The Pennsylvania State University

The Graduate School

Department of Agricultural Economics and Rural Sociology

ESTIMATING MIGRATION TO PENNSYLVANIA COUNTIES DUE TO MARCELLUS SHALE DRILLING ACTIVITY

A Thesis in

Rural Sociology and Demography

by

Andrew Kelly

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2012

The thesis of Andrew Kelly was reviewed and approved* by the following:

Mark Leach
Assistant Professor of Rural Sociology and Demography
Thesis Adviser

Diane McLaughlin
Professor of Rural Sociology and Demography

Kathryn J. Brasier
Associate Professor of Rural Sociology

Clare Hinrichs
Associate Professor of Rural Sociology
Coordinator, Graduate Program in Rural Sociology

*Signatures are held on file in the Graduate School

ABSTRACT

In 2003, the first natural gas well was drilled into the Marcellus Shale in Washington County, Pennsylvania. From then to the beginning of 2010, over 1,200 Marcellus wells were drilled across Pennsylvania. This type of rapid natural resource extraction has occurred in other parts of the country, and there is extensive literature about this “boomtown” phenomenon. However, most of the prior research has focused on the social impacts of rapid population change associated with the resource extraction rather than attempting to quantify the change that has taken place. Due to the unique nature of economic development associated with natural resource extraction development, current methods of population estimation may fail to fully capture the population change which is occurring. Using the U.S. Census Bureau’s population estimates for Pennsylvania counties, I find that after controlling for other factors that may affect population change and county population estimate accuracy, net migration estimates between 2000 and 2010 are underestimated relatively more in the northeastern counties of Pennsylvania, which is where most of the drilling activity has taken place thus far. This may indicate that population change is occurring at a rate faster than expected in this area and that population estimates are not fully capturing the change in this region. The same cannot be said for the southwestern counties of the state, which have also experienced extensive drilling activity. Considering the Bureau’s population estimates are used for allocating funds related to various government programs, these findings have important implications for state and local government agencies in areas which have experienced drilling activity, as well as those which expect to in the future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to acknowledge my thesis adviser, Dr. Mark Leach, for his support not only with this thesis but during my graduate school experience as a whole. I feel fortunate to have had Mark as a mentor and to have worked with him during my time at Penn State. I would also like to thank my other committee members, Dr. Diane McLaughlin and Dr. Kathy Brasier, for their critical feedback on this work. Their perspectives and expertise on the topic have added to this study immensely.

I am also grateful to the faculty and staff of the Penn State Rural Sociology program, who have provided a wealth of knowledge and guidance throughout my educational process. This experience has benefitted me both personally and professionally. Finally, I am thankful for the support of my family, particularly my parents, grandparents, and my wife Emily. My parents and grandparents have always supported me in everything I do, and this endeavor has been no different. Emily has also given me support and encouragement during the trying times throughout the process. Without her, this would not have been possible.

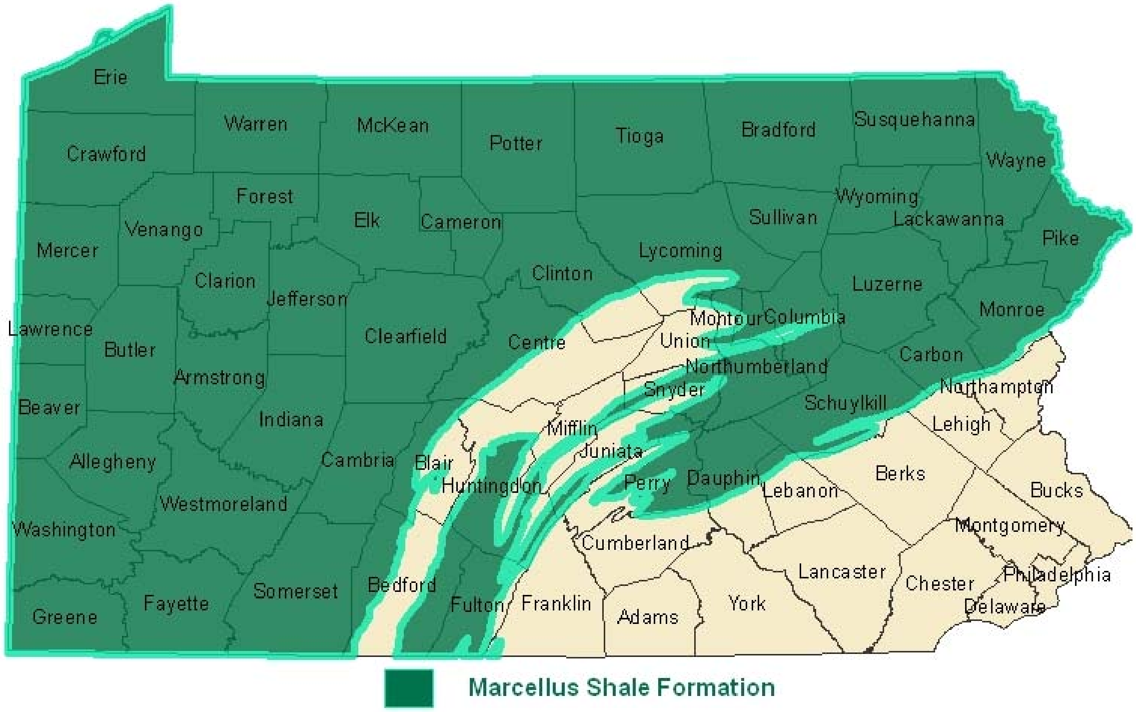
Chapter 1

Introduction

Beginning in 2003, Pennsylvania counties started experiencing dramatic growth in deep drilling activity aimed at extracting natural gas from the Marcellus Shale. Pennsylvania's Department of Environmental Protection reports that over 1,200 Marcellus Shale wells were drilled in the Commonwealth between 2003 and the beginning of 2010, but the number of wells across counties varies substantially. Of the 67 counties in the Commonwealth, nearly half contained at least one well while roughly three quarters of the total drilling activity had occurred in just seven counties (Bradford, Fayette, Greene, Susquehanna, Tioga, Washington, and Westmoreland) as of the beginning of 2010. Figure 1 shows an overlay of the Marcellus Shale formation across Pennsylvania. All seven high-activity counties are located in either the northeast or southwest region of the state and are within the Marcellus Shale portion of the state.

Development of Marcellus Shale wells requires a workforce, of course, to construct and manage drilling rigs. In addition, natural resource extraction often generates demand for locally manufactured parts and supplies, construction of new roads, homes and hotels, and provision of food, healthcare and educational services. Emerging research in particular Pennsylvania counties suggests that many workers and their families are migrating to fill such jobs and causing rapid population and demographic change (Ward et al., 2011). This is consistent with past research dedicated to the "boomtown" phenomenon, which describes the rapid population growth that often accompanies the development of natural resource

Figure 1. Marcellus Shale Formation in Pennsylvania



Source: Pennsylvania Department of Environmental Protection Office of Oil and Gas Management

extraction (for examples, see Gilmore and Duff, 1975; Gilmore, 1976; Cortese and Jones, 1977; Albrecht, 1978).

The extent to which development of the Marcellus Shale drilling is affecting county-level population change across all Pennsylvania regions remains largely unknown, however. Understanding this is critical because state, county, and municipal officials rely on timely and accurate population estimates to project demand for essential public services such as healthcare, schooling, and even waste collection and disposal. In addition, municipalities where drilling has not yet begun would likely benefit from incorporating accurate projections of the effects of drilling activity in their plans.

The main objective of this study is to assess how Marcellus Shale drilling activity may be affecting net migration in Pennsylvania counties. This will be done through an evaluation of U.S. Census Bureau county net migration estimates between 2000 and 2010 with respect to Marcellus Shale drilling activity. Other factors known to influence migration patterns will also be taken into account. This approach allows for a more nuanced view of the issue than simply looking at how county population totals changed from 2000 to 2010.

To accomplish the main objective of the study, this thesis first determines whether migration associated with Marcellus Shale drilling development is fully captured in the Bureau's net migration estimates. This is done through answering the following questions:

1. What type of Marcellus Shale drilling activity measure is most important when considering population change and migration in Pennsylvania counties?
2. Did greater than expected net migration occur in counties and regions of Pennsylvania where drilling activity has been relatively high?

The results of the study suggest that drilling activity has a regional effect on net migration to Pennsylvania counties. Based on a regional measure of drilling activity, greater than expected net migration did occur in counties within the northeastern region of the state, an area which contains the high drilling activity counties of Bradford, Susquehanna, and Tioga. This result holds true even after accounting for other factors known to influence migration.

Thesis Overview

This thesis is made up of five chapters along with an appendix and references at the end. This first chapter has briefly described the Marcellus Shale natural gas drilling development, as well as the "boomtown" phenomenon, and why there is the need for research on how it affects population change. Chapter 2 contains a review of the literature dedicated to boomtown migration and its effects as well as the characteristics of recent population change trends in Pennsylvania counties. This literature review serves as the basis for the independent variables used in the analysis. Chapter 3 contains a description of the data sources used, as well as the methodology for this study. Chapter 4 presents the empirical results of the study. The fifth and final chapter contains the main findings from this study, limitations of the research, policy implications, and possible avenues for future research on this topic.

Chapter 2

Literature Review

In this chapter, I begin by outlining the boomtown model and how it applies to migration due to the Marcellus Shale drilling development. There are three main components of population change: births, deaths, and migration. It is the working assumption of this study that most of the population change caused by the Marcellus Shale drilling activity will be due to migration rather than births and deaths. This assumption is supported by previous boomtown migration research which has shown the rapid influx of people which can accompany natural resource extraction. Following this, internal migration flows in general, and recent Pennsylvania county population change patterns specifically, are discussed. I conclude the chapter with my hypotheses for this study.

The Boomtown Model and Marcellus Shale Drilling Activity

In the rural sociological literature, “boomtowns” are defined as areas which experience rapid population growth over a short period of time due to increased economic activity associated with natural resource extraction. These areas usually have low population totals prior to the resource extraction and relatively little demographic diversity (Albrecht, 1978; Freudenburg, 1986). The study of this phenomenon began in the 1970s and 1980s with the examination of several small communities in the western United States experiencing significant population growth due to a dramatic increase in economic activity from energy extraction (Jacquet, 2009). These early studies detailed the economic and social impacts rapid population change had on the receiving communities. When natural resource extraction begins in a community, population levels and economic activity will vary with the amount of extraction being undertaken (Brown et al., 2005). It will also vary depending on the stage of development in which the community finds itself. For example, natural gas extraction from the Marcellus Shale requires a relatively larger workforce for the initial drilling of gas wells as compared to the workforce required in subsequent developmental stages (Marcellus Shale Education and Training Center, 2011). This influx of workers to drill wells is referred to as the "boom" in the overall boom-bust cycle which often accompanies the natural resource extraction process in general (Galston and Baehler, 1995). Workforce requirement reports estimate that the total number of hours needed to drill one Marcellus Shale well corresponds to 12.9 full-time equivalent jobs. Compare this to the 0.19 full-time equivalent jobs needed to extract the gas in the subsequent production phase, and it is clear that most of the labor is needed up front during the drilling of the wells (Marcellus Shale Education and Training Center, 2011). These drilling and production jobs are referred to as "direct labor", because these jobs are involved with work done directly on well sites or work done directly for energy companies (Marcellus Shale Education and Training Center, 2009).

The rapid influx of people brought on during the drilling stages will contain not only workers related to the natural resource extraction industry but also migrants moving to fill positions which have been created to meet the economic needs of the extraction industry workers themselves (The Perryman Group, 2008). These types of jobs are referred to as indirect and induced labor. Indirect labor includes jobs dedicated to supplying well sites with parts and materials needed for direct laborers to perform their tasks. Induced labor refers to jobs in industries not directly involved with the drilling or supplying materials to well sites. This includes jobs related to public services and housing construction, among others (Marcellus Shale Education and Training Center, 2009). The number of indirect and induced jobs created by drilling activity is harder to quantify than direct labor employment. However, the Pennsylvania Economy League (2008) estimates that 1.52 indirect and induced jobs are created for every direct job created in the Pennsylvania oil and gas industry as a whole. It seems reasonable that this would also apply to Marcellus Shale drilling activity specifically. The influence of direct labor totals on the demand for indirect and induced (non-direct) jobs means these non-direct job totals will also fluctuate as direct labor requirements change with the cyclical boom and bust nature of the overall natural resource extraction process. This adds to the overall impact of the boom phase of development on population change in, and migration to, the communities and areas where it is taking place.

A major criticism of early boomtown studies is that they examined communities after the boom of the natural resource extraction development had already begun (see Wilkinson et al., 1982). Therefore, their findings could not be compared to a pre-boom baseline, because there was no baseline measurement for the comparison. The first to conduct a study of the issue using data during as well as before the initial population boom was Brown and colleagues (1989) in Delta, Utah. They found that changes in community attitudes towards the resource extraction changed even before the extraction began, and that this pre-boom change accounted for most of the overall attitude change over the entire boom-bust cycle of development. Their results highlighted the importance of establishing expectations prior to the boom period in order to properly assess subsequent changes in social well-being. The same principle applies to this study of estimating migration to Pennsylvania counties resulting from Marcellus Shale drilling activity. Census Bureau county population estimates are used as an approximate baseline for population levels. A comparison of these estimates to actual 2010 Census county population counts will give an insight into how the beginning of the natural gas extraction boom in Pennsylvania has affected population totals at the county level.

The "bust" of the boom-bust cycle corresponds to the decrease in overall worker requirements in the post-drilling (production) phase of resource extraction. As previously mentioned, the worker requirement needed per well is much less once a well has been drilled and it begins producing gas. Although the overall work requirement is less, these jobs are more permanent than well drilling jobs, and they are cumulative in nature, meaning that the total number of production jobs increases as active, producing wells increase. In other words, the total number of active wells can increase from year to year, even while the number of wells being drilled in consecutive years may not increase. This would lead to an increase from year to year production job totals but no increase in drilling job totals because the number of wells in the drilling phase did not increase.

The Marcellus Shale drilling activity remains in the "boom" phase of development, as evidenced by the increase in the number of Marcellus wells drilled statewide from 324 in 2008 to 807 in 2009, 1,577 in 2010, and 1,936 in 2011 (PA Department of Environmental Protection, 2012). This means overall direct and non-direct job totals have likely increased with the rise in well drilling activity. However, it is not yet understood to what extent the drilling activity is affecting population change at the county level. The trend of well increases up to and including 2010 is particularly important within the scope of this study. It allows for an analysis of the issue by comparing Census Bureau population estimates to actual Census population totals for Pennsylvania counties in 2010, a time at which the boom phase of the Marcellus Shale drilling development was in motion. Differences between the estimated and actual population for 2010 will give insights into the extent to which migration is occurring in high drilling activity counties and regions of the state.

Internal Migration Flows and Pennsylvania County Population Change

Within this study's scope of analyzing greater than expected migration to Pennsylvania counties experiencing Marcellus Shale drilling activity, it is important to understand the determinants which drive migration in general. It is also important to take into account the population change trends across Pennsylvania counties in recent decades. Taking this into account in the overall analysis will help to separate migration created by the economic activity surrounding the drilling from migration attributable to other influences. For macro-level analyses such as this one, aggregate data that describe differences between regions are generally used to model migration flows. From a purely economic standpoint, the influencing factors include wages, employment opportunities, and housing costs in the respective regions. Individuals are drawn from one region to the other either through higher wages, greater employment opportunities, or lower housing costs. This is considered the disequilibrium perspective, which is one where wages, labor, or cost of living disequilibrium is eventually remedied by a movement of people from one region to the other (Lee, 1966; Greenwood, 1997). The main weakness of this model is that it does not account for migration which occurs in a pattern opposite to what would be expected under certain economic conditions. For example, it could not explain migration of people from higher wage regions to lower wage ones.

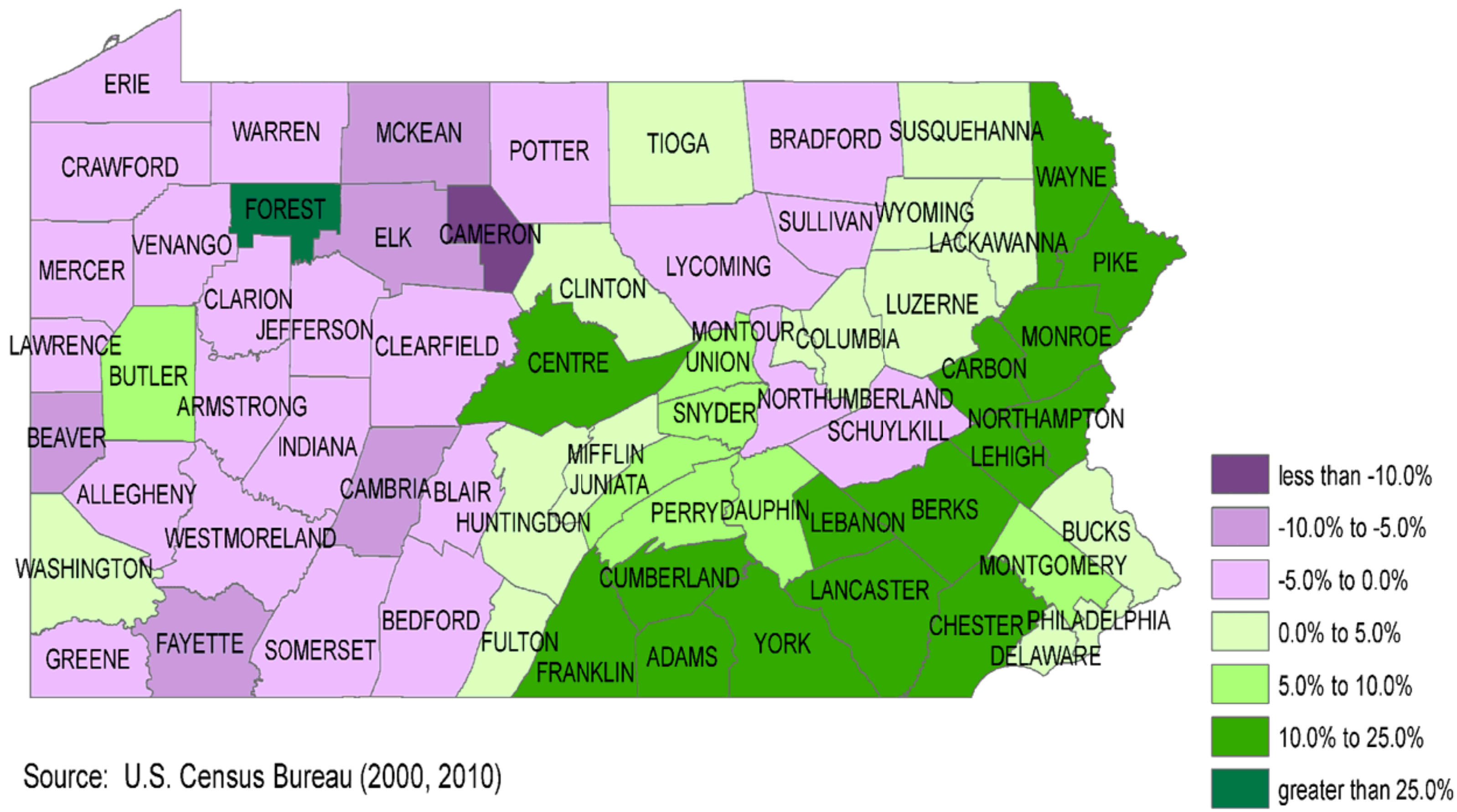
The equilibrium perspective extends the classical disequilibrium macroeconomic model to include non-economic factors which may influence migration decisions. It assumes that, given economic factors are similar across regions, non-economic factors play a bigger role in shaping migration flows (Greenwood, 1997). These non-economic factors can include proximity to amenities such as healthcare and other services as well as the level of natural amenities provided by the landscape in an area. Natural amenities have been shown to encourage population growth even in rural areas which would otherwise not experience growth (McGranahan, 1999). Also, McGranahan (2008) showed that migration to high natural amenity areas was suppressed by a shortage in housing opportunities represented by relatively high housing prices. Although it has been shown that natural landscape amenity levels can influence migration patterns, they are not expected to play much of a role in shaping migration in the context of this study. This is because the natural amenity levels are not relatively high in Pennsylvania as compared to other areas of the country, and there is not much variation in levels between different regions of the state. Access to social and cultural institutions provided by larger cities can also greatly influence population growth (Glaeser and Shapiro, 2001). Greenwood and colleagues (1991) showed that social amenity differences between regions influenced migration as long as they were not already captured in economic differences between regions such as wages and housing costs.

It is also important to understand that macro-level migration patterns are shaped by migration decisions made at the individual level. From this point of view, the classical microeconomic model is commonly used to explain an individual’s decision making process. This model explains micro-level migration as a comparison of the costs and benefits of the decision to migrate (Sjaastad, 1962). When the benefits of moving, such as an increase in wages or a decrease in the cost of living, outweigh the costs like moving expenses or the fracturing of social ties, the individual will decide to move. Researchers often view these micro-level decisions within the context of macro-level influences on the decision-maker (Findley, 1987; Massey, 1990). For example, the macro-level factor of available amenities may play a greater role in attracting certain demographics of people who can utilize those amenities.

Both the micro-level and macro-level perspectives are important within the scope of this study. At the individual level, Marcellus workers face a cost/benefit analysis, weighing the greater wage and employment opportunities presented by the drilling development against costs such as moving distance and expenses, as well as the possible separation from family members for long periods of time. When combined, these individual migration decisions form a macro-level pattern which fits in with the boomtown model. Workers are moving to regions of Pennsylvania which have been experiencing drilling activity, because that is where the labor is needed.

These perspectives also extend to population change in Pennsylvania counties in general. Figure 2 shows overall population change for Pennsylvania counties from 2000 to 2010. There are several stark trends illustrated in the figure. First, most of the high-growth counties are located in the southeast and extreme northeast portions of the state. The two high-growth counties in the northeast corner of the state, Monroe and Pike, saw their populations increase by 22.5 and 23.9 percent respectively over the decade. This is largely due the counties' proximity to the New York metropolitan area (Frank, 2012). A significant number of people employed closer to the center of downtown New York chose to relocate to Monroe and Pike counties during the decade to live in communities which they believe

Figure 2. Pennsylvania Percent Change in Population by County: 2000 to 2010



Source: U.S. Census Bureau (2000, 2010)

provide a higher quality of life than a community closer to the city, but which are also within driving distance of their workplace. Commuting access to the city provided by Interstate 80, which runs through northern Monroe County, makes this type of move feasible. The dramatic population increase in these two counties extends back even further than the most recent decade, as each county has experienced at least a 20 percent increase in population in each decade since the 1970 Census (U.S. Census Bureau, 2012). Overall, this type migration flow is consistent with a larger national trend of migration from inner cities to surrounding suburban areas (Mieszkowski and Mills, 1993).

A number of counties located in the southeastern portion of Pennsylvania also saw large population increases over the last decade. As Figure 2 shows, the highest growth counties in this region are more specifically located on the eastern portion of the state's southern border, and along part of the Pennsylvania/New Jersey border. The relatively higher growth in counties on the southern border has been attributed to their proximity the Baltimore/Hagerstown metropolitan area in Maryland. There is evidence that migration to these counties in the southeastern region of the state is occurring for both quality of life and financial reasons. Some people are moving because of the comparatively lower property tax rates afforded in Pennsylvania than in New Jersey and Maryland (New Jersey Star-Ledger, 2011).

Hispanic immigration has also driven population change in some areas of the southeastern part of the state. It has been a driver of population change in other parts of the United States since the 1960s, but in more recent decades, due to the passage of various pieces of legislation which have deterred Hispanic immigration to traditional receiving states, Hispanic immigrants have begun to settle in alternative, non-traditional destinations

(Zuñiga and Hernández-León, 2005; Massey, 2008). These new destinations include non-metropolitan areas as compared to traditional urban centers. This migration flow is not expected to be a significant driver of migration to or from many Pennsylvania counties, as Hispanics only made up 3.2% and 5.7% of the state's population in 2000 and 2010 respectively (U.S. Census Bureau, 2000, 2010). However, research has been conducted in Berks County which has a relatively high Hispanic concentration, so it is important to take this into account for the purposes of this study (Oropesa and Jensen, 2010).

The remaining high growth county, Forest County, can be ignored for the purposes of this discussion, as a State Correctional Institution was opened there during the decade. Therefore, the large population increase in this county between 2000 and 2010 is largely due to the transfer of prisoners to the new complex at its opening in October of 2004 (Barnes, 2004). If this complex had not been opened, it is likely Forest County would have experienced the same population decline as other counties in this north-central/northwestern portion of the state experienced over the last decade. The population decline in this area of the state extends over the last several decades as a whole as well. A large amount of this is due to highly educated college graduates leaving the area to gain employment opportunities in urban areas (IssuesPA, 2003). These individuals are able to see greater returns on their educational investment in the form of higher wages by moving closer to urban areas either within or outside Pennsylvania. Also, many of the highly technical jobs to be filled by educated individuals are located only in urban areas. This "brain drain" leaves behind a population with relatively lower education levels than the urban destination areas (Carr and Kefalas, 2009). This persistent, as well as recent, rural and non-metropolitan population decline is important within the scope of this study. Much of the Marcellus Shale drilling development has taken place in rural areas of Pennsylvania. Therefore, any population increase associated with the drilling development could be masked by the movement of younger, college-educated individuals out of these rural areas. It is important to take this into account when attempting to assess the effect of the Marcellus Shale drilling activity on population change.

The second largest metro area in the state, Pittsburgh, also experienced population decline over the last decade. Along with this overall decline, there has been a pattern of migration from Allegheny County (Pittsburgh) to surrounding counties over the past several decades (Briem, 2011). The biggest beneficiaries of this movement have been Butler, Washington, and Westmoreland counties. This suburbanization has been facilitated by the vast transportation infrastructure available in the southwestern portion of the state. The numerous interstate highways in the region have enabled people to live outside of downtown Pittsburgh even if their jobs are located downtown. Overall there are many dynamics shaping population change and migration patterns at the county level across Pennsylvania, and it is important to keep these dynamics in mind for the purposes of this study which analyzes migration due to the Marcellus Shale drilling development specifically.

Hypotheses

Given the past boomtown literature, the general expectation of this study is that migration will play a greater role in population change in Pennsylvania counties due to the Marcellus Shale drilling activity than population change due to births and deaths. Therefore, the following hypotheses of this study focus on the effect Marcellus Shale drilling could have on migration. The first hypothesis is that a regional measure of drilling activity will capture migration processes better than more quantitative measures of drilling activity. This hypothesis assumes that drilling activity in a county will have an effect on migration patterns in that county as well as surrounding counties. In other words, county lines may be arbitrary boundaries when considering the settlement of drill rig workers or other migrants associated with the drilling activity. Proximity to counties which have experienced extensive drilling activity may matter when considering migration patterns, even if a given county has not experienced a significant amount of drilling.

The second hypothesis of this study is that greater than expected in-migration occurred in Pennsylvania counties within high drilling regions of the state from 2000 to 2010. This expectation of greater than expected in-migration is supported by past boomtown literature which has documented rapid population change associated with natural resource extraction. It is also supported by more recent work specifically dedicated to the Marcellus Shale region of Pennsylvania which has documented the social and economic impacts of a rapid influx of migrants (Brasier et al., 2011; Christopherson and Rightor, 2011; Ward et al., 2011). Given these hypotheses, the next chapter details the data and methodology used to test them both, while accounting for other factors known to influence migration patterns.

Chapter 3

Data and Methodology

In this chapter, I describe the data sources used in this study, how the dependent variable and each independent variable are constructed, and the methodology used in the analysis. Counties within Pennsylvania are the level of analysis for this study. Forest County is excluded from the analysis due to the unique nature of its population growth during the 2000s. In 2004, a State Correctional Institution was opened in Marienville, PA, which is located in Forest County, and prisoners began being transferred to the location in October of that year (Barnes, 2004). For population estimate purposes, these prisoners are counted as migrants to the county. Therefore, this justifies removing Forest County from the analysis, as its net migration estimate and actual net migration total for the decade were biased by the inclusion of the prisoner population. No other county in the state experienced this type of unique event during the decade.

Data

The major data source for this study is the U.S. Census Bureau's County Population Estimates. The Census Bureau produces population estimates for each year between decennial Censuses at the national, state, and county levels. These estimates are needed for intercensal years to monitor population totals when no full count of the population is available. A population estimate is a calculation of the population at a point in time in the past, as compared to a population projection, which is an estimate of the expected population at some point in the future (Smith, Tayman, and Swanson, 2001). Therefore, the estimate for a given year is usually released in the following year. For example, the 2009 county population estimates were released in 2010. For the purposes of this study, these estimates provide a benchmark for determining whether population growth is occurring at rates faster or slower than expected at the county level due to the Marcellus Shale drilling development. This data source is used in the calculation of the dependent variable.

The second data source used is the Census Bureau's 2000 and 2010 decennial Censuses. The Census provides information on demographic, economic, and other characteristics of the population which are useful for migration related studies. County level data from the Census is used for this study to determine actual net migration totals for Pennsylvania counties during the decade of the 2000s. This source provides county population counts for 2000 and 2010 as well as demographic and economic county measures from 2000. The measures from year 2000 will help determine if certain counties were poised to experience population growth or decline based on certain population demographics or economic conditions at the beginning of the decade.

The third data source used in this study is the vital statistics database from the Pennsylvania Department of Health. The information used from this database includes birth and death totals for each county for each year between 2000 and 2009. 2010 totals were not available at the time the study was conducted, so those totals are estimated using the available 2009 data. The totals from these years are used to calculate the actual population increase or decrease due to births and deaths for each county between decennial Censuses (from April 1, 2000 to April 1, 2010). Calculating the amount of population change in counties between Censuses due to births and deaths will separate this type of population change from the population change due to net migration. This will then allow for a direct comparison between Marcellus Shale drilling activity and county population change due to net migration, rather than overall county population change for the decade.

The fourth data source used is the natural amenity scale from the United States Department of Agriculture's Economic Research Service (ERS). The natural amenity scale contains a numerical measure of the natural amenities present in each county in the lower 48 states. The measure has been used in the past to study migration patterns, specifically to and from rural areas (McGranahan, 1999; McGranahan, 2008). It is incorporated into this analysis to control for the relative level of natural amenities a county possesses, which in turn may influence the amount of migration to or from a county.

The fifth data source used in this study is the metropolitan/non-metropolitan county classification from the Office of Management and Budget. This was created in June 2003 using 2000 Census data and classifies counties based on population size and worker commuting patterns to adjacent counties. There are three county classifications: metropolitan, non-metropolitan adjacent, and non-metropolitan non-adjacent counties.

The sixth data source used comes from the Pennsylvania Department of Environmental Protection's Office of Oil and Gas Management. This division of the DEP provides information on the number of wells drilled and permits issued each year by county, and distinguishes whether each well or permit is for Marcellus Shale drilling or non-Marcellus Shale drilling. Therefore, this data allows for a direct measurement of Marcellus-specific drilling activity in each county.

Main Measures

Dependent Variable - Percent Net Migration Estimate Error

By definition, a population estimate is the calculation of the population at a prior date, as compared to a population projection, which is the calculation of the population at a date in the future (Smith, Tayman, and Swanson, 2001). The 2010 Census Bureau county population estimates provide a unique opportunity to assess the accuracy of the Bureau's estimates. The 2010 county estimates are produced by the Bureau prior to the release of the actual 2010 Census count for each county, and therefore, the estimates can be compared to the actual county Census counts to assess their accuracy.

To estimate county population totals, the U.S. Census Bureau uses the component of population change method, where they separately estimate births, deaths, net domestic migration, and net international migration for each county (U.S. Census Bureau, 2009). Administrative records and other data sources are used to estimate each of these four components of population change. Vital statistics from state government agencies are used to help estimate birth and death totals. As previously mentioned, most of the population change due to the economic activity surrounding the Marcellus Shale drilling development is expected to come from net migration rather than births and deaths, so these birth and death estimates are not analyzed directly in this study. However, they are important for separating county population change due to net migration from overall county population change.

Considering most of the population change associated with Marcellus Shale drilling activity is expected to be due to net migration, the main concern with the Census Bureau's population estimates methodology for the purposes of this study lies with the way they estimate net migration. Particularly, the way they estimate domestic migration for the working age population is of importance. Estimates for net domestic migration are broken down into two sub-estimates: under age 65 and age 65 and over net migration estimates. Data from Federal income tax returns produced by the Internal Revenue Service (IRS) is used to estimate net migration for the under age 65 population. The total number of in-migrants, out-migrants, and non-migrants for each county are calculated by matching individual tax filings by social security number for consecutive years and comparing addresses on the two returns. If the individual's address did not change between filings, then that person and the number of dependents listed on the filing are considered non-migrants for that year for the county in which their address is located. If the individual's address was in another county on the first year's filing and in a given county for the second year's filing, that person and the dependents listed on the filing are considered in-migrants for that year for the given county. Finally, if the individual's address was in a given county for the first year's filing and in another county for the second year's filing, that person and the dependents listed on the filing are considered out-migrants for that year for the given county. Net migration rates are calculated using these non-migrant, in-migrant, and out-migrant tabulations, and the rates are applied to the overall population estimate calculations. The over age 65 net domestic migration estimates are calculated using Medicare enrollment data. Consecutive year changes in a county's Medicare enrollment are used in conjunction with total Medicare enrollees at the beginning of the period to create an over 65 net migration rate in a similar way that IRS data is used to calculate the rate for the under 65 population. This rate is then used to estimate the over 65 net migration total for the county.

Finally, net international migration is estimated using 2000 Census, American Community Survey (ACS), and Armed Forces data. County totals are calculated by applying information from the 2000 Census to state and national immigration and emigration totals. These county totals are then compared to the state totals to ensure that the sum of all county net international migration totals matches the overall state total.

Since the under age 65 net domestic migration methodology is of most importance to this study, it is important to consider the accuracy of this component of the Bureau’s population estimates for Pennsylvania counties, particularly counties associated with the Marcellus drilling development. To assess the Bureau’s net migration estimate totals from 2000 to 2010, the population counts from the 2010 decennial Census can be used. People who resided in different counties in their 2000 and 2010 Censuses, respectively, are considered migrants at the county level for the decade, for Census purposes. However, there are several reasons Marcellus-related migrants may be counted as migrants in the decennial Census, yet not be fully captured as migrants by the Census Bureau’s method of estimating net domestic migration using IRS tax filing data. One reason is the difference in the way residence and, in turn, migration is measured between the decennial Census and the Bureau’s population estimates. The decennial Census form asks for the number of people living at a location on April 1, 2010, while the Bureau uses IRS administrative data to track working-age migration for its population estimates. If there is no change of address between a filer’s consecutive year tax forms, that individual is counted as a non-migrant in the county of residence listed on their tax filings. Therefore, if someone did not change addresses on subsequent tax forms, but was living at a residence on April 1, 2010 (either permanently or temporarily) different from the one they resided in on April 1, 2000, it is possible they will not be counted as a migrant in the IRS administrative data but will list that new residence in their 2010 Census response and be counted as a migrant for Census purposes. This type of error may be particularly appropriate at describing much of the Marcellus drilling workforce. Many workers are living in hotels or temporary trailers near their worksite in Pennsylvania while maintaining permanent residences in the areas from which they moved (Christopherson and Rightor, 2011). This type of error can also apply to individuals who moved and are filing for the first time, such as younger people obtaining a first job in an area where they did not previously reside. Demographers refer to this type of error as measurement error, in that the Bureau's method of capturing migration would not actually measure these people as migrants.

A second possible source of difference between estimated and actual net migration totals for a county is the amount of actual migration which takes place within the last two years of the decade. The IRS data used for the Bureau's net migration estimates is the most recent data available, which is usually lagged by up to two years (U.S. Census Bureau, 2009). For example, the 2010 population estimates were produced using comparisons of 2007 and 2008 tax returns, which do not capture migration in 2009 and 2010. Therefore, any differences between 2008 migration rates the rates in 2009 and 2010 will not be captured in the Bureau's population estimates, but will be reflected in actual net migration totals drawn from the 2010 Census population count. This source of error does not seem like it would be unique to counties associated with the Marcellus drilling activity, as any factor causing an increase or decrease in migration in 2009 and 2010 as compared to 2008 could create it. However, it could be particularly appropriate for counties which began experiencing Marcellus drilling activity prior to 2008 and may have been poised to experience relatively higher in-migration levels in 2009 and the beginning of 2010 than they did in previous years. For these reasons, it is important to take into account previous patterns of migration in Pennsylvania counties and the previously mentioned factors which influence migration in general to properly assess what could be causing errors in the Bureau's net migration estimates due to this reason. Demographers refer to this as coverage error, because data limitations in the Bureau's estimates methodology will cause those people moving in the last two years of the decade to not be covered. Another possible source of coverage error is people who did not or were not required to file a tax return in a given year. These people would not be counted as non-migrants, in-migrants, or out-migrants because they would not show up in the IRS data.

Since the Census Bureau uses a component of population change method to calculate their county population estimates, the corresponding formula for population change over a time period used in the estimates is:

$$EP_{t+n} = P_t + EB_{t-t+n} - ED_{t-t+n} + ENM_{t-t+n}$$

where EP_{t+n} is the estimated population at time t+n, P_t is the actual population count at time t, usually referred to as the "base population" from which subsequent population estimates are generated, EB_{t-t+n} is the estimated number of births between times t and t+n, ED_{t-t+n} is the estimated number of deaths between times t and t+n, and ENM_{t-t+n} is the estimated net migration total between times t and t+n. Estimated births and deaths can be combined into a "natural increase" component and indicated by an ENI_{t-t+n} term. Substituting this term in, the corresponding formula is then:

$$EP_{t+n} = P_t + ENI_{t\text{-}t+n} + ENM_{t\text{-}t+n}$$

To calculate the estimated net migration component, the formula can be transformed algebraically to:

$$ENM_{t\text{-}t+n} = EP_{t+n} - P_t – ENI_{t\text{-}t+n}$$

Actual net migration components can also be calculated with the same formula using an actual population count at time t+n (P_{t+n}), if available, and actual birth and death totals for the time period (B_{t+n} and D_{t+n}), instead of an estimated population count and estimated birth and death totals. Usually, an actual population count will only be available in a Census year. The difference between this actual net migration total using Census population counts and the estimated total from the Census Bureau's population estimates can be dependent on several sources of error, as previously mentioned.

With the decennial Census, every ten years there is an opportunity to assess the accuracy of the Census Bureau's population estimates. This is because the Bureau produces a 2010 population estimate prior to the release of the 2010 Census population count, which can be compared to the 2010 Census count when it is released. Estimates of net migration and natural increase are also produced by the Bureau allowing for the assessment of net migration and natural increase estimates separately along with the assessment of the population estimates as a whole. The dependent variable for this study, county level Percent Net Migration Estimate Error (PNMEE), is calculated as the difference between a county's estimated net migration total from 2000 to 2010 and the county's actual net migration calculated using the 2000 and 2010 Census, divided by the 2000 population total for the county. Therefore, the dependent variable is calculated as follows:

$$PNMEE_{2000\text{-}2010} = (ENM_{2000\text{-}2010} - NM_{2000\text{-}2010}) / P_{2000}$$

and substituting the formulas for estimated and actual net migration, respectively:

$$PNMEE_{2000\text{-}2010} = [(EP_{2010} - P_{2000} - ENI_{2000\text{-}2010}) - (P_{2010} - P_{2000} - NI_{2000\text{-}2010})] / P_{2000}$$

This result is then multiplied by 100 for percent interpretation purposes. The PNMEE is constructed to capture the types of net migration estimation error previously mentioned, as well any other possible sources which would cause an error in the estimate of net migration for the decade. It will determine if faster or slower than expected migration occurred in a county during the decade. Typically this type of algebraic percent error is larger for estimates of areas with smaller populations (Smith, Tayman, and Swanson, 2001), as a small absolute error in the estimated net migration total can lead to a larger percent error in an area with a lower population because of the subsequent lower denominator in the algebraic percent error formula.

The overall general interpretation of the PNMEE is relatively straightforward. A negative value indicates an underestimation of net migration for a county during the decade, while a positive value indicates an overestimation of net migration for the decade. The more specific interpretation of the PNMEE is a little more nuanced. Both an underestimation and overestimation of net migration can be caused in two different ways. An underestimation of net migration (negative PNMEE) can be caused by either an underestimation of in-migration for a county or the overestimation of out-migration. Both have the same effect on the PNMEE. When considering the Census Bureau's method of estimating net migration, an underestimation of in-migration seems much more likely than an overestimation of out-migration if measurement error is expected as the main driver of the PNMEE. As previously mentioned, an underestimation of in-migration could be caused by the failure of in-migrants to be identified as movers on subsequent IRS tax filings, but identifying as a new resident of a Pennsylvania county in the 2010 Census. Conversely, an overestimation of out-migration would be caused by more out-migrants being identified through IRS tax filings than those who identified as residents of a different county in the 2010 Census than the one they listed on their 2000 Census. In other words, a person would have to list a new address on their tax form, yet identify at their previous address in the 2010 Census. This does not seem likely. If coverage error due to data limitations is expected as the main driver of a negative PNMEE, then either an underestimation of in-migration or an overestimation of out-migration could be causing the PNMEE. The interpretation of the PNMEE in this case will be largely dependent on past trends in population change for a county or region and how those trends might have changed in the last two years of the past decade.

In a similar way, if measurement error is driving the PNMEE, an overestimation of net migration (positive PNMEE) can be caused by either an overestimation of in-migration for a county, not very likely, or the underestimation of out-migration, more likely as just explained by the way IRS tax return data is used to measure migration. If coverage error is driving the positive PNMEE, then either an overestimation of in-migration or an underestimation of out-migration could be causing the PNMEE, and it is necessary to consider past population change patterns of that given county. Whether net migration is underestimated or overestimated, it is important to consider how measurement and coverage error may be producing the result when interpreting the dependent variable, and in turn, interpreting the effects of the independent variables on the dependent variable.

Main Independent Variable - Marcellus Shale Drilling Activity (M)

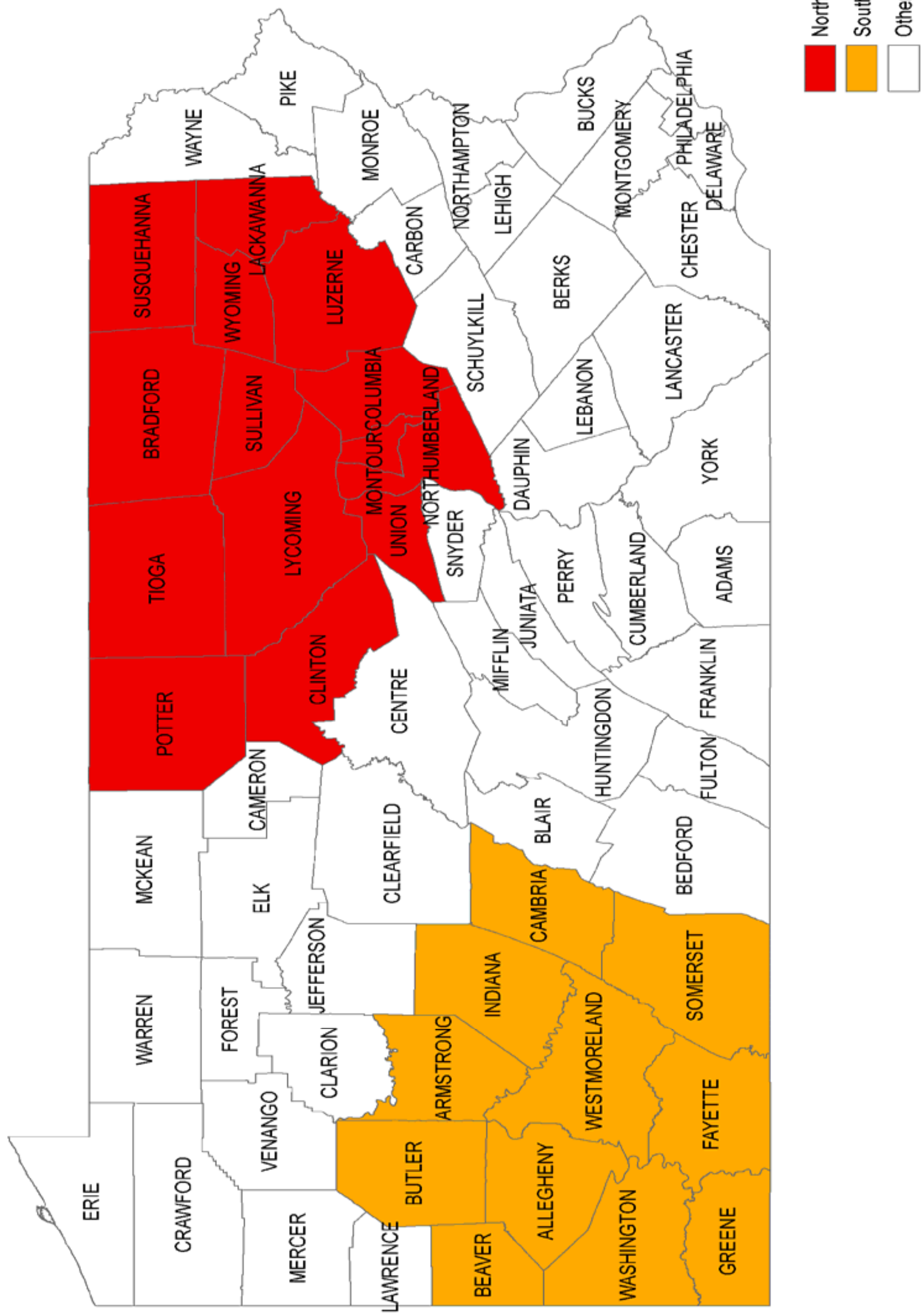
The main objective of this study is to estimate how Marcellus Shale drilling activity has affected net migration in Pennsylvania counties. Therefore, a measure of drilling activity in counties needs to be established. Three possible ways to measure this activity at the county level are examined in this study: total number of wells drilled, total number of drilling permits issued, and geographic place. The measure of total wells drilled in a county is a measure of physical drilling activity which has taken place in the county prior to the beginning of 2010. This type of measure is tested under the assumption that the actual process of drilling wells will influence the PNMEE due to the rapid influx of workers needed to work on the well rigs and any other migrants moving to the area because of increased economic activity. The second overall type of measure of drilling activity, number of drilling permits issued, is similar to the first type of measure, except it also includes well sites which were approved in a county as of the beginning of 2010, but had not yet been drilled. This measure is tested under the assumption that using a measure including only the number of wells drilled does not fully capture the drilling-related activity which is taking place within a county, and therefore it will not be highly associated with faster than expected migration to that county. In other words, the assumption for this second type of measure is the total drilling permits issued, including those permits issued for wells which were already drilled, is a better indicator of possible drilling rig worker migration, because it not only takes into account drilling which has already taken place, but also includes a count of wells which are likely to be drilled in the near future.

For the wells drilled and permits issued measures, four types of variable construction are tested. The first type of construction is simply a count of total wells drilled or permits issued as of the beginning of 2010. The assumption here is that the effect of drilling activity on net migration is linear. The second type is also a continuous measure, but the total wells and permits are logged. The assumption of this type of variable construction is that the effect of drilling activity on net migration is log-linear. In other words, the effect of an additional well or permit will matter less as the total number of wells or permits increases. For example, a well drilled county which has already had 100 wells drilled will matter less for net migration purposes than a well drilled in a county which has not previously had one drilled. The third type of construction is wells drilled or permits issued per a county’s area in square miles. This will test if well or permit density in a county is a more appropriate measure of drilling activity with respect to migration. The fourth type of well and permit variable construction is to create a dichotomous measure of Marcellus drilling activity using threshold levels. For example, if the selected well or permit threshold is 0, then any county with a non-zero well or permit total will have a value of 1 for the dichotomous variable and any county with no wells or permits will have a value of 0 for the dichotomous variable. The assumption here is that the effect of the drilling activity is not linear or even log-linear. Instead, a certain threshold level of activity may capture net migration processes better.

Different levels are tested for both the wells drilled and permits issued dichotomous threshold measures.

The third overall type of measure is based on the geographic location of the county compared to where most of the drilling has taken place thus far in Pennsylvania. High drilling activity has taken place in the northeastern and southwestern parts of the state. Therefore, counties from those two regions are indicated in the *Marcellus region* dichotomous variable. This variable is also split into two separate dichotomous variables, *Northeast Marcellus region* and *Southwest Marcellus region*, indicating whether each county is in the northeast or southwest region of the Marcellus formation, respectively. This allows for the measurement of greater than expected migration in each region separately. The Northeast Marcellus region is made up of Bradford, Clinton, Columbia, Lackawanna, Luzerne, Lycoming, Montour, Northumberland, Potter, Sullivan, Susquehanna, Tioga, Union, and Wyoming counties. The Southwest Marcellus region is comprised of Allegheny, Armstrong, Beaver, Butler, Cambria, Fayette, Greene, Indiana, Somerset, Washington, and Westmoreland counties. The geographic locations of the Northeast and Southwest Marcellus regions are shown in Figure 2. This type of variable construction will capture any “spillover effect” drilling activity in one county may have on the settlement of migrants in adjacent counties. In other words, it has the advantage of including greater than expected migration to counties which have not experienced extensive drilling or permitting but are adjacent to counties which have, if it is actually occurring. Theoretically this may make sense, as county lines may be arbitrary boundaries when considering the settlement patterns of drill rig workers and other migrants moving due to increased economic activity in an area.

Figure 3. Northeast and Southwest Marcellus Region

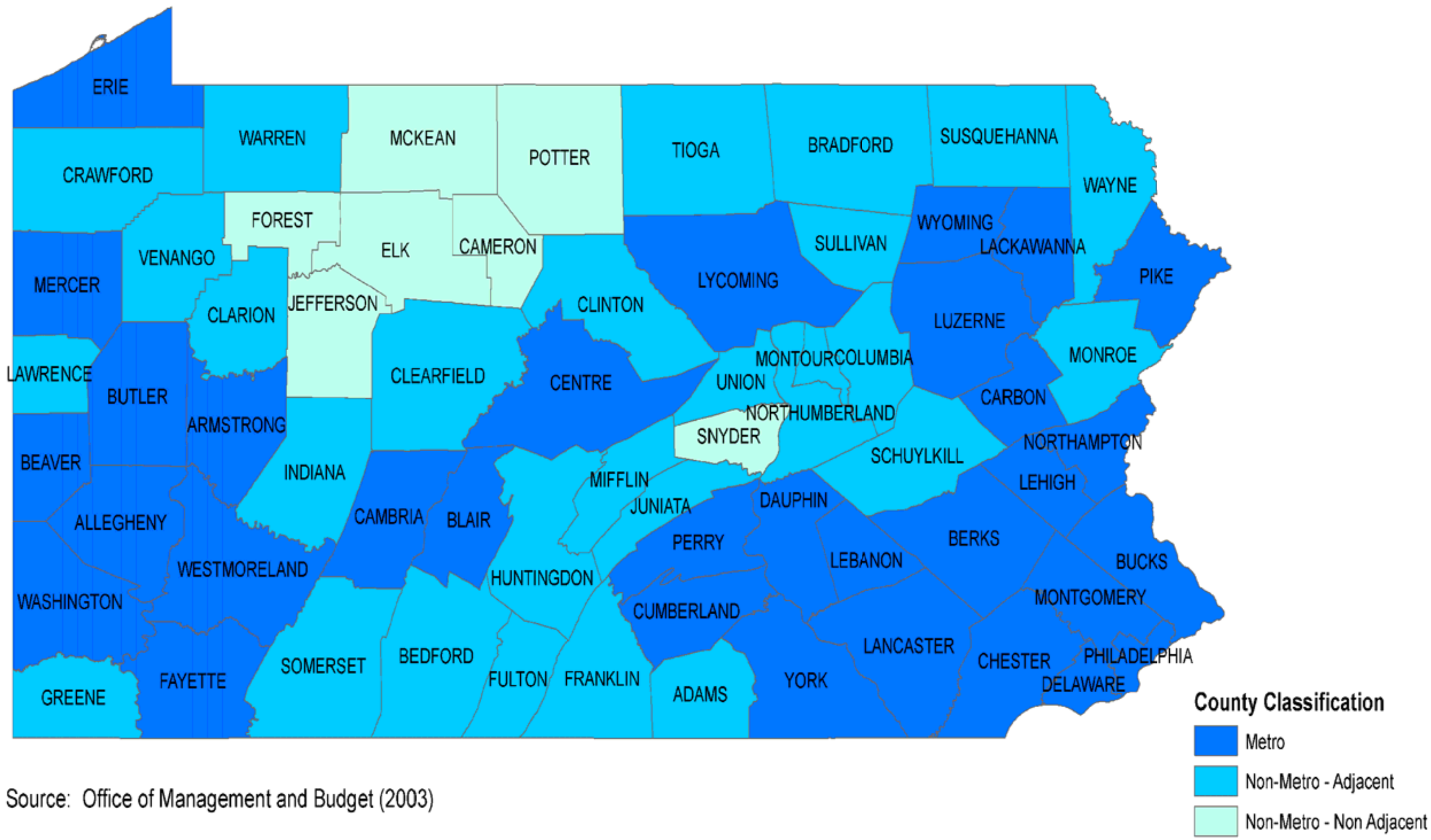


Additional Independent Variables

Amenity Measures

The first two amenity measures used in this study, *non-metropolitan adjacent* and *non-metropolitan non-adjacent*, help indicate a county's proximity, or lack thereof, to urban centers which may contain the benefits of social and cultural amenities. These are based on the 2003 Office of Management and Budget classification system developed using 2000 Census data. Metropolitan counties consist of the central metropolitan counties which contain an urbanized area of 50,000 people or more, as well as counties where 25 percent or more of its workers commute to a central metropolitan county, or if 25 percent or more of the employment in the county consists of people commuting from a central metropolitan county (Office of Management and Budget, 2003). In other words, these additional metropolitan counties are ones which are "economically tied" to the central metropolitan counties (Office of Management and Budget, 2003). The counties not meeting one of these metropolitan criteria are classified as either non-metropolitan adjacent or non-metropolitan non-adjacent. Non-metropolitan adjacent counties are those counties which are geographically next to a metropolitan county and have at least 2 percent of its workers commuting to a metropolitan county. Non-metropolitan non-adjacent counties are those non-metropolitan counties which do not meet the non-metropolitan adjacent criteria (Office of Management and Budget, 2003). Figure 3 shows a map of Pennsylvania counties by metropolitan/non-metropolitan classification.

Figure 4. Pennsylvania Metropolitan/Non-Metropolitan County Classification



Source: Office of Management and Budget (2003)

The second amenity measure, *natural amenity scale*, is a quantitative measure of the natural amenities present in a county. McGranahan (1999) used a measure of natural amenities calculated by the Economic Research Service (ERS) in his study of population change in non-metropolitan counties from 1970 to 1996. This measure was produced by the ERS in 1999, and it is comprised of county level scores based on the following six measures: mean January temperature, mean hours of sunlight in January, mean temperature in July, mean relative humidity in July, topographic variation, and the percent of the county's area which is covered by water (McGranahan, 1999). The overall raw standardized amenity score is used in this study instead of the ERS standardized rank scale to allow for a more precise calculation of the natural amenities present in each county. This also allows for more variation in the variable, as raw standardized scores for counties in Pennsylvania range from -2.78 to 1.26, while the standardized rank scale only ranges from 2 to 4, with all but two counties scoring a 3 or 4 on the rank scale.

Demographic Determinants

The following variables are all calculated from data contained in the 2000 U.S. Census. In order to take into account the racial/ethnic composition of each Pennsylvania county, *percent Hispanic*, is included in the analysis. It is calculated as the total number of people in the county who identified as Hispanic, divided by the total county population, multiplied by 100. Two variables are included in the analysis to take into account the age structure of each county. The first, *percent old*, is the number of people in the county age 65 or older divided by the total county population, multiplied by 100. The second variable relating to age structure, *percent young*, is the number of people in the county aged 0 through 14 divided by the total county population, multiplied by 100. A variable to take into account the marital characteristics of counties, *percent single*, is the number of single people in a county age 15 and older divided by the total county population age 15 and older, multiplied by 100. An education variable, *percent high school education or less*, is the number of people in a county age 25 and older with a high school education or less divided by the total county population age 25 and older, multiplied by 100.

Economic Determinants

Several variables measuring the aggregate economic conditions of each county are also incorporated into the analysis. These are all based on 2000 data, but come from different data sources. The first variable, *homeownership percentage*, is calculated as the number of occupied housing units being owned divided by the total number of occupied housing units, multiplied by 100. This is included to control for the propensity of a county's population to own a home, which may indicate a more stable population that is relatively less likely to migrate as a whole. The second variable, *percent in poverty*, is calculated as the number of families in poverty in the county divided by the total families in the county, multiplied by 100, using 2000 U.S. Census data. The third, *annual rent/wage ratio*, is the median value of all the rent responses given in the 2000 Census for each respective county divided by the median value of all wage responses given in the Census for each respective county, multiplied by 100. Median rent and wages are used instead of average values to construct this variable to account for the possibility of severe high or low outliers in the overall data, which is common when monetarily measured variables such as rent and wages are used in an analysis. Originally, rent and wages were to be included in the model as separate variables, but a collinearity analysis revealed that the two were highly correlated and multicollinearity was present. The correlation between the two variables was over 0.8, and the variance inflation factors from test multivariate regression runs were over 10 for both, indicating that the two variables are highly correlated and multicollinearity exists between the two (Williams, 2008). Therefore, it was necessary to combine the two into one ratio. These were the only two variables for which multicollinearity was present. Other correlations between economic variables were somewhat high, but not to the extent that they cause variance inflation factors over 10 in the test multivariate regression runs. A complete table of bivariate correlations for all independent variables can be found in the Appendix. The fourth and final economic measure, *unemployment rate*, is taken from the 2000 decennial Census and is calculated as the total population over 16 unemployed divided by the total population over 16 in the labor force, multiplied by 100.

Methodology

The analysis that follows first assesses whether greater than expected migration, as measured by Percent Net Migration Estimate Error (PNMEE), is associated with Marcellus Shale drilling activity using each of the four drilling activity variable types described above. If a general relationship exists, the analysis will continue with a multivariate regression analysis to understand if an association remains after controlling for other county-level determinants of migration. To explore a general relationship, a series of correlation analyses are used to compare the PNMEE of counties to the different Marcellus drilling activity measures.

If there is an association between one of the Marcellus Shale drilling activity measures and greater than expected migration as measured by the PNMEE, an ordinary least squares (OLS) multivariate analysis will be used to determine if this relationship remains true after controlling for other county-level migration determinants. To accomplish this, I will create a multivariate regression model using the Marcellus Shale drilling activity measure (M) determined in the first objective as the main independent variable. The model will also include the predictors mentioned earlier which are known to influence migration and population estimate accuracy: the level of amenities present in a county (A), the demographic factors of the county at the beginning of the decade (D), and the economic characteristics of the county at the beginning of the decade (E). The model is defined as:

$$PNMEE_{2000\text{-}2010} = f(M, A_{2000}, D_{2000}, E_{2000})$$

Groups of independent variables will be added to the model as blocks to help determine the impact of each group on the overall model. The Marcellus Shale drilling activity measure will first be added to the model by itself. Next, the non-metropolitan adjacent and non-adjacent indicators and the natural amenities scale will be added, followed by the group of demographic variables (percent Hispanic, percent old, percent young, percent single, and percent high school education or less), with the group of economic variables (homeownership percentage, percent in poverty, annual rent/wage ratio, and unemployment rate) added last.

The study concludes with a discussion of possible influential observations (counties) in the multivariate regression model and a robust regression analysis. Since the regression sample size is relatively small (N=66) it is possible the PNMEE of a few counties could dramatically influence the overall results of the regression model. The robust regression analysis accounts for influential observations by weighting each observation based on the likelihood it is influential. The higher the likelihood the observation is highly influential, the lower the weight it is given in the robust regression model. Two measures of influence will be used to identify possible highly influential observations, studentized residuals for overall observation influence and the DFBeta value for the main independent variables in the analysis. The studentized residuals are the residual difference between a county's PNMEE and its predicted PNMEE based on the regression model coefficients, divided by an estimate of the residual's standard deviation. Therefore, the studentized residual represents the number of standard deviations the observation's residual is from 0. They can assume positive or negative value, and the cutoffs for identification of possible high influence in the analysis are values less than -2.5 or greater than 2.5. The variable DFBeta value for an observation is the scaled difference between a variable's coefficient in the full regression model and the variable's coefficient in a model with that observation (county) deleted from the analysis. This is a measure of the effect an observation has on one coefficient, compared to the studentized residual, which is a measure of the effect an observation has on all coefficients in the model. DFBeta can have positive or negative values, and as a general rule of thumb, any observation with a DFBeta absolute value of over 2/sqrt(n) is considered a highly influential observation (Belsley, Kuh, and Welsch, 1980). In this analysis of 66 counties, that cutoff value is roughly 0.25. Therefore, any observations with a studentized residual less than -2.5 or greater than 2.5 or with a DFBeta less than -0.25 or over 0.25 are identified as possible influential observations. If the substantive robust regression results are similar to those of the regular regression results, it is reasonable to assume there are no observations which are heavily influencing the substantive conclusions drawn from the regular regression model.

Chapter 4

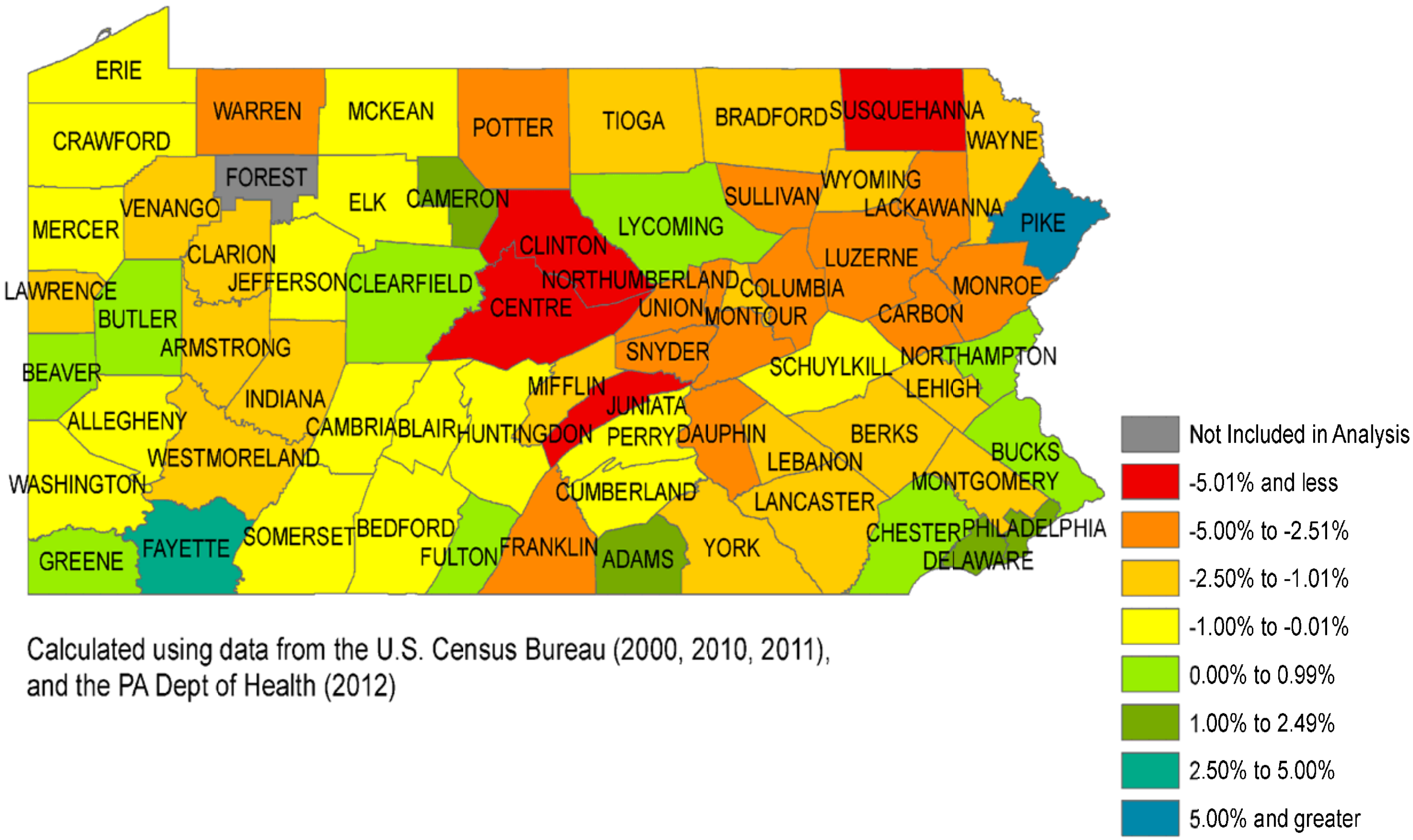
Empirical Results

In this chapter I show the results of the analysis. First, a map of the dependent variable, Percent Net Migration Estimate Error, is shown in Figure 5, and the descriptive statistics are presented in Table 1. Next, the bivariate analysis between the PNMEE and the different constructions of the main independent variable, Marcellus Shale drilling activity, is shown in Table 2. Following this are the results of the Ordinary Least Squares (OLS) multivariate regression model in Table 3. The chapter concludes with an influential observation analysis and robust regression analysis, shown in Table 4 and Table 5 respectively.

Descriptive Statistics

Figure 5 shows a map of PNMEE values across Pennsylvania counties, and Table 1 shows the descriptive statistics of all variables used in the analysis. There are several points worth noting that this figure and table illustrate. First, there was a severe net migration overestimation in Pike County for the decade, and in fact, this county represents the high end of the PNMEE range, with a value of 5.92 percent. Given the recent history of population change in Pike County, it seems reasonable that migration to the county could have slowed in 2009 and 2010 compared to earlier years in the decade, considering the economic recession began in 2008. Most of the population change in this county in recent decades has been due to people moving from inner New York City to suburban areas in the

Figure 5. Percent Net Migration Estimate Error (PNMEE) by County



Calculated using data from the U.S. Census Bureau (2000, 2010, 2011), and the PA Dept of Health (2012)

Table 1. Descriptive Statistics (N = 66)

	Mean	Standard Deviation	Minimum	Maximum
Dependent Variable				
Percent Net Migration Estimate Error (PNMEE) (%)	-1.35	2.23	-7.28	5.92
Population Estimate Measures				
Percent Natural Increase Estimate Error (%)	-0.05	0.22	-0.96	0.31
Drilling Measures[a]				
Wells B.O.Y. 2010	19.42	47.86	0	229
Permits B.O.Y. 2010	40.92	95.37	0	486
Wells per Square Mile	0.02	0.06	0	0.32
Permits per Square Mile	0.05	0.10	0	0.46
Marcellus Region	0.38	0.49	0	1
Northeast Marcellus Region	0.21	0.41	0	1
Southwest Marcellus Region	0.17	0.38	0	1
Rurality/Amenity Measures				
Non-Metropolitan Adjacent[b]	0.42	0.50	0	1
Non-Metropolitan Non-Adjacent[b]	0.09	0.29	0	1
Natural Amenity Score[c]	-0.16	0.78	-2.78	1.26
Demographic Measures (year 2000)[d]				
Percent Hispanic (%)	1.87	2.27	0.30	10.20
Percent Old (Age 65 and over) (%)	16.14	2.24	10.37	21.87
Percent Young (Age 0 to 14) (%)	19.25	1.60	15.00	22.40
Percent Single (%)	42.50	4.24	35.40	63.20
Percent H.S. Education or Less (%)	62.92	8.81	36.60	77.20
Economic Measures (year 2000)[d]				
Homeownership Percentage (%)	74.78	4.49	59.30	84.80
Percent in Poverty (%)	7.55	2.52	2.80	18.40
Annual Rent/Wage Ratio (%)	14.51	1.47	12.64	22.23
Unemployment Rate (%)	5.56	1.55	3.01	10.90

Sources: a - PA Department of Environmental Protection (2012); b - Office of Management and Budget (2003); c - Economic Research Service (1999); d - U.S. Census Bureau (2000)

Note: Forest County excluded from analysis

county. However, with the economic recession starting in 2008, this type of move may not have been as feasible for as many people after that year. Therefore, in-migration to the county may have been overestimated for the decade. The county's relatively high annual rent/wage ratio of 18.56 percent reflects a high demand for housing, which supports the idea that buying a house in the county may have become less feasible after the economic downturn.

In the central part of the state, there were three counties with significant net migration underestimation: Centre, Clinton, and Juniata. Centre County contains Penn State University, and the net migration underestimation could be due to more students than expected formally relocating their address to the county or large numbers of students remaining in the county after graduating. This may not have been captured accurately by the Census Bureau's estimation methodology if a large amount of this type of migration occurred in the last two years of the decade. This could have occurred if the economic recession of 2008 caused student to remain in Centre County because they were unable to find jobs after graduating. Clinton and Juniata counties are relatively small in population compared to other counties in the state, meaning a smaller absolute net migration underestimation could cause a more negative PNMEE than in other counties in the state, but their highly negative PNMEE is still worth noting.

Overall, net migration estimates for counties in the western portion of the state were more accurate than estimates for eastern counties. Estimates were most accurate for counties in the north-central and south-central portions of the state, as well as the northwestern corner. One western county, Fayette, did have a relatively large net migration overestimation with a PNMEE of 3.10 percent. One possibility for this overestimation could have to do with coverage error due to the prevalence of poverty in the county. Fayette County had the second highest poverty rate in the state in 2000 according to the 2000 Census, with 13.9% of its residents living below the poverty line. This could have made it difficult to track their migration patterns if these people were, in fact, less likely to be captured in tax return data when moving due to a lack of income. Cameron County also had a relatively large PNMEE, and its net migration was overestimated for the decade as well. Similar to Clinton and Juniata counties, this could be due to the fact that Cameron County had the second smallest population out of all counties in the state in 2000, and a small absolute error led to a relatively large percent error.

Table 1 shows the mean of the PNMEE is -1.35 percent, representing a general underestimation in the amount of net migration for all Pennsylvania counties from 2000 to 2010. Comparing the PNMEE mean to the Percent Natural Increase Estimate Error mean, it is clear that the errors in the Census Bureau’s estimates of net migration are the main cause of overall population estimate error. The mean of the natural increase error is -0.05 percent, nearly 0, and the standard deviation of 0.22 percent and range, -0.96 percent to 0.31 percent, are both much smaller than that of the PNMEE. This Percent Natural Increase Estimate Error is not included in the remainder of the analysis, but it is shown here to confirm the expectation that net migration estimate error is the main cause of overall population estimate error, suggesting that recent dynamics that attract new migrants, such as the Marcellus Shale drilling development, may be unaccounted for in these estimates. The drilling measures show that permits are issued much more readily in Pennsylvania counties than wells are drilled. The mean total permits issued as of the beginning of 2010 is nearly triple the mean wells drilled at the same point in time. Therefore, these two measures may be capturing two different Marcellus Shale development processes. The mean wells and permits per square mile also reflect this difference. The geographic Marcellus region measure contains 25 counties, corresponding to roughly 38 percent of the counties in the state. These are roughly evenly distributed between the Northeast Marcellus region and the Southwest Marcellus region, with 14 and 11 counties in each region, respectively. There are no unusual characteristics revealed in the descriptive statistics of the demographic or economic measures. Overall it appears recent population change patterns across Pennsylvania counties impacted the Census Bureau's net migration estimates for the decade, and that there was a general underestimation of net migration for counties in the state as a whole. A table of all county PNMEEs is located in the Appendix.

Bivariate Results

Table 2 shows the results of the bivariate correlation analysis between the PNMEE and the various types of Marcellus Shale drilling activity variable construction. A table showing well and permit totals for all counties in the analysis can be found in the Appendix. The wells drilled measures included total wells drilled in a county, logged total wells drilled in a county, wells per county area in square miles, and various thresholds of wells drilled in a county as of the beginning of 2010. None provided much explanation for the PNMEE observed in Pennsylvania counties. The signs of the correlations for different drilling measures were not consistent, and none of the correlations were strong enough to show statistical significance. The strongest correlation between the PNMEE and wells drilled was for the highest threshold level of 100 wells. However, even this correlation is not

Table 2. Bivariate Correlations of Percent Net Migration Estimate Error (N = 66)		
Drilling Measure Criteria	**Counties Meeting Criteria**	**Pearson Correlation Coefficient**
Total Wells Drilled	66	0.0063
Log of Total Wells Drilled	66	-0.0066
Wells per Square Mile	66	0.0441
Total Wells Drilled > 0	27	-0.0347
Total Wells Drilled >= 10	17	-0.0814
Total Wells Drilled >= 20	12	0.0557
Total Wells Drilled >= 50	7	0.0037
Total Wells Drilled >= 100	5	-0.1162
Permit Measure Criteria		
Total Permits Issued	66	-0.0392
Log of Total Permits Issued	66	-0.0457
Permits per Square Mile	66	-0.0125
Total Permits Issued > 0	32	-0.0979
Total Permits Issued >= 10	21	-0.0996
Total Permits Issued >= 20	19	-0.0970
Total Permits Issued >= 50	13	-0.0197
Total Permits Issued >= 100	8	0.0386
Total Permits Issued >= 200	5	-0.1162
Geographic Measure Criteria		
Marcellus Region	25	-0.1972
Northeast Marcellus Region	14	-0.4641***
Southwest Marcellus Region	11	0.2524*

+ significant p<0.10; * significant p<0.05; ** significant p<0.01; *** significant p<0.001

Source: PA Department of Environmental Protection Office of Oil and Gas Management (2012)

Note: Forest County excluded from analysis

statistically significant, due in part to the fact that only five counties met the 100 well criteria.

The permits issued measures included total permits issued in a county, logged total permits issued in a county, permits issued per county area in square miles, and various thresholds of permits issued in a county as of the beginning of 2010. These permit measures showed slightly stronger correlations than the well measures overall, but none were statistically significant. The over 200 permits issued measure showed the strongest correlation out of all permit measures, with a negative correlation of -.1162. However, this is the same statistically insignificant result shown for the over 100 wells drilled measure, as the same five counties meet this over 200 permits issued criteria.

The geographic measures by far show the strongest correlations with the PNMEE. The overall Marcellus region measure containing 25 counties had a stronger, negative correlation than any of the drilling or permit measures, indicating that a geographic measure captures errors in county net migration estimation better than quantitative well or permit measures. The negative value also indicates there is a trend in net migration underestimation for counties within the Marcellus region, and that this trend likely represents an underestimation of in-migration to those counties. The correlation was marginally significant, having a p-value of .1125. However, when this variable construction is broken down into its Northeast Marcellus and Southwest Marcellus components, there are stark differences between the two sub-regions. The Northeast Marcellus region was highly, negatively correlated with the PNMEE, indicating a strong trend of net migration underestimation in these 14 counties. In other words, more in-migration occurred than was expected to counties in this region during the decade, and this in-migration is associated with the geographic measure of Marcellus Shale drilling activity. It appears that an underestimation of net migration into the Northeast Marcellus region is driving the negative correlation for the Marcellus region as a whole. This is because the Southwest Marcellus region, by contrast, is positively and significantly correlated with the PNMEE. This reflects the fact that net migration for the decade was overestimated for 8 of the 11 counties in the region, as shown in Figure 3. There are several general differences between the Northeast and Southwest Marcellus regions which may be playing into this PNMEE correlation difference. First, the average year 2000 population for counties in the Southwest Marcellus region, 253,999, was much larger than that of counties in the Northeast Marcellus region, 79,154. Therefore in general, a population increase in a Southwest Marcellus county would have a smaller percentage increase than a similar size population increase in a Northeast Marcellus county because of the relatively smaller size of the Northeast Marcellus counties. Second, the transportation infrastructure in the Southwest Marcellus region is much greater than that of the Northeast Marcellus region. As noted in the earlier description of recent population change patterns across Pennsylvania, this infrastructure in the southwest part of the state has allowed for a steady stream of suburbanization from Pittsburgh and Allegheny County to surrounding counties. This could be masking some of the effect of Marcellus drilling activity on net migration for high-activity counties in the region such as Washington and Greene.

Multivariate Regression Results

Table 3 shows the OLS multivariate regression results for the Percent Net Migration Estimate Error using the Northeast and Southwest Marcellus region measures of drilling

Table 3. OLS Regression Analysis of Percent Net Migration Estimate Error (N = 66)

	Model 1	Model 2	Model 3	Model 4
Marcellus Drilling Measures				
Northeast Marcellus Region	-2.319***	-2.158***	-1.934**	-2.251**
	(.0003)	(.0011)	(.0048)	(.0022)
Southwest Marcellus Region	0.908	0.624	0.929	-0.197
	(.1812)	(.3816)	(.2456)	(.8361)
Rurality/Amenity Measures				
Non-Metropolitan Adjacent[a]		-0.958+	-0.697	-1.526*
		(.0744)	(.3092)	(.0314)
Non-Metropolitan Non-Adjacent[a]		-0.1511	-0.416	-0.912
		(.8677)	(.6866)	(.3610)
Natural Amenity Scale[b]		0.247	0.189	0.303
		(.4626)	(.5766)	(.4009)
Demographic Characteristics[c]				
Percent Hispanic			-0.045	-0.067
			(.7387)	(.6441)
Percent Old (age 65 and over)			0.187	0.053
			(.1745)	(.7122)
Percent Young (age 0 to 14)			0.497*	0.392+
			(.0158)	(.0916)
Percent Single			0.049	-0.087
			(.4588)	(.5697)
Percent H.S. Education or Less			-0.004	-0.028
			(.9134)	(.6076)
Economic Characteristics[c]				
Homeownership Percentage				0.055
				(.6038)
Percent in Poverty				0.066
				(.7941)
Annual Rent/Wage Ratio				-0.007
				(.9792)
Unemployment Rate				0.657*
				(.0214)
Constant	-1.008**	-0.535	-15.052*	-11.026
	(.0018)	(.2007)	(.0282)	(.4597)
R-squared	0.238	0.283	0.361	0.477
Adjusted R-squared	0.213	0.223	0.244	0.334

+ significant p<0.10; * significant p<0.05; ** significant p<0.01; *** significant p<0.001
Coefficient p-values shown in parentheses below coefficient value

Sources: a - Office of Management and Budget (2003); b - Economic Research Service (1999); c - U.S. Census Bureau (2000)
Note: Forest County excluded from analysis

activity. Model 1 includes the two Marcellus Shale region measures of the state separately, and the results are similar to those shown in the bivariate analysis. The significant negative coefficient for the constant in this model indicates there is a general trend in underestimating net migration for all Pennsylvania counties which is not captured in the Northeast Marcellus and Southwest Marcellus variables. This reflects the -1.35 percent PNMEE mean for all counties shown in Table 1. There is a strongly significant negative effect shown for the Northeast Marcellus region dichotomous variable. This shows that, on average, net migration for counties in the Northeast Marcellus region was underestimated by 2.32 percent more than counties not in the region. The Southwest Marcellus region variable shows a general overestimation of net migration for counties in the region, but its relationship with the PNMEE is not significant as it was in the previous bivariate correlation analysis. This is because the Northeast and Southwest Marcellus regions are now being accounted for simultaneously. This model containing only Marcellus Shale Region predictors accounts for 23.8 percent of the overall variance in the PNMEE for all observations.

Model 2 shows the addition of the non-metropolitan county classifications and natural amenity scale variables. The non-metropolitan adjacent coefficient is statistically significant at the 0.10 level. Its negative value indicates a general net migration underestimation for non-metropolitan counties adjacent to metropolitan areas, compared to metropolitan counties. When comparing the county classification map in Figure 2 with the PNMEE map in Figure 4, the negative non-metropolitan adjacent coefficient represents the general net migration underestimation for counties in the central and northeastern portion of the state. This could include population movement from the New York metropolitan area to Wayne and Monroe counties, as well as migration to Harrisburg (Dauphin County) and surrounding counties like Juniata, Snyder, and Northumberland. The natural amenity scale coefficient is relatively small and statistically insignificant in explaining the accuracy of county level net migration estimates. This was an expected result, as Pennsylvania counties do not score particularly high on this scale, and there is little variation in amenity scores across counties. The Northeast Marcellus coefficient remains highly statistically significant and negative in this model, with its coefficient only being reduced from -2.32 to -2.16, indicating a remaining effect of the Marcellus drilling activity on the underestimation of in-migration in the Northeast Marcellus region. The reduction of the constant in Model 2 indicates that the general trend in net migration underestimation shown by the constant of Model 1 has been partially explained by the addition of the non-metropolitan county classification and natural amenity variables.

Model 3 includes county demographic characteristics as predictors of the PNMEE. The results show percent young as a statistically significant predictor, with a coefficient value of 0.497. The positive coefficient indicates the larger the percentage of a county's population in the young category (age 0 to 14) in 2000, the larger the amount of net migration overestimation for the county for the decade, controlling for the other variables included in Model 3. There is no clear substantive interpretation of this coefficient from examining the PNMEE map in Figure 4. However, one possibility is that the higher the percentage of young people in the county in the year 2000, the greater potential there is for these younger people to move from their parents' residences by the end of the decade. Also, these young movers may be filing tax returns for the first time, and therefore would not register as migrants under the tax return methodology. This measurement error could cause net migration overestimation in those counties by underestimating the amount of out-migration in a county. This idea follows the “brain drain” migration flow mentioned in the earlier discussion of Pennsylvania population change patterns (Carr and Kefalas, 2009). The non-metropolitan adjacent coefficient is no longer significant now that county demographic characteristics are accounted for, and the coefficient magnitudes of the percent Hispanic, percent old, percent single, and percent high school education or less variables are all relatively small and statistically insignificant. The Northeast Marcellus region indicator magnitude remains relatively large and statistically significant. Its coefficient magnitude is reduced slightly, but the negative value still indicates significant net migration underestimation for counties within the region, which again likely means that in-migration was underestimated during the decade for those counties.

Model 4 includes economic predictors of the PNMEE along with all previously included variables. The only economic independent variable that was statistically significant is the unemployment rate. Its positive coefficient of 0.657 indicates the higher the unemployment rate for a county in 2000, the greater net migration was overestimated in that county. Theoretically this may make sense, as higher unemployment trends could be a driver of faster than expected out-migration of people looking for employment opportunities. In other words, a higher unemployment rate in 2000 could indicate counties which were more dramatically affected by the economic downturn later in the decade in 2008. This more dramatic impact could have, in turn, increased out-migration from those counties in the final few years of the decade. The percent young remains marginally significant with a positive coefficient value, indicating a similar, but slightly weaker, effect as what was shown in Model 3. Interestingly, the non-metropolitan adjacent coefficient becomes statistically significant again now that county economic characteristics are controlled for. The Northeast

Marcellus county indicator remains negative and statistically significant. Again, this suggests that greater than expected in-migration occurred in counties experiencing or close to Marcellus drilling activity in the northeast part of the state after accounting for other amenity, demographic, and economic indicators known to affect migration patterns. The addition of county economic characteristics to the model also increased the R-squared and adjusted R-squared by roughly 12 and 9 percent respectively.

Influential Observation Analysis Results

Table 4 shows the results of the influential observation analysis performed on the 66 Pennsylvania counties included in the analysis. The table includes the studentized residuals and Northeast Marcellus DFBeta for possible highly influential observations. The criteria used for the inclusion of an observation in Table 4 was any county with a studentized residual more than 2.5 or less than -2.5 or with a DFBeta less than -0.25 or greater than 0.25. Three counties in the Northeast Marcellus region, Potter, Susquehanna, and Union represent three possible influential observations. Susquehanna County is by far the most influential county within the Northeast Marcellus region based on the DFBeta value. Its relatively large, negative DFBeta value indicates that its presence in the model has a large influence on the overall negative Northeast Marcellus coefficient. In other words, if it were removed from the analysis, the value of the Northeast Marcellus coefficient would become less negative. The same can be said for Potter County, which is also in the Northeast Marcellus region. Although Pike County is not in the Northeast Marcellus region, its relatively large, positive PNMEE has the same effect on the Northeast Marcellus coefficient as Potter and Susquehanna counties, as evidenced by its negative Northeast Marcellus DFBeta value. In

Table 4. Possible Influential Observations

County	Northeast Marcellus Region?	PNMEE	Predicted PNMEE Value	Residual	Studentized Residual	Northeast Marcellus DFBeta
Centre	No	-5.63	-3.63	-2.00	-1.673	0.369
Juniata	No	-7.17	-2.09	-5.08	-3.361	0.415
Pike	No	5.92	2.38	3.54	3.098	-0.633
Potter	Yes	-3.60	-2.18	-1.42	-1.019	-0.367
Susquehanna	Yes	-7.28	-3.51	-3.77	-2.518	-0.863
Union	Yes	-4.26	-6.37	2.10	1.529	0.285

other words, because net migration was dramatically overestimated for Pike County, a county not in the Northeast Marcellus region, this contributes to the general net migration underestimation effect of being located the Northeast Marcellus region. The remaining three influential counties, Centre, Juniata, and Union, all have the opposite effect on the Northeast Marcellus coefficient as the previously mentioned counties. Centre and Juniata counties have this effect, because they have large, negative PNMEEs and are not in the Northeast Marcellus region. If these two counties were removed from the analysis, it would make the Northeast Marcellus coefficient more negative, as reflected by their positive Northeast Marcellus DFBeta values. Although Union County has a negative PNMEE and it is in the Northeast Marcellus region, its PNMEE predicted by the model was more negative than its actual PNMEE, indicated by its positive residual value. Therefore, it does not have the same effect on the Northeast Marcellus coefficient as Potter and Susquehanna counties do, even though it has a negative PNMEE like those counties. In other words, if it were removed from the analysis, the Northeast Marcellus coefficient would become more negative. Overall it does appear there are several counties within and outside the Northeast Marcellus region which could be dramatically affecting the multivariate regression results shown in Table 3. Therefore, it is important to determine whether or not these counties are creating the entire effect the Northeast Marcellus region has on the PNMEE.

Robust Regression Results

To take into account the fact that several counties appear to be heavily influencing the OLS regression results, a robust regression analysis is presented here. This method assigns weights to each observation in an attempt to minimize the impact highly influential observations have on the overall results. The higher the influence an observation potentially has on the OLS regression model, the lower the weight given to that observation in the robust regression model. Table 5 shows a comparison of the Model 4 OLS regression results and the results of the robust regression model. The robust results show that the effect of the Northeast Marcellus region on the PNMEE decreased in magnitude, but it is still negative and statistically significant, indicating the effect of Marcellus Shale drilling activity on net migration underestimation remained even after highly influential counties were taken into account. This robust regression result shows that the overall Northeast Marcellus coefficient result produced by the OLS regression model was not created entirely by a few highly influential counties. There is in fact a larger trend of net migration underestimation due to Marcellus Shale drilling activity across counties within the region.

When using a robust regression analysis, it is important to keep in mind that the standard errors of all independent variable coefficients will decrease. This is because observations are weighted based on their likelihood of high influence, giving lower weights to observations which are outliers either on the dependent variable or on one or more

Table 5. OLS vs. Robust Regression Results (N = 66)		
	OLS	Robust
Marcellus Drilling Measures		
Northeast Marcellus Region	-2.251**	-1.471*
	(.0022)	(.0122)
Southwest Marcellus Region	-0.197	0.209
	(.8361)	(.7929)
Rurality/Amenity Measures		
Non-Metropolitan Adjacent[a]	-1.526*	-0.549
	(.0314)	(.3429)
Non-Metropolitan Non-Adjacent[a]	-0.912	-0.584
	(.3610)	(.4821)
Natural Amenity Scale[b]	0.303	0.197
	(.4009)	(.5119)
Demographic Characteristics (year 2000)[c]		
Percent Hispanic	-0.067	0.003
	(.6441)	(.9786)
Percent Old (Age 65 and over)	0.053	0.001
	(.7122)	(.9950)
Percent Young (Age 0 to 14)	0.392+	0.450*
	(.0916)	(.0188)
Percent Single	-0.087	0.089
	(.5697)	(.4871)
Percent H.S. Education or Less	-0.028	-0.044
	(.6076)	(.3312)
Economic Characteristics (year 2000)[c]		
Homeownership Percentage	0.055	0.065
	(.6038)	(.4578)
Percent in Poverty	0.066	0.277
	(.7941)	(.1913)
Annual Rent/Wage Ratio	-0.007	-0.557*
	(.9792)	(.0200)
Unemployment Rate	0.657*	0.242
	(.0214)	(.2978)
Constant	-11.026	-10.617
	(.4597)	(.3930)
R-squared	0.477	0.361
Adjusted R-squared	0.334	N/A

+ significant p<0.10; * significant p<0.05; ** significant p<0.01; *** significant p<0.001

Coefficient p-values shown in parentheses below coefficient value

Sources: a - Office of Management and Budget (2003); b - Economic Research Service (1999); c -U.S. Census Bureau (2000)

Note: Forest County excluded from analysis

independent variables. In turn, this reduces the variation and the standard errors of the independent variable coefficients. This can increase the statistical significance level of coefficients between OLS and robust analyses without dramatically changing the actual magnitude of the coefficient. This is what occurs with the percent young coefficient. The coefficient magnitude only changed from 0.392 to 0.450 between models, but the p-value decreased from 0.09 to 0.02. This similar magnitude suggests that the effect of the variable on the PNMEE is roughly the same between OLS and robust models. Interestingly, the non-metropolitan adjacent coefficient is no longer significant in the robust model. The change in the magnitude of this coefficient is driven by several metropolitan counties having large positive PNMEEs (Pike and Fayette) and several non-metropolitan adjacent counties having large negative PNMEEs (Monroe and Susquehanna). All four counties are down-weighted in the robust model, reducing the non-metropolitan adjacent coefficient. The change in the annual rent/wage ratio coefficient value and statistical significance between the OLS and robust models is also dramatic. This is largely due to the high influence of Pike County on the OLS model. In a similar influential observation analysis performed for this specific independent variable that was performed for the Northeast Marcellus region variable, the Pike County DFBeta for this variable was 2.12. This is an extremely large DFBeta, which is why there is such a large change in the rent/wage ratio coefficient between the OLS and robust models. With Pike County accounted for in the robust model, the rent/wage ratio is shown to have a negative effect on the PNMEE, indicating the higher the rent/wage ratio, the more net migration was underestimated for the decade. This could represent the higher demand for housing reflected by higher rent and housing costs in areas which experienced high in-migration prior to the beginning of the decade, then continued to experience it during the decade.

The most important finding from this robust regression model is that the coefficient of the main independent variable, Marcellus Shale drilling activity, remains significant and negative after accounting for highly influential counties. Therefore, the substantive conclusions that can be drawn from this are the same as in Model 4, and in-migration was, in fact, underestimated by a significant amount for counties within the Northeast Marcellus region of the state. This is an important finding in terms of quantifying population change due to the drilling activity for counties within this region of the state, as well as with respect to what counties and regions poised to experience drilling activity in the future can expect in terms of population change associated with the drilling activity.

Chapter 5

Discussion

Main Findings

The most important finding of this research is that, while the U.S. Census Bureau tended to underestimate county-level net migration in general for Pennsylvania counties during the 2000s, underestimation of net migration was relatively greater for counties in the Northeast Marcellus region of Pennsylvania. This confirmed both of the initial hypotheses of the study. A regional measure was the most appropriate for capturing migration processes with respect to Marcellus Shale drilling activity, and this type of measure captured a general trend of net migration underestimation for counties in the Northeast Marcellus region. The final model in the multivariate regression analysis showed that net migration was underestimated by more than an additional two percent in counties in the Northeast Marcellus region. Considering the past boomtown literature documenting the possibility of a rapid population influx accompanying natural resource extraction, it is reasonable to assume that this net migration underestimation was caused by greater than expected in-migration to the region due to the economic development surrounding the drilling activity, rather than slower than expected out-migration. The same result was not found for the Southwest Marcellus region. This could be due to several reasons. First, the average population for counties in the Southwest Marcellus region is much greater than that of counties in the Northeast Marcellus region. Therefore, a similar size population influx to counties in the Southwest Marcellus region would cause a smaller percent change in population than it would for counties in the Northeast Marcellus region. Second, the greater transportation infrastructure provided by the Pittsburgh metropolitan area in the southwestern part of the state presents a greater likelihood that workers are not living as close to well sites as they are in the Northeast Marcellus region. The enhanced transportation infrastructure better enables workers to travel greater distances in less time to get to work. If this is occurring, it makes it less likely specific counties in the region have experienced concentrated population influxes.

I also found that this relationship between the Northeast Marcellus region and in-migration underestimation remains after taking into account several counties within and outside the Northeast Marcellus region which highly influence the results. In-migration in the Northeast Marcellus counties of Susquehanna, Clinton, and Potter was underestimated by much more than most counties in the state, but the robust regression analysis confirmed a broader trend of net migration underestimation across all counties within the Northeast Marcellus region.

Limitations of study

The main limitation of this study is that it does not directly assess the number of people moving to the Northeast Marcellus region of Pennsylvania due to the Marcellus Shale drilling activity. The methodology used here is unable to distinguish a Marcellus-related migrant from a non-Marcellus related one. Therefore, the study is unable to determine how many Marcellus-related migrants are captured in the Census Bureau’s net migration estimates and how many are not. Consequently, it is also unable to distinguish between net migration estimation error caused by Marcellus-related migrants and error caused by non-Marcellus-related migrants. Although these limitations exist, the method used in this study does provide a more nuanced look at migration to and from Pennsylvania counties than simply looking at the change in the Census population count from 2000 to 2010.

Policy Implications

The results of this study showed counties that experienced or were near counties that experienced extensive Marcellus Shale drilling activity had greater than expected levels of in-migration during the 2000s compared to what was estimated by the Census Bureau. This pattern may continue throughout the next decade. The main policy implication drawn from these results is that the Pennsylvania state government as well as local municipality governments should take this into account when assessing Census Bureau county level population estimates going forward. This will help ensure that funds dedicated toward government programs designed to provide services such as healthcare and education to individuals are allocated at the levels necessary to administer the programs properly. The results are also important for areas which have not experienced drilling yet but expect to in the future. This may help them better prepare for the possibility of increased population levels brought on by the economic activity surrounding the drilling.

Future Research

Future research should continue to track population change in areas of Pennsylvania which continue to experience heavy amounts of Marcellus Shale drilling activity, or begin to experience drilling activity in the future. It will be important for this research to take into account that the Census Bureau’s population estimates may not fully capture migrants moving to these areas. This is due to potential weaknesses in the Bureau’s net migration estimate methodology at capturing the movement of Marcellus-related migrants.

Another potentially fruitful line of research would be to conduct a spatial analysis of net migration across Pennsylvania counties. The results of this study showed that drilling activity in one county may affect net migration in nearby counties. A spatial regression model would explicitly take into account how the characteristics of one county can affect net migration in neighboring counties.

Overall, future research on population change due to the Marcellus Shale drilling activity should examine in more detail how the population characteristics have changed in Pennsylvania counties within the Northeast Marcellus region of Pennsylvania. This could either be a look at only counties within the region, or a comparison to other counties in the state. Potential data to look at include changes in county age structure, unemployment rates, and industry specific occupation levels. A good place to start this research would be examining detailed 2000 and 2010 Census data to determine how the population changed over the decade. Microdata from the American Community Survey (ACS) may also provide useful insights into how the populations of these counties have changed over time.

Appendix – Additional Tables

Table 6. Full Correlation Matrix

	PNMEE	Northeast	Southwest	Natural Amenity Scale	Non-Metropolitan Adjacent	Non-Metropolitan Non-Adjacent	Pct Hispanic
Northeast	-0.4641						
Southwest	0.2524	-0.2321					
Natural Amenity Scale	0.0680	0.0875	0.2621				
Non-Metropolitan Adjacent	-0.3083	0.2295	-0.1371	0.0721			
Non-Metropolitan Non-Adjacent	0.0181	-0.0352	-0.1414	-0.2254	-0.2715		
Pct Hispanic	0.1190	-0.1869	-0.2444	0.0130	-0.2451	-0.1675	
Pct Old	0.0063	0.1745	0.2417	0.1843	0.0762	0.1313	-0.1675
Pct Young	0.2777	-0.2163	-0.2719	-0.1400	-0.1766	0.0865	0.3761
Pct Single	0.0715	-0.0535	0.0774	0.0773	-0.2517	-0.0587	0.1681
Pct HS or Less	-0.1363	0.0880	0.0093	0.0035	0.4749	0.2140	-0.3220
Homeownership Pct	0.0975	-0.0531	0.0134	0.0295	0.2863	0.1312	-0.3132
Pct in Poverty	0.1217	0.0518	0.2573	-0.0067	0.1861	0.0168	-0.1223
Annual Rent/ Wage Ratio	0.1436	0.0436	-0.0966	0.1340	-0.1757	-0.1297	0.4570
Unemployment Rate	0.1971	0.1115	0.3011	0.1029	0.1679	0.0069	-0.1133

Table 6. Full Correlation Matrix (Continued)							
	Pct Old	Pct Young	Pct Single	Pct HS or Less	Homeownership Pct	Pct in Poverty	Rent/Wage Ratio
Northeast							
Southwest							
Natural Amenity Scale							
Non-Metropolitan Adjacent							
Non-Metropolitan Non-Adjacent							
Pct Hispanic							
Pct Old							
Pct Young	-0.4625						
Pct Single	0.0467	-0.2909					
Pct HS or Less	0.3848	-0.2138	-0.1703				
Homeownership Pct	0.1762	0.1936	-0.7853	0.3158			
Pct in Poverty	0.2551	-0.1696	0.5445	0.4805	-0.2511		
Annual Rent/ Wage Ratio	-0.2352	0.0985	0.6102	-0.2972	-0.4703	0.3931	
Unemployment Rate	0.3764	-0.3665	0.5191	0.2496	-0.1928	0.7465	0.3959

Table 7. County Percent Net Migration Estimate Error (PNMEE)

County	PNMEE (%)
Adams County	1.28
Allegheny County	-0.08
Armstrong County	-1.86
Beaver County	0.75
Bedford County	-0.57
Berks County	-1.10
Blair County	-0.90
Bradford County	-2.24
Bucks County	0.08
Butler County	0.51
Cambria County	-0.21
Cameron County	1.66
Carbon County	-2.72
Centre County	-5.63
Chester County	0.27
Clarion County	-1.46
Clearfield County	0.85
Clinton County	-6.40
Columbia County	-3.45
Crawford County	-0.86
Cumberland County	-0.86
Dauphin County	-3.37
Delaware County	1.05
Elk County	-0.53
Erie County	-0.27
Fayette County	3.10
Franklin County	-2.52
Fulton County	0.99
Greene County	0.95
Huntingdon County	-0.46
Indiana County	-1.65
Jefferson County	-0.71
Juniata County	-7.17
Lackawanna County	-2.97
Lancaster County	-1.90
Lawrence County	-1.54

*Note: Forest County excluded from analysis

Table 7. County Percent Net Migration Estimate Error (PNMEE) (Continued)

County	PNMEE (%)
Lebanon County	-2.39
Lehigh County	-1.66
Luzerne County	-2.94
Lycoming County	0.32
McKean County	-0.87
Mercer County	-0.70
Mifflin County	-1.32
Monroe County	-2.91
Montgomery County	-1.38
Montour County	-2.14
Northampton County	0.35
Northumberland County	-3.03
Perry County	-0.81
Philadelphia County	2.36
Pike County	5.92
Potter County	-3.60
Schuylkill County	-0.85
Snyder County	-3.28
Somerset County	-0.94
Sullivan County	-4.85
Susquehanna County	-7.28
Tioga County	-2.14
Union County	-4.26
Venango County	-1.58
Warren County	-3.12
Washington County	-0.53
Wayne County	-1.34
Westmoreland County	-1.15
Wyoming County	-1.60
York County	-1.36

*Note: Forest County excluded from analysis

Table 8. Wells Drilled and Permits Issued by County as of January 1, 2010

County	Wells Drilled	Permits Issued
Adams County	0	0
Allegheny County	4	7
Armstrong County	30	61
Beaver County	0	6
Bedford County	0	0
Berks County	0	0
Blair County	0	2
Bradford County	185	486
Bucks County	0	0
Butler County	35	73
Cambria County	2	6
Cameron County	5	8
Carbon County	0	0
Centre County	13	55
Chester County	0	0
Clarion County	8	14
Clearfield County	29	84
Clinton County	12	49
Columbia County	0	2
Crawford County	0	0
Cumberland County	0	0
Dauphin County	0	0
Delaware County	0	0
Elk County	19	48
Erie County	0	0
Fayette County	85	127
Franklin County	0	0
Fulton County	0	0
Greene County	182	242
Huntingdon County	0	0
Indiana County	12	33
Jefferson County	5	7
Juniata County	0	0
Lackawanna County	1	28
Lancaster County	0	0
Lawrence County	0	0

Source: PA Department of Environmental Protection Office of Oil and Gas Management (2012)
*Note: Forest County excluded from analysis

Table 8. Wells Drilled and Permits Issued by County as of January 1, 2010 (Cont.)

County	Wells Drilled	Permits Issued
Lebanon County	0	0
Lehigh County	0	0
Luzerne County	0	1
Lycoming County	39	172
McKean County	13	26
Mercer County	0	0
Mifflin County	0	0
Monroe County	0	0
Montgomery County	0	0
Montour County	0	0
Northampton County	0	0
Northumberland County	0	0
Perry County	0	0
Philadelphia County	0	0
Pike County	0	0
Potter County	22	51
Schuylkill County	0	0
Snyder County	0	0
Somerset County	7	21
Sullivan County	0	1
Susquehanna County	124	226
Tioga County	139	328
Union County	0	0
Venango County	0	0
Warren County	2	6
Washington County	229	398
Wayne County	1	3
Westmoreland County	77	119
Wyoming County	2	11
York County	0	0

Source: PA Department of Environment Protection Office of Oil and Gas Management (2012)
*Note: Forest County excluded from analysis

References

Albrecht, S. L. (1978). Socio-cultural factors and energy resource development in rural areas in the West. *Journal of Environmental Management, 7*(1), 90-93.

Barnes, T. (2004). State announces plan to mothball 'Western Pen'. *Pittsburgh Post-Gazette*. Retrieved from http://www.postgazette.com/stories/local/uncategorized/state-announces-plan-to-mothball-western-pen-528273/.

Belsley, D. A., Kuh, E., & Welsh, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons, Inc.

Brasier, K. J., Filteau, M. R., McLaughlin, D. K., Jacquet, J., Stedman, R. C., Kelsey, T. W., & Goetz, S. J. (2011). Residents' Perceptions of Community and Environmental Impacts from Development of Natural Gas in the Marcellus Shale: A Comparison of Pennsylvania and New York Cases. *Journal of Rural Social Sciences, 26*(1), 32-61.

Briem, C. (2011). Migration *Trends in the Pittsburgh Region: Update Through 2010*. Pittsburgh, PA: University Center for Social and Urban Research. Retrieved from http://www.ucsur.pitt.edu/files/frp/MigrationReport2011.pdf

Brown, R. B., Geertsen, H. R., & Krannich, R. S. (1989). Community Satisfaction and Social Integration in a Boomtown: A Longitudinal Analysis. *Rural Sociology, 54*(4), 568-586

Brown, R. B., Dorins, S. F., & Krannich, R. S. (2005). The Boom-Bust-Recovery Cycle: Dynamics of Change in Community Satisfaction and Social Integration in Delta, Utah. *Rural Sociology, 70*(1), 28-49.

Carr, P. & Kefalas, M. (2009). *Hollowing out the Middle: The Rural Brain Drain and What It Means for America.* New York: Beacon Press.

Christopherson, S. & Rightor, N. (2011). How Should We Think About the Economic Consequences of Shale Gas Drilling? (Working Paper Series). Retrieved March 27, 2012 from http://www.greenchoices.cornell.edu/downloads/development/marcellus/Marcellus_SC_NR.pdf

Cortese, C., & Jones, B. (1977). The sociological analysis of boomtowns. *Western Sociological Review, 8*(1), 76-90.

Economic Research Service (1999). *Natural Amenities Scale*. Retrieved from http://www.ers.usda.gov/data-products/natural-amenities-scale.aspx

Findley, S. E. (1987). An Interactive Contextual Model of Migration in Ilocos Norte, the Philippines. *Demography, 24*(2), 163-190.

Frank, H. (2012). Coming or going? City migrants to Poconos trail off. *Pocono Record*. Retrieved from http://www.poconorecord.com/apps/pbcs.dll/article?AID=/20120603/NEWS/206030325

Freudenburg, W. (1986). Social Impact Assessment. *Annual Review of Sociology, 12*(1), 451-478.

Galston, W. A. & Baehler K. J. (1995). *Rural Development in the United States: Connecting Theory, Practice, and Possibilities.* Washington, DC: Island Press.

Gilmore, J. S. (1976). Boom towns may hinder energy resource development. *Science, 191*(4227), 535-540.

Gilmore, J. S. & Duff, M. K. (1975). *Boom Town Growth Management: A Case Study in Rock Springs – Green River, Wyoming.* Boulder, CO: Westview Press.

Glaeser, E. & Shapiro, L. (2001). Is there a new urbanism? The growth of US cities in the 1990s (Working paper series No. W8357). Retrieved March 28, 2012 from National Bureau of Economic Research website: http://www.nber.org/papers/w8357.

Greenwood, M. J. (1997). Internal Migration in Development Countries. In M. R. Rosenberg and O. Stark (Eds.), *Handbook of population and family economics* (pp. 647-720). Amsterdam: North Holland.

Greenwood, M. J., Hunt, G., Rickman, D. S., & Treyz, G. I. (1991). Migration, Regional Equilibrium, and the Estimation of Compensating Differentials. *American Economic Review, 81*(5), 1382-1390.

IssuesPA (2003). Brain Drain Versus Brain Gain: Pennsylvanians on the Move? Yes, but. Retrieved from http://issuespa.org/content/brain-drain-versus-brain-gain-pennsylvanians-move-yes

Jacquet, J. (2009). *Energy Boomtowns & Natural Gas: Implications for Marcellus Shale Local Governments & Rural Communities* (NERCRD Rural Development Paper No. 43). University Park, PA: Northeast Regional Center for Rural Development. Retrieved from http://nercrd.psu.edu/publications/rdppapers/rdp43.pdf

Lee, S. (1966). A Theory of Migration. *Demography, 3*(1), 47-57.

Marcellus Shale Education and Training Center (2009). Marcellus Shale Workforce Needs Assessment. Retrieved from http://www.msetc.org/docs/NeedsAssessmentwithcover.pdf

Marcellus Shale Education and Training Center (2011). Marcellus Shale Workforce Needs Assessment. Retrieved from http://www.msetc.org/docs/PennsylvaniaStatewideWorkforceAssessmentv1_Final.pdf

Massey, D. S. (1990). Social Structure, Household Strategies, and the Cumulative Causation of Migration. *Population Index, 56*(1), 3-26.

Massey, D. S. (2008). *New Faces in New Places: The Changing Geography in American Immigration*. New York: Russell Sage Foundation.

McGranahan, D. (1999). *Natural Amenities Drive Rural Population Change*. Agricultural Economic Report No. 781. U.S. Department of Agriculture, Economic Research Service, Washington, DC.

McGranahan, D. (2008). Landscape Influence on Recent Rural Migration in the U.S. *Landscape and Urban Planning, 85*, 228-240.

Mieszkowski, P., & Mills, E. S. (1993). The Causes of Metropolitan Suburbanization. *The Journal of Economic Perspectives, 7*(3), 135-147.

New Jersey Star-Ledger (2011). N.J. residents are leaving for Pennsylvania in increasingly large numbers. *New Jersey Real-Time News*. Retrieved May 25, 2012 from http://www.nj.com/news/index.ssf/2011/12/nj_residents_are_leaving_for_p.html

Office of Management and Budget (2003). OMB Bulletin No. 03-04. Retrieved from http://www.whitehouse.gov/sites/default/files/omb/bulletins/b03-04_attach.pdf

Oropesa, R. S. & Jensen, L. (2010). Dominican Immigrants and Discrimination in a New Destination: The Case of Reading, Pennsylvania. *City & Community, 9*(3), 274-298.

Pennsylvania Department of Environmental Protection (2012). Office of Oil and Gas Management. Retrieved March 6, 2012 from http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?/Oil_Gas/Wells_Drilled_By_County

Pennsylvania Department of Health (2012). Pennsylvania Vitals Statistics Database. Retrieved from http://www.portal.state.pa.us/portal/server.pt?open=514&objID=596032&mode=2

Pennsylvania Economy League (2008). *The Economic Impact of the Oil and Gas Industry in Pennsylvania*. Retrieved from http://alleghenyconference.org/PEL/PDFs/EconomicImpactOilGasInPA1108.pdf

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy, 70*(5), 80-93.

Smith, S. K., Tayman, J., & Swanson, D. A. (2001). *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic Publishers.

The Perryman Group. (2008). *Drilling for Dollars: An Assessment of the Ongoing and Expanding Economic Impact of Activity in the Barnet Shale on Forth Worth and Surrounding Area.* Retrieved from http://www.bseec.org/images/summaryreport.pdf.

U.S. Census Bureau. (2000). Census 2000 Summary File 1 & 3, Detailed Tables.

U.S. Census Bureau. (2009). Methodology for the State and County Total Resident Population Estimates (Vintage 2009). Retrieved from http://www.census.gov/popest/methodology/2009-st-co-meth.pdf

U.S. Census Bureau. (2010). Census 2010 Summary File 1.

U.S. Census Bureau. (2011). 2010 County Total Population Estimates. Retrieved from http://www.census.gov/popest/

U.S. Census Bureau, American Factfinder (2012). Profile of General Population and Housing Characteristics. Retrieved from http://factfinder2.census.gov/

Ward, M. M., Kelsey, T. W., Brundage, T., Ladlee, J., Lorson, J., Michael, L., & Murphy, T. (2011). *Local Business Impacts of Marcellus Shale Development: The Experience in Bradford and Washington Counties, 2010* [Fact Sheet]. Retrieved from pubs.cas.psu.edu/freepubs/pdfs/ee0005.pdf

Wilkinson, K. P., Thompson, J. G., Reynolds, Jr., R. R., & Ostresh, L. M. (1982). Local Social Disruption and Western Energy Development: A Critical Review. *The Pacific Sociological Review, 25*(3), 275-296.

Williams, R. (2008). *Multicollinearity.* University of Notre Dame.

Zuñiga, V. & Hernández-León, R. (2005). *New Destinations of Mexican Immigration in the United States: Community Formation, Local Response and Inter-Group Relations*. New York: Russell Sage Foundation.