.70	0.090	-0.088	1.210	*
2012	0.588	0.348	1.69	0.091	-0.095	1.271	*
2013	0.416	0.348	1.20	0.231	-0.266	1.098	
2014	-0.754	0.357	-2.11	0.035	-1.455	-0.054	**

Mean dependent var	-0.138	SD dependent var	0.659
R-squared	0.563	Number of obs	2477
F-test	61.207	Prob > F	0.000
Akaike crit. (AIC)	3104.255	Bayesian crit. (BIC)	3354.291

*** *p<0.01*, ** *p<0.05*, * *p<0.1*

3.9 Conclusions

This chapter began with a discussion of the dual, persistent, multi-decade declines in economic dynamism and the propensity to migrate. The aim of the empirical investigation was to explore the link between these two declines and how their relationship varies based on the economic characteristics of the migration destination. The theoretical work discussed the role of information as an influence on the decision to migrate internally within the U.S. by appearing as a psychic cost in the utility function. We discussed how economic dynamism generates information useful to a local economy through the birth, expansion, contraction, and death of establishments. This concept embedded the available quantity of information on dynamism as a component of the psychic costs potential migrants face when considering a location change. The impact of the available information on dynamism on migration was then tested empirically.

We have seen that information on dynamism does show up as statistically significant within a first differenced panel-data regression, at both the metro county and MSA levels. However, there was a clear tension between the two levels of aggregation when it came to the role of the informational variables in the regression results. The more disaggregated metro county level results appeared to show a stronger role for the information on dynamism variables in influencing migration decisions. While the results of the F-tests showed substantial significance for the full sets of informational variables in each model, we used section 3.8 to dig further into the MSA level results and uncover some additional insights.

Plotting the average in-migration rates of the single-county MSAs and the multi-county MSAs showed a persistent gap between the two groups, with the multi-county MSAs experiencing a persistently higher in-migration rate. To search for the contributing factors to explain this gap we decomposed the difference using a three-fold Blinder-Oaxaca decomposition. We found evidence of significant contributions from all three parts of the decomposition: the endowment effect, the coefficient effect, and the interaction effect. A large portion of the gap was driven by differences in the size and impact of population differences and difference in the first lag of the contraction rate. While helpful in explaining the gap between the two groups, these variables provide limited policy prescriptions for places looking to increase their in-migration rate.

While the work done in this chapter can help from an explanatory perspective, it should not be seen to be predictive or prescriptive in nature. One interesting insight that was raised in the robustness checks section was the apparent role of the smaller MSAs in driving the MSA level results. The single-county MSAs showed a much larger role for the set of information on dynamism variables as potential drivers of in-migration compared to the larger, multi-county MSAs. One potential reason for this is that in electing to move to a single-county MSA, a potential migrant must do additional research to ensure congruence with the smaller metro's more singular identity. This is in comparison to the larger metro areas which, by nature of their size, offer a more heterogeneous menu of potential destinations within the confines of a single designated MSA. This results in a greater need for information by the individual wanting to move to a smaller MSA.

Over time technology has increased the amount of income, wage, and economic information available to potential migrants prior to making the migration decision. One impact may be that by increasing the base level available information around economic dynamism and other dimensions of potential destinations we are able to reduce the number of migrations that are unsuccessful and thereby reduce reverse migration. Further work would involve dividing the sample into pre- and post-recession partitions to see if the model provides a better explanation in one period relative to the other. Additionally, future explorations could also address the how the quantity of information on dynamism interacts with the speed of the information generating process to impact migration.

Conclusion

The preceding chapters of this dissertation strived to answer the question "what makes a place attractive?" by examining three different aspects of the regional economics literature. Each chapter approached this question through a separate lens and highlighted some of the challenges unique to the regional literature. This investigation considered the relationship between employment portfolio measures of risk and return with Colorado migration dynamics, how to measure economic resilience, and the role of information generated through dynamism in the decision to migrate. As this work concludes, it is important to summarize the main insights and lessons that were shared over the course of this dissertation.

In Chapter 1 we examined the role that regional employment theory can have in explaining migration into Colorado at the county level. In doing so we found that Colorado counties show clear evidence of the labor pooling effect of agglomeration economies discussed by Marshall (1920) and Krugman (1992). This can be seen in Figure 1-2, where the scatterplot of portfolio risk versus portfolio return shows that employees in Colorado's largest counties are achieving higher risk adjusted returns to employment compared to employees in medium or small population counties. Evidence of this effect is also visible in heat maps (Figures 1-3 and 1-4), which display the employment portfolio return and risk measures from a spatial perspective. Additionally, while it is possible to construct measures of portfolio risk and return at the Supersector level, and while such a model is likely more precise given the decreased AIC compared to the aggregate model with population cuts, the tradeoff comes at the expense of the increased explanatory difficulty created by introducing 20 new independent variables into the model. This realization, coupled with a lack of consistency in the Supersector measures across varying population cuts, supports the idea that the best way local and regional governments can attract migrants is by focusing on providing an overall stable economic environment, rather than by trying to focus on any one specific sector. Finally, this chapter highlights that, from the regional economic perspective, when investigating county level questions, the aggregate portfolio measures provide the best balance of explanatory power and complexity, while industry or Supersector level measures are best saved for industry specific case studies.

Chapter 2 explored the concept of regional economic resilience and the associated difficulty of defining and measuring this idea. The concept of economic resilience will continue to benefit from work designed to provide a cohesive definition and empirical construct that is agreed upon by the academic community. We provided seven different dependent measures of resilience; however, the correlation matrix of Figure 2-4 shows a limited association between these definitions. Precision and clear definitions are of primary importance for works published in this domain. While definitions show limited correlation, it is possible to identify determinants of resilience that are robust across definitions. This was accomplished by testing the significance of the independent variables across the regression models. However, as seen in the empirical results, the predicted impact of a given determinant may differ depending on if the measure of resilience is income based or employment based. The Variance-Mean Ratio model considered in Chapter 2 provides a way to measure stability but does not provide an intuitive way to determine if an economy is experiencing a positive or negative growth path. We showed that the same VMR can be associated with very different experiences, as seen in Figure 2-5. Additionally, stability may be synonymous with stagnation, if it means missing out on the broader recovery experienced at the national level. These results provide new insight to existing measures of resilience and will help clarify the empirical construct of resilience in the literature. The findings generated in this chapter can provide insight and guidance for local policy makers as they construct policies and development policies to balance growth with resilience, with the understanding that the policies may need to be distributed intertemporally. Additionally, it may be feasible to identify policies that positively impact resilience and growth simultaneously, such as encouraging inflows in the segment of the population between 18 and 35. Similarly, finding ways to shift employment from manufacturing to more service-based sectors may result in both increased resilience and increased economic growth.

Finally, Chapter 3 considered the role of information on dynamism as a psychic cost and how this impacted migration into counties across the United States. When considering all the models investigated, the included informational variables provided a statistically significant contribution to the model of migration. This provides evidence that information from dynamism is incorporated into the decision to migrate and should be considered as part of psychic costs when formulating an individual's utility function. Additionally, F-tests on the lag structure of these variables presented evidence that historical information on dynamism, not just the most recent observation, plays an important role in this process. These tests also showed joint significance, both within an individual information variable and across the entire set of variables representing information on dynamism. The difference in average annual in-migration rates between the single- and multi-county MSAs was explored by both a Blinder-Oaxaca decomposition on the migration rates and by running the first difference model. A T-test showed that there was a statistically significant difference in means between the two groups. Moreover, the Blinder-Oaxaca decomposition showed that significant contributions to this difference came from each of the three portions of the decomposition: the endowment effect, the coefficient effect, and the interaction effect. The main drivers of the approximately half a percentage point difference in mean in-migration rates appear to be driven by differences in population and the first lag of the contraction rate. The regression results for the two groups highlighted a more pronounced role for the information on dynamism variables in the decision-making process of migrating to the single-county MSAs, as compared to the larger, multi-county MSAs. One potential reason for this, and an avenue for future research, is that multi-county MSAs provide a heterogeneous experience across different portions of the MSA, while single-county MSAs contain a more singular, cohesive identity. Therefore, information plays a stronger role when making the choice to move to a smaller MSA, as the potential migrant must ensure a higher degree of matching with the MSA's core identity.

While the goal of any dissertation is to move the literature forward, it also serves as a catalyst for future research. Avenues for additional research certainly came to light over the course of the three prior chapters. In Chapter 1 we raised the point that industry level portfolio measures would be most useful when considering industry use cases, and as seen in Watson & Deller (2022) industry specific questions around economic stability are already being pursued as they examine tourism, resilience, and recessions. Using portfolio measures in these types of investigations would be a way of marrying concepts explored in Chapters 1 and 2. Additionally, the work of Chapter 1 could be furthered by more fully addressing the role of distance in migration decisions. Slicing the data set into distance bands could provide more insight into how the role of information from dynamism varies by move length. Most of the empirical frameworks in this dissertation utilized exogenous controls in their specification. The inclusion of more endogenous controls, such as the amount of art and cultural activities produced in a location could add an additional degree of awareness into what makes a place attractive. This could follow in the vein of Falck, Fritsch, & Heblich (2011) which examined the potential benefits of subsidizing cultural amenities to attract high human capital individuals. Lastly, Chapter 3 stressed the role of the quantity of information on dynamism in the decision to migrate but left unexplored the role the speed of information generation may have on these decisions. Additionally, more work could be done on determining which segment of the business cycle is best suited to use this type of migration model.

The overarching goal across the three chapters has been to explore the concepts of why individuals are attracted to a particular location and what factors work to determine attractiveness. The themes of economic risk and return, economic resilience, and the certainty needed to overcome the costs of leaving one's existing support system all play an important role in the migration decision. The insights uncovered in the preceding chapters, and summarized in this conclusion, show how this investigation has helped advance the literature in these areas. In a discipline that works to comprehend the motivations of individuals and groups, a more thorough understanding of the role that locations play in maximizing utility is a worthwhile pursuit. The literature continues to gleam new understandings into these themes and will provide insights to help governments and policymakers best portray themselves to those in search of new adventures in their ideal place.

Bibliography

Akerlof, G. A., 1970, The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, *84*(3), 488-500.

An, G., & Becker, C. M. (2013). Uncertainty, insecurity, and emigration from Kazakhstan to Russia. *World Development*, *42*, 44-66.

Anam, M., & Chiang, S. H. (2007). Rural–urban migration of family labor: A portfolio model. *The Journal of International Trade & Economic Development*, *16*(3), 325-335.

Anam, M., Chiang, S. H., & Hua, L. (2008). Uncertainty and international migration: an option cum portfolio model. *Journal of Labor Research*, *29*, 236-250.

Andersson, F., Burgess, S., & Lane, J. I. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, *61*(1), 112-128.

Arzaghi, M., & Rupasingha, A. (2013). Migration as a Way to Diversify: Evidence from Rural to Urban Migration in the US. *Journal of Regional Science*, *53*(4), 690-711.

Barth, J., Kraft, J., & Wiest, P. (1975). A portfolio theoretic approach to industrial diversification and regional employment. *Journal of Regional Science*, 15, 9-15.

Barry, F., & Kearney, C. (2006). MNEs and industry structure in host countries: A portfolio analysis of Irish manufacturing. *Journal of International Business Studies*, 37, 392-406.

Bauernschuster, Stefan et al (2014) Why are Educated and Risk-Loving Persons More Mobile Across Regions? *Journal of Economic Behavior and Organization*, 98, 56-69.

Bentivogli, C., & Pagano, P. (1999). Regional disparities and labour mobility: The euro-11 versus the USA. *Labour*, *13*(3), 737-760.

Blanchard, O., & Katz, L. F. (1992). Regional evolutions, brooking papers on economic activity. *Economic Studies Program, The Brookings Institution*, *23*(1), 76.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436-455.

Borjas, G. J. (1987). *Self-selection and the earnings of immigrants* (No. w2248). National Bureau of Economic Research.

Borjas, G. J., Bronars, S. G., & Trejo, S. J. (1992). Self-selection and internal migration in the United States. *Journal of urban Economics*, *32*(2), 159-185.

Bound, J., & Holzer, H. J. (2000). Demand shifts, population adjustments, and labor market outcomes during the 1980s. *Journal of labor Economics*, *18*(1), 20-54.

Bunten, D., Weiler, S., Thompson, E., & Zahran, S. (2015). Entrepreneurship, information, and growth. *Journal of regional science*, *55*(4), 560-584.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The review of economic studies*, *47*(1), 225-238.

Chandra, S. (2002). A test of the regional growth-instability frontier using state data. *Land Economics*, 78, 442-462.

Chandra, S. (2003). Regional economy size and the growth-instability frontier: Evidence from Europe. *Journal of Regional Science*, 43, 95-122.

Conroy, M. (1974). Alternative strategies for regional industrial diversification. *Journal of Regional Science*, 14, 31-46.

Cooke, T. J. (2011). It is not just the economy: Declining migration and the rise of secular rootedness. *Population, space and place*, *17*(3), 193-203.

County Distance Database. Retrieved from https://www.nber.org/research/data/county-distance-database

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., & Webb, J. (2008). A place-based model for understanding community resilience to natural disasters. *Global environmental change*, *18*(4), 598-606.

Damm, A. P. (2009). Determinants of recent immigrants' location choices: quasi-experimental evidence. *Journal of population Economics*, *22*(1), 145-174.

Dao, M., Furceri, D., & Loungani, M. P. (2014). *Regional labor market adjustments in the United States and Europe* (No. 14-26). International Monetary Fund.

Dao, M., Furceri, D., & Loungani, P. (2017). Regional labor market adjustment in the United States: trend and cycle. *Review of Economics and Statistics*, *99*(2), 243-257.

Davies, P. S., Greenwood, M. J., & Li, H. (2001). A conditional logit approach to US state-to-state migration. *Journal of Regional Science*, *41*(2), 337-360.

Davis, S. J., Faberman, R. J., & Haltiwanger, J. (2012). Labor market flows in the cross section and over time. *Journal of Monetary Economics*, *59*(1), 1-18.

Davis, S. J., & Haltiwanger, J. (2014). *Labor market fluidity and economic performance* (No. w20479). National Bureau of Economic Research.

Deaton, B. J., Morgan, L. C., & Anschel, K. R. (1982). The Influence of Psychic Costs on Rural-Urban Migration. *American Journal of Agricultural Economics*, *64*(2), 177-187.

Delgado, Mercedes, Michael Porter, and Scott Stern. (2010) Clusters and Entrepreneurship. *Journal of Economic Geography*, 10, 495-518.

Deller, S. C., Conroy, T., & Watson, P. (2017). Women business owners: A source of stability during the great recession?. *Applied Economics*, *49*(56), 5686-5697.

Deller, S. C., & Watson, P. (2016a). Did regional economic diversity influence the effects of the great recession?. *Economic Inquiry*, *54*(4), 1824-1838.

Deller, S., & Watson, P. (2016b). Spatial variations in the relationship between economic diversity and stability. *Applied Economics Letters*, *23*(7), 520-525.

Di Caro, P. (2017). Testing and explaining economic resilience with an application to Italian regions. *Papers in Regional Science*, *96*(1), 93-113.

Dostie, B., & Léger, P. T. (2009). Self-selection in migration and returns to unobservables. *Journal of Population Economics*, *22*, 1005-1024.

Ellison, Glenn; Glaeser, Edward L. and Kerr, William R. What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, June 2010, 100(3), pp. 1195-1213.

Etezady, A., Shaw, F. A., Mokhtarian, P. L., & Circella, G. (2021). What drives the gap? Applying the Blinder–Oaxaca decomposition method to examine generational differences in transportation-related attitudes. *Transportation*, *48*, 857-883.

Evenhuis, E. (2020). New directions in researching regional economic resilience and adaptation. In *Handbook on regional economic resilience* (pp. 69-86). Edward Elgar Publishing.

Fairlie, R. (2020). The impact of COVID-19 on small business owners: Evidence from the first three months after widespread social-distancing restrictions. *Journal of economics & management strategy*, *29*(4), 727-740.

Falck, Oliver; Fritsch, Michael and Heblich, Stephan. The Phantom of the Opera: Cultural Amenities, Human Capital, and Regional Economic Growth. *Labour Economics*, December 2011, 18(6), pp. 755-66.

Fingleton, B., Garretsen, H., & Martin, R. (2012). Recessionary shocks and regional employment: evidence on the resilience of UK regions. *Journal of regional science*, *52*(1), 109-133.

Fingleton, B., & Palombi, S. (2013). Spatial panel data estimation, counterfactual predictions, and local economic resilience among British towns in the Victorian era. *Regional Science and Urban Economics*, *43*(4), 649-660.

Foster, K. A. (2007). A case study approach to understanding regional resilience. Working Paper 2007–08, Institute of Urban and Regional Development, Berkeley

Freddie Mac, 30-Year Fixed Rate Mortgage Average in the United States [MORTGAGE30US], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MORTGAGE30US, March 27, 2023.

Glaeser, E. L. (Ed.). (2010). *Agglomeration economics*. University of Chicago Press.

Glaeser, Edward and Matthew Resseger (2010) The complementarity between cities and skills. Journal of Regional Science, 50(1), 221-244.

Goetz, Stephan J. 1999. “Migration and Local Labor Markets.” In Web Book of Regional Science. West Virginia University, Regional Research Institute. Available online at: http://www.rri.wvu.edu/WebBook/Goetz/contents.htm

Greenwood, M. J. (1975). Research on internal migration in the United States: a survey. *Journal of Economic Literature*, 397-433.

Grogger, J and G H Hanson. 2011. “Income Maximization and the Selection and Sorting of International Migrants.” *Journal of Development Economics* 95 (1): 42–57

Guimaraes, P., Figueirdo, O., & Woodward, D. (2003). A tractable approach to the firm location decision problem. *Review of Economics and Statistics*, *85*(1), 201-204.

Guimaraes, P., Figueiredo, O., & Woodward, D. (2004). Industrial location modeling: Extending the random utility framework. *Journal of Regional Science*, *44*(1), 1-20.

Gulati, B., & Weiler, S. (2021). Risk, Recessions, and Resilience: Towards Sustainable Local Labor Markets through Employment Portfolio Analysis. *Sustainability*, *13*(14), 7926.

Han, Y., & Goetz, S. J. (2015). The economic resilience of US counties during the great recession. *The Review of Regional Studies*, *45*(2), 131-149.

Han, Y., & Goetz, S. J. (2019). Predicting US county economic resilience from industry input-output accounts. *Applied Economics*, *51*(19), 2019-2028.

Harrison, B., Kelley, M. R., & Gant, J. (1996). Specialization versus diversity in local economies: The implications for innovative private-sector behavior. *Cityscape*, 61-93.

Hlavac, M. (2022). oaxaca: Blinder-Oaxaca decomposition in R. Social Policy Institute, Bratislava, Slovakia. R package version 0.1.5, https://CRAN.R-project.org/package=oaxaca.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual review of ecology and systematics*, *4*(1), 1-23.

Hong, S., & Xiao, Y. (2016). The influence of multiple specializations on economic performance in US Metropolitan Areas. *Sustainability*, *8*(9), 963.

Hyatt, H., McEntarfer, E., Ueda, K., & Zhang, A. (2018). Interstate Migration and Employer-to-Employer Transitions in the United States: New Evidence From Administrative Records Data. *Demography*, *55*(6), 2161-2180.

Hyatt, H. R., & Spletzer, J. R. (2013). The recent decline in employment dynamics. *IZA Journal of Labor Economics*, *2*(1), 5.

Ihrke, David. 2014. Reason for Moving: 2012 to 2013. Current Population Reports. P20-574. U.S. Census Bureau, Washington, DC.

Internal Revenue Service. (2022). Publication 501 (2022), Dependents, Standard Deduction, and Filing Information

Jackson, R. (1984). An evaluation of alternative measures of regional industrial diversification. *Regional Studies*, 18, 103-112.

Jacobs, J. (1969). Strategies for helping cities. *The American Economic Review*, *59*(4), 652-656.

Kaplan, G., & Schulhofer-Wohl, S. (2017). Understanding the long-run decline in interstate migration. *International Economic Review*, *58*(1), 57-94.

Kennan, J., & Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, *79*(1), 211-251.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American statistical association*, *50*(272), 1168-1194.

Kitsos, A., & Bishop, P. (2018). Economic resilience in Great Britain: the crisis impact and its determining factors for local authority districts. *The Annals of Regional Science*, *60*(2), 329-347.

Kluge, J. (2018). Sectoral diversification as insurance against economic instability. *Journal of Regional Science*, *58*(1), 204-223.

Kort, J. (1981). Regional economic instability and industrial diversification in the U.S. *Land Economics*, 57, 596-608.

Krugman, P. (1992). *Geography and trade*. MIT press.

Krugman, Paul (2010) The New Economic Geography, Now Middle Aged. Paper presentation at the Association of American Geographers.

Lande, P. (1994). Regional industrial structure and economic growth and instability. *Journal of Regional Science*, 34, 343-360.

Lang, W. W., & Nakamura, L. I. (1993). A model of redlining. *Journal of Urban Economics*, *33*(2), 223-234.

Lavelle, M., & Kepner, E. (2022, July 7). U.S. migration patterns before and after the start of the Covid-19 pandemic. *Midwest Economy Blog*. https://www.chicagofed.org/publications/blogs/midwest-economy/2022/migration-before-and-during-pandemic

Lee, N. (2014). Grim down South? The determinants of unemployment increases in British cities in the 2008–2009 recession. *Regional Studies*, *48*(11), 1761-1778.

Low, S. A., & Weiler, S. (2012). Employment risk, returns, and entrepreneurship. *Economic Development Quarterly*, *26*(3), 238-251.

Maddox, J. G. (1960). Private and social costs of the movement of people out of agriculture. *The American Economic Review*, *50*(2), 392-402.

Magrini, S. (2004). Regional (di) convergence. In *Handbook of regional and urban economics* (Vol. 4, pp. 2741-2796). Elsevier.

Marré, A. W., & Rupasingha, A. (2020). School quality and rural in-migration: Can better rural schools attract new residents?. *Journal of Regional Science*, *60*(1), 156-173.

Marshall, Alfred (1920). *Principles of Economics.* London, UK: MacMillan and Co.

Martin, R. (2012). Regional economic resilience, hysteresis and recessionary shocks. *Journal of economic geography*, *12*(1), 1-32.

Martin, R., & Sunley, P. (2015). On the notion of regional economic resilience: conceptualization and explanation. *Journal of economic geography*, *15*(1), 1-42.

Matti, J., & Neto, A. B. F. (2022). Consolidated city–county governments and economic stability. *Economics of Governance*, 1-24.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of public economics*, *3*(4), 303-328.

McGlade, J., Murray, R., Baldwin, J., Ridgway, K., & Winder, B. (2006). Industrial resilience and decline: A co-evolutionary framework. *Complexity and co-evolution: Continuity and change in socio-economic systems*, 147-176.

McGranahan, David and Wojan, Timothy. Recasting the Creative Class to Examine Growth Processes in Rural and Urban Counties. *Regional Studies*, April 2007, 41(2), pp. 197-216.

Modica, M., & Reggiani, A. (2015). Spatial economic resilience: overview and perspectives. *Networks and Spatial Economics*, *15*(2), 211-233.

Molloy, R., & Smith, C. (2019, October). US Internal Migration: Recent Patterns and Outstanding Puzzles. In *Prepared for the Federal Reserve Bank of Boston's 63rd Economic Conference. "A House Divided: Geographical Disparities in Twenty-First Century America," October* (pp. 4-5).

Molloy, R., Smith, C. L., & Wozniak, A. K. (2014). *Declining migration within the US: the role of the labor market* (No. w20065). National Bureau of Economic Research.

Molloy, R., Smith, C. L., & Wozniak, A. (2017). Job changing and the decline in long-distance migration in the United States. *Demography*, *54*(2), 631-653.

Nakosteen, R. A., & Zimmer, M. (1980). Migration and income: the question of self-selection. *Southern Economic Journal*, 840-851.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, 1-32.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693-709.

O'Connell, P. G. (1997). Migration under uncertainty: "Try your luck" or "Wait and see". *Journal of Regional Science*, *37*(2), 331-347.

Ormerod, P. (2010). Resilience after local economic shocks. *Applied Economics Letters*, *17*(5), 503-507.

Partridge, Mark (2010) The dueling models: NEG vs amenity migration in explaining US engines of growth. *Annals of Regional Science*, 89(3), 513-536.

Pierce, K. (2015). SOI migration data: a new approach: Methodological improvements for SOI's United States Population Migration Data, calendar years 2011-2012. *Statistics of Income. SOI Bulletin*, *35*(1).

Pimm, S. L. (1984). The complexity and stability of ecosystems. *Nature*, *307*(5949), 321-326.

Ray, D. M., MacLachlan, I., Lamarche, R., & Srinath, K. P. (2017). Economic shock and regional resilience: Continuity and change in Canada's regional employment structure, 1987–2012. *Environment and Planning A*, *49*(4), 952-973.

Rhoades, S. A. (1993). The herfindahl-hirschman index. *Fed. Res. Bull.*, *79*, 188.

Romer, R. (2001) *Advanced Macroeconomics*. New York: McGraw Hill.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, *3*(2), 135-146.

Rupasingha, Anil, Yongzheng Liu and Mark Partridge. 2015. "Rural bound: Determinants of metro to non-metro migration in the United States." *American Journal of Agricultural Economics* 97(3):680-700.

Saks, R. E., & Wozniak, A. (2011). Labor reallocation over the business cycle: New evidence from internal migration. *Journal of Labor Economics*, *29*(4), 697-739.

Schwartz, A. (1973). Interpreting the effect of distance on migration. *Journal of political economy*, *81*(5), 1153-1169.

Simmie, J., & Martin, R. (2010). The economic resilience of regions: towards an evolutionary approach. *Cambridge journal of regions, economy and society*, *3*(1), 27-43.

Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of political Economy*, *70*(5, Part 2), 80-93.

Spelman, W. (2006). Growth, stability, and the urban portfolio. *Economic Development Quarterly*, 20, 299-315.

Todaro, M. P. (1969). A model of labor migration and urban unemployment in less developed countries. *The American economic review*, *59*(1), 138-148.

Toukabri. A., & Delbé, C. (2022, May 26). New data reveal most populous cities experienced some of the largest decreases. *America Counts: Stories*. https://www.census.gov/library/stories/2022/05/population-shifts-in-cities-and-towns-one-year-into-pandemic.html

U.S. Census Bureau and U.S. Department of Housing and Urban Development, Monthly Supply of New Houses in the United States [MSACSR], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MSACSR, March 27, 2023.

U.S. Office of Management and Budget. (2010). 2010 standards for delineating metropolitan and micropolitan statistical areas; Notice. *Federal Register*, *75*(123), 37246-37252.

Watson, P., & Deller, S. (2022). Tourism and economic resilience. *Tourism Economics*, *28*(5), 1193-1215.

Weiler, S. (2006). A park by any other name: National Park designation as a natural experiment in signaling. *Journal of Urban Economics*, *60*(1), 96-106.

Wheaton, W. C., & Lewis, M. J. (2002). Urban wages and labor market agglomeration. *Journal of Urban Economics*, *51*(3), 542-562.

Wundt, B. (1992). Reevaluating alternative measures of industrial diversity as indicators of regional cyclical variations. *Review of Regional Studies*, 22, 59-73.

Zabel, J. E. (2012). Migration, housing market, and labor market responses to employment shocks. *Journal of Urban Economics*, *72*(2-3), 267-284.

Appendix A

Table A-1: RUCC model results

County to County In-Migration Count	Coef.	St.Err.	t-value	p-value	95% Conf	Interval	Sig
Overall Mean Employment Growth Rate	0.343	0.051	6.70	0.000	0.243	0.444	***
Overall Std. Deviation of Employment Growth Rate	-0.526	0.029	-17.84	0.000	-0.584	-0.468	***
Distance between Counties (100s mi.)	-2.366	0.213	-11.11	0.000	-2.783	-1.948	***
Distance between Counties (100s mi.) Squared	0.001	0.000	9.34	0.000	0.001	0.001	***
Median Household Income ($1,000s)	0.007	0.006	1.12	0.262	-0.005	0.020	
USDA Natural Amenity Scale	-0.122	0.032	-3.88	0.000	-0.184	-0.061	***
Population Density (Persons/square mile)	0.000	0.000	4.89	0.000	0.000	0.001	***
Rural-Urban Continuum Codes							
2	-0.122	0.267	-0.46	0.647	-0.645	0.401	
3	-1.548	0.319	-4.85	0.000	-2.173	-0.923	***
4	-2.713	0.328	-8.27	0.000	-3.355	-2.070	***
5	-1.588	0.185	-8.57	0.000	-1.951	-1.225	***
6	-3.568	0.346	-10.32	0.000	-4.246	-2.890	***
7	-3.597	0.264	-13.64	0.000	-4.114	-3.080	***
8	-5.031	0.528	-9.52	0.000	-6.066	-3.995	***
9	-5.462	0.406	-13.46	0.000	-6.257	-4.667	***

Mean dependent var	88.562	SD dependent var	2052.423
Number of obs	40387	Chi-square	2839.434
Prob > chi2	0.000	Akaike crit. (AIC)	3911041.803

**** p<0.01, ** p<0.05, * p<0.1*

Table A-2: Full Supersector model results

County to County In-Migration Count	Coef.	St.Err.	t-value	p-value	95% Conf	Interval	Sig
Mean Employment Growth Rate							
Natural Resources and Mining	-0.142	0.023	-6.05	0.000	-0.188	-0.096	***
Construction	0.281	0.050	5.56	0.000	0.182	0.379	***
Education and Health Services	-0.073	0.024	-2.98	0.003	-0.120	-0.025	***
Financial Activities	0.056	0.041	1.35	0.178	-0.025	0.137	
Government	0.173	0.138	1.25	0.211	-0.098	0.443	
Information	-0.104	0.045	-2.32	0.020	-0.192	-0.016	**
Leisure and Hospitality	0.068	0.084	0.81	0.418	-0.097	0.233	
Manufacturing	-0.065	0.030	-2.15	0.032	-0.125	-0.006	**
Other Services	0.016	0.048	0.34	0.737	-0.077	0.109	
Professional and Business Services	-0.034	0.043	-0.80	0.421	-0.118	0.049	
Trade, Transportation, and Utilities	0.051	0.041	1.24	0.215	-0.029	0.131	
Std. Deviation of Employment Growth Rate							
Natural Resources and Mining	0.040	0.007	6.01	0.000	0.027	0.053	***
Construction	-0.067	0.011	-5.94	0.000	-0.090	-0.045	***
Education and Health Services	0.070	0.037	1.92	0.055	-0.001	0.142	*
Financial Activities	-0.046	0.036	-1.27	0.204	-0.116	0.025	
Government	-0.136	0.065	-2.08	0.038	-0.264	-0.008	**
Information	0.020	0.015	1.30	0.192	-0.010	0.049	
Leisure and Hospitality	-0.109	0.038	-2.88	0.004	-0.184	-0.035	***
Manufacturing	-0.008	0.019	-0.41	0.679	-0.045	0.030	
Other Services	0.011	0.014	0.80	0.422	-0.016	0.038	
Professional and Business Services	-0.029	0.014	-2.00	0.045	-0.056	-0.001	**
Trade, Transportation, and Utilities	0.081	0.051	1.59	0.112	-0.019	0.180	
Distance between Counties (100s mi.)	-2.125	0.158	-13.48	0.000	-2.433	-1.816	***
Distance between Counties Squared (100s mi.)	0.001	0.000	11.12	0.000	0.001	0.001	***
Median Household Income ($1,000s)	0.041	0.008	5.12	0.000	0.025	0.057	***
USDA Amenity Scale	-0.058	0.051	-1.14	0.255	-0.159	0.042	
Population Density	0.000	0.000	3.37	0.001	0.000	0.001	***
Population Cuts							
< 15,000	0.000	.	.	.	.	.	
15,000 to 75,000	1.070	0.223	4.80	0.000	0.633	1.507	***
> 75,000	2.492	0.221	11.30	0.000	2.060	2.925	***
Mean dependent var		88.562	SD dependent var			2052.423	
Number of obs		40387	Chi-square			7869.594	
Prob > chi2		0.000	Akaike crit. (AIC)			2,756,162.293	

**** p<0.01, ** p<0.05, * p<0.1*

Appendix B

A model of migration incorporating information:

This section lays out a theoretical model that incorporates an information generating process into a model of migration. In this framework we will show that the greater the amount of information generated by a potential migration location, the smaller the variance. Our model draws heavily from two papers: embedding the information generating process of Lang & Nakamura (1993) into the utility function outlined in An & Becker (2013). The basic utility maximization problem exists where our individual or family must select the optimal location i^* to maximize the discounted lifetime utility function. If the discounted lifetime utility of another county minus the cost of moving exceeds the discounted lifetime utility of remaining in the origination location, then migration will occur.

$$m_{oi^*} = 1 \; if \; i^* \neq o \; maximizes \; \sum_{t=0}^{T} \beta^t u\big(y_{i*}(t)\big) - c_{oi^*} \geq \sum_{t=0}^{T} \beta^t u\big(y_o(t)\big) \quad \text{(B.1)}$$

$$m_{oi^*} = 0 \; if \; the \; originating \; county \; is \; the \; solution \; above$$

However, migration is a risky endeavor and individuals will face uncertainty over future utility flows, often stemming from uncertainty over the path of future income flows. The utility maximization problem should be restated as:

$$\sum_{t=0}^{T} \beta^t \Big(Eu\big(y_{i*}(t)\big) - u\big(y_o(t)\big)\Big) - c_{oi^*} \geq 0 \qquad \text{(B.2)}$$

This requires us to posit some version of how expectations are formed regarding future income flows in potential migration destinations. One way is to start with the current incomes in location i and project future flows based off what we know about past observations from location i.

$$E\big(y_{i,t+n}\big)|i,t = y_{i,t} + \sum_{\tau=t-p}^{k} \delta_\tau f(y_\tau - y_t) \qquad \text{(B.3)}$$

While the income in a specific location is likely to be observed most accurately by the people already living there, a potential migrant can use past observations of income to impute a best guess of what income is likely. This prediction becomes more accurate as larger pools of past information are available for analysis.

$$y_{idt} = y_{dt} + \epsilon_{idt}$$

Assume the potential migrant has secured a job in location i with an income of y_{idt}. The observed income consists of two parts. y_{dt} is the log of the average income for the potential migration destination in year t, while ϵ_{idt} is the random noise contained in the income. This reflects any bargaining power the individual may have, for example negotiating a wage at a new job. To simplify, assume that y_{dt} is a random walk where

$$y_{dt} = y_{dt-1} + \eta_t$$

ϵ_{idt} and η_t are assumed to be independent, normally distributed, mean-zero random variables with variances σ_ϵ and σ_η. We also assume that the potential migrant does not know the true mean value y_{dt}, they only observe their own income. Because of this, the migrant needs to form an expectation of what the path of future income values will be using a greater pool of information than just their own income. Because of the consequences resulting from a poor prediction of the future path of income, an individual may not be willing to move based off a single data point, as y_{idt} reflects a single observation. Individuals may want to gain a sense of the other jobs available in the region, their corresponding incomes, and the paths of these variables over time.

Potential migrants form their estimates of the incomes in other markets using the information available, which draws from past observations of income. We designate the optimal estimated income as $\mu_{id,t}|_{t-1}$. As we will show below, the precision of estimated income depends on the number of available past observations. Because information is noisy, particularly with wages in imperfectly competitive markets with regional differences, a greater number of observations leads to a more precise estimate of income in a potential migration destination.

$$\mu_{i,t}|_{t-1} = \mu_{i,t-1}|_{t-2} + \alpha_{t-1}\left(m_{t-1} - \mu_{t-1|t-2}\right) \quad \text{(B.4)}$$

Where

$$m_{t-1} = \frac{\sum_j \mu_{jit-1}}{N_{it-1}} \tag{B.5}$$

m_{t-1} is the mean of the incomes observed for the period t-1.

$\mu_{t-1|t-2}$ is the best estimated income for an individual living in location i last period.

N_{t-1} is the number of observations in location i last period.

$$\alpha_{t-1} = \frac{N_{it-1} s_{it-1|t-2}}{\sigma_\epsilon + N_{it-1} s_{it-1|t-2}} \tag{B.6}$$

$$s_{it|t-1} = \frac{\sigma_\epsilon \alpha_{t-1}}{N_{it-1}} + \sigma_\eta \tag{B.7}$$

$s_{it|t-1}$ measures the estimation error variance of $y_{id,t}|_{t-1}$. As the number of observed data points in a potential destination increases the variance is reduced, leading to a more certain estimate of the income one could expect in a given location.

$$\frac{ds_{it|t-1}}{dN_{dt-1}} = -\frac{\sigma_\epsilon s^2_{t-1|t-2}}{(\sigma_\epsilon + N_{dt-1} s_{it-1|t-2})^2} < 0 \tag{B.8}$$

Because the number of observed incomes can vary significantly by location, locations that generate more information (in the form of a high number of observations) are less risky as a potential migration location than destinations with fewer observations. The marginal effect of an increase in N_{dt-1} on $s_{dt|t-1}$ will be greatest in those markets with the fewest number of observations.

After conducting their analysis, the current period information set of the potential migrants includes the current income, y_{idt}, the predicted income of the location, $\mu_{i,t}|_{t-1}$, and the variance of the predicted income, $s_{it|t-1}$. The current income also contains information that can be used to update the predicted income. Let

$$\mu_{t|jt} = E(y_t | y_{jt}, I_{t-1})$$

$$s_{t|jt} = E\left[\left(y_t - \mu_{t|jt}\right)^2 \middle| y_{jt}, I_{t-1}\right]$$

The minimum mean squared error estimators would give

$$\mu_{t|jt} = \mu_{i,t|t-1} + \gamma_t\left(y_{jt} - \mu_{t|t-1}\right) \tag{B.9}$$

Where $\gamma_t = \frac{s_{it|t-1}}{\sigma_\epsilon + s_{it|t-1}}$,

$$s_{t|jt} = \sigma_\epsilon \gamma_t = \frac{\sigma_\epsilon s_{it|t-1}}{\sigma_\epsilon + s_{it|t-1}} \tag{B.10}$$

Equation (9) is the potential migrant's best estimate of the true income given their current income. Equation (10) gives the estimation error variance. Between them, Equations (9) and (10) define the information process. Equation (10) shows that $s_{t|jt}$ increases as $s_{it|t-1}$ increases. Using Equation (8) we can see that a decrease in N_{dt-1} increases $s_{t|jt}$. If fewer observations are available at time t-1 then potential migrants will be less certain of the true mean income in time t.

This process can be repeated using past time periods to generate additional information to be used in Equation (3). Rewriting (3) as

$$E\left(y_{idt,t+n}\right) | i, t = y_{idt} + \sum_{\tau=t-p}^{k} \delta_\tau f\left(\mu_\tau - \mu_{t|jt}\right)$$

tells us that the expected income at time $t + n$ is equal to the current income plus an adjustment based on the discounted sum of the one period differentials between the best estimates and the true income in the migration destination for k observed periods. The discount function δ_τ weighs periods closer to the present more heavily and decays the value of past periods.