An Analysis of Gross In- and Out-Migration Flows for Pennsylvania Counties: A closer look at the relative importance of labor market and amenity differentials between origin and destination counties

Martin Shields
Assistant Professor of Agricultural and Regional Economics
mshields@psu.edu

Stephan Goetz
Professor of Agricultural and Regional Economics and
Director, Northeast Regional Center for Rural Development

Latika Bharadwaj
Graduate Research Assistant

The Pennsylvania State University

V 1.0: Please do not cite

Paper prepared for presentation at the annual meetings of the Western Regional Science Association Meetings, Maui, Hawaii, February 26-28 2004.

Abstract

Population growth in Pennsylvania has been relatively stagnant for the past 50 years. This has had important consequences, not the least of which is the loss of 2 congressional seats after the 2000 US Census. In this paper we study gross migration flows in and out of Pennsylvania. Drawing on previous work, we offer a theoretical framework that suggests amenities and labor market conditions play an important part in individual migration decisions. We then turn our attention to empirical analysis of these factors' influence on county-to-county migration patterns, using data from the 2000 Census. While migration determinants have been oft-addressed in the literature, there are a number of critical specification issues that have received scant attention in econometric studies. The consequence is that our understanding of the relative importance of various hypothesized factors may not be complete. In this paper, we offer empirical evidence from new parametric and non-parametric models that account for both "preference bundling" and zero-migration flows.

1. Problem

Pennsylvania's population growth has long been stagnant. Whereas the state's population (12.2 million) has grown by only 4 percent since 1970, the US has grown by 38 percent. This, however, is consistent with long term trends. Since 1950, Pennsylvania's population has increased by just 17 percent, while the US has grown by 86 percent. This has a number of important implications, not the least of which is the fact that Pennsylvania lost two congressional seats from 1990-2000.

One of the most important causes of this decline is a substantial net out-migration from the state. According to the US Census Bureau, Pennsylvania ranked 5th among all states in net out-migration for the period 1995-2000. From 1995-2000, 669,000 people moved into the state, while 800,000 left the state, resulting in a net loss of 131,000 people. The state's in-migration rate of 58 people per 1000 residents ranked 6th lowest in the nation.

In this paper we take a closer look at gross out-migration and in-migration for Pennsylvania for the period 1995-2000. Specifically, we examine the Census Bureau's county-to-county migration data in a family of econometric models in an attempt to better understand the relative importance of various factors hypothesized to affect household location decisions, particularly amenities and employment opportunities. Overall, we find that out-migration from Pennsylvania is driven more by relatively poor economic performance, especially in the state's rural areas, than it is by differences in natural amenities. This is an important finding in that it suggests that job creation remains an important strategy for reversing out-migration patterns.

This paper is divided into four remaining sections. In the next section we briefly review the migration literature, emphasizing recent findings on amenities and employment. We then offer descriptive statistics on the role of three relative variables in out- and immigration: amenities, unemployment and employment growth. In section four develop two econometric models, one of gross in-migration and one of gross out-migration, each examining the three relative variables as well as several other local characteristics. In the final section we offer preliminary conclusions and directions for future research.

2. Literature

Regional migration theory is rooted in the expected utility framework of microeconomic theory. In a household location framework, the problem is one of choosing the location (r) that maximizes the household's utility subject to regional prices and income:

$$v_r(\mathbf{p}, y; \theta) = \max u_r(\mathbf{x}; \theta) \quad st\ \mathbf{p}_r\mathbf{x} = y_r$$

The function $v(\mathbf{p},y;\theta)$ is the indirect utility function, which gives the maximum utility attainable at given local prices (p) and income (y). Accounting for moving costs, if a household expects to attain greater utility in a different region, there is a high likelihood it will move. Simply put, the probability that a household will move to region j from region k:

$$\Pr(\textit{migrate}) = \Pr(E(v_j(\mathbf{p},y;\theta)) \succ E(v_k(\mathbf{p},y;\theta)))$$

where $E(v_r(\mathbf{p},y;\theta))$ is the expected indirect utility of the household. Given this simple framework, the relevant migration factors are those that affect the consumption bundle (e.g., regional prices and wages) as well as taste and preference arguments of the household utility function (e.g., expected employment stability and local amenities).

Nearly all studies that investigate regional migration focus on how regional differences in expected income and amenities determine migration behavior. Equilibrium in such a system is characterized in terms of zero net migration, where expected income differentials compensate for differences in local amenities (Graves, 1979; Molho, 1995). Greenwood (1975; 1985), Clark (1982) and Mueller (1982) each give detailed histories of the theoretical and empirical developments in migration theory.

Empirical Evidence on the Income and Employment Determinants of Migration

The role of expected income in the household migration decision depends on differential economic opportunities that consist of two parts—the expected wage and the probability of receiving that wage (Treyz et al, 1993). Regarding expected wages, migration theory focuses on earnings differentials across regions. Appealing to the neoclassical theory of factor mobility, it is argued that labor responds to wage rate differentials by moving until a new equilibrium is reached (e.g., Borts and Stein, 1964; Smith, 1974; 1975). While regional wage differentials have (haltingly) declined over time (Dickie and Gerking, 1989), empirical evidence suggests they have not yet fully equilibrated (Eberts and Schweitzer, 1994). Still, it is clear that migration has led to a decline in inter-regional wage variations, though the process has been fairly lengthy (Greenwood and Hunt, 1984).

One difficulty with these early migration studies is that they assume full employment, with wages changing until the market clears (Isserman et al, 1986). Absent these conditions it is important to examine the probability that a household will receive the regional wage when investigating expected income. Accordingly, the second component of relative economic opportunity is the probability of getting a job.

While the basic concept is simple, determining the probability of employment is tremendously difficult. As noted by Isserman et al (1986), information is needed

on job vacancies and the number of people seeking jobs (including discouraged workers who would reenter the job market should a job become available); data not generally available at any level. Because of these difficulties, a number of proxies for opportunity have been examined as determinants of migration. These opportunity measures include population (Greenwood and Sweetland, 1972), the employment-to-population ratio (Dahlberg and Holmlund, 1978), and the number of new hirings (Fields, 1976; 1979).

The most prevalent measures of opportunity, though, are employment and employment growth (e.g., Muth, 1971; Duffy and Greenwood, 1980; Plaut, 1982; Bartik, 1993). Muth (1971) provides an early investigation into the importance of job opportunities in explaining net migration. Using data for urban areas in the 1950s, Muth finds that both jobs and wages are important in the household migration decision. Treyz et al (1993) provide recent support for the importance of relative regional wages and employment opportunities on migration in the US.

When examining the effect of employment on migration it is necessary to also consider local unemployment. In the household's expected utility decision, areas with high relative local unemployment offer a lower expected probability of employment, leading to lower expected earnings. Thus, regions with high unemployment are unlikely to attract in-migrants, while current residents that are currently unemployed may move elsewhere. Graves (1979) provides evidence that in-migration is minimal in areas with relatively high local unemployment rates.

Empirical Evidence on the Location Specific Determinants of Migration

A second aspect considered in migration studies is the importance of location-specific amenity factors such as weather and public services. The theoretical argument is that amenities provide non-earnings-based utility to households, drawing new residents to the region (Graves, 1979; Graves and Linneman, 1979). In this framework, equilibrium is achieved as people move into an amenity rich region, increasing the local labor supply and subsequently reducing wages to the point that regional wage differentials exactly compensate for local amenities.

Graves (1979) provides one of the earliest examinations into the importance of weather in household location decisions. Examining net population migration in the 1960s, Graves demonstrates that when income levels and unemployment rates are taken into account, certain climatological amenity variables are important. These variables include the influence of heating and cooling-degree days, annual temperature variance, relative humidity and wind speed. Other researchers have investigated the importance of local public services on the migration decision. In a survey of migration and the local level of local public services Charney (1993) concludes that higher levels of public expenditures on a number of goods serve as an incentive for migration. Of course, higher

expenditures could also mean higher taxes, a factor that can discourage the household location decision (Yinger, 1982).

In sum, the theoretical and empirical evidence suggests that expected earnings, regional employment opportunities and locational amenities influence migration. From this, the basic migration modeled looks like:

$$migration = m(empgrow, relwage, relunemp, \mathbf{A})$$

where *empgrow* is local employment growth, *relwage* is the relative average local wage, *relunemp* is relative local unemployment, and **A** is a vector of location-specific amenities.

3. Some Basic Facts about Pennsylvania In- and Out-Migration

This section under development: See the early overheads for the gist of this discussion.

- 2000 state population: 12,281,054
  - 6th most populous state (4.4 percent of US in 2000)
- 1990-2000 Growth: 3.4 percent (ranked 48th)
  - US growth 1990-2000: 13.1 percent
- 15.3 percent of residents live in 'rural' counties

- Pennsylvania Has Had Relatively Slow Job Growth
  - Since 1969 state employment has grown 33 percent while US has grown 84 percent
- Yet the States' Unemployment Rate Closely Mirrors the US

- Basic Migration Statistics
  - Pennsylvania net-migration rate -11.4 (37^{th} overall)
  - Pennsylvania in-migration rate 58.1 per 1000 1995 residents (46^{th} overall, US average 109.4, standard deviation 46.0)
  - Pennsylvania out-migration rate 69.5 per 1000 1995 residents (47^{th} overall, US Average 106.3, 40.1)
- *This suggests it may not be so much people leaving—out-migration is low—rather it is that no one is moving here. Why not?*

- About 37 percent of Pennsylvania out-migrants moved to contiguous states.
  - The top overall destination was Florida (12 percent)
- About 50 percent of in-migrants moved from contiguous states
  - Primarily New York (17 percent), New Jersey (17 percent) and Maryland (8 percent).

4. Conditional Expectations

As the recent literature suggests, regional scientists continue to struggle with sorting out the relative importance of amenities and employment opportunities in household migration decisions. In this section, we provide nonparametric analysis of select relative conditions in order to provide an initial investigation into these factors. Specifically, we examine the proportion of movers who migrated to "better" places, considering four separate regional factors.

The first step is to create a county specific data set for each US county. This dataset contains information on amenities, employment growth and unemployment rates. We then merge this data set with the county-to-county migration data from the 2000 census, which tracks the number of people moving between 1995 and 2000. Here, then we have several attributes of the origin and destination counties and a count of the number of people who moved 1) from one county in Pennsylvania to each county outside of the state; and 2) from a county outside of the state to a county in Pennsylvania.

The next step is to examine the propensity for members of the population to seek to improve their lot with respect to each of the hypothesized factors. Specifically, we examine the frequency of the two possible outcomes:

$IMPROVE_{i,j} = 1$ if $CONDITION_i > CONDITION_j$; else
$IMPROVE_{i,j} = 0$

Here, *CONDITION* refers to characteristics of region *i* and *j*. The variable *IMPROVE* takes on a value of one if the destination county is "better" than the origin county with respect to a particular indicator. For example, if a destination county has greater amenities than the origin county, then the move is said to be improving.

In the third step, we sum the number of people who make an improving move and divide it by the total number of people who move to determine the proportion of the population that potentially seeks a particular attribute.

For the *CONDITION* variable we consider four attributes: unemployment rate, employment growth, amenities and housing values. We measure the labor market indicators at the commuting zone level,[1] and the amenities and housing

[1] In defining labor markets, we utilize the delineation of "commuting zones" forwarded by the Economic Research Service (ERS) of the USDA. In classifying 741 commuting zones (483 of which are non-metro; containing no MSAs), ERS researchers utilized a hierarchical cluster algorithm to examine the journey-to-work commuting flow data for counties from the 1990 Census. In a nutshell, the method analyzes the proportional frequency of inter-county commutes, and

value variables are measured at the county-level. Because the Census data shows that a large proportion of out-of-state moves are made to contiguous states, we also offer conditional rates based on contiguity to Pennsylvania.

The unemployment rate is for 1995, and is derived from the Bureau of Labor Statistics Local Area Unemployment Statistics (LAUS). Here, we sum up the number of unemployed workers in the commuting zone and divide it by the commuting zone labor force.

The employment growth variable is the average annual percent change in total employment from 1993-1995 for the commuting zone. This is drawn from Bureau of Economic Analysis Regional Economic Information System (BEA-REIS), with county totals once again aggregated to the commuting zone level.

The natural amenities data was obtained through the ERS (1999) and assigns a score based on environmental qualities that people are purported to prefer. The components of this measure that enhance a location's livability are: warm winter, winter sun, temperate summer, low summer humidity, topographic variation, and water area. These six variables were scored and combined to form a natural amenities scale, which is continuous and is used in the econometric specifications. From this scale, a discrete ranking system was computed with a range of 1 to 7, where 1 is a low natural amenity rank and 7 is a high rank.

The housing value data is the median housing value in the county in 1990 and is drawn from the 1990 US Census. For all variables, summary statistics and data definitions and sources are provided in Table 1.

Table 1. Summary Statistics for Select Variables (weighted by migration flow)

	mean	std dev
gross out-migration	11,841	20,367
County-to-county outflow	39.8	156.0
unemployment rate difference	-0.68	11.03
growth rate difference	1.22	10.44
amenity difference	0.50	
median home value difference	28.30	460.62

From this table, we see that the average outflow to another state from a Pennsylvania county was 11,841 people, while the average flow, when observed was about 40 people. We also see that the destination counties, on (weighted)

---

defines a commuting zone as having met a threshold of interaction. A map of all commuting zones is available at the ERS web site (www.ers.usda.gov/briefing/rurality/lmacz/czones.htm). Tolbert and Sizer (1994) provide a detailed description of the methodology.

average, had lower unemployment rates, higher 3-year average employment growth rates, better amenities and higher median home values.

The findings of this analysis are shown in Table 2. With respect to people leaving Pennsylvania, our simple bi-variate analysis suggests that relative labor market conditions may matter, especially for those not moving to adjacent states. Here, 87 percent of all out-migrants moving to non-adjacent states moved to places with higher 3-year employment growth rates than the labor markets they left, while 74 percent moved to labor markets with lower unemployment rates. These trends are similar for moves to adjacent states, though not as strong.

Table 2. Percent of Pennsylvania Migrants Making "Improving" Moves

	All States	Non-adjacent States
Out-migrants		
Unemployment Rate	64%	74%
Employment Growth	72%	87%
Amenities	41%	55%
Home values	34%	36%
In-migrants		
Unemployment Rate	57%	28%
Employment Growth	36%	19%
Amenities	21%	18%

Turning to amenity differentials, the results of the simple analysis are ambiguous, with slightly more than half (55 percent) of all out-migrants making 'amenity improving' moves. Finally, lower relative median housing values do not seem strongly related to out-migration decisions, with only 34 percent of all out-migrants moving to counties with lower housing values.

The results of a similar analysis for Pennsylvania in-migrants offer much less support for the hypotheses that regional differences matter. Of all in-migrants, only 36 percent moved into labor markets with faster employment growth; whereas 57 percent moved into labor markets with lower unemployment rates. Only 21 percent of all in-migrants made amenity improving moves.

Looking at moves into Pennsylvania from non-adjacent states, 18 percent were amenity improving, 19 percent were employment growth improving and 28 percent were unemployment rate improving.

Overall, then, we find preliminary evidence that people leaving Pennsylvania were acting fairly consistently with the hypotheses, whereas observed moved into Pennsylvania provided much less support.

5. Econometric Models

In this section we return to the theoretical migration framework described above, offering an econometric model of gross in- and out-migration flows for Pennsylvania. Specifically, we model the following:

> *number of migrants$_{i,j}$ = f(unemployment rates differences$_{i,j}$, employment growth rate differences$_{i,j}$, natural amenity differences$_{i,j}$, housing value differences$_{i,j}$, population$_j$, adjacency).*

Here, we have separate equations for the number of out-migrants from Pennsylvania county *i* to non-Pennsylvania county *j*, and the number of in-migrants to Pennsylvania county *i* from non-Pennsylvania county *j.*

Most measures look at the differences between a particular indicator in the destination county and the origin county. Population (1995) in the origin county is included to control for the pool of potential migrants. Adjacency is a dummy variable that takes a value of 1 if the destination/origin county is in a state contiguous to Pennsylvania. Its inclusion is designed to controls for "attachment" factors, such as proximity to family and friends.

For the out-migration equation we enlist two estimation procedures. First, we consider only those instances when there is an observed migration flow between two counties (i.e., migrants$_{i,j}$ > 0). For Pennsylvania, this criteria allows for 20,167 observations. The general model we estimate, then is:

$$y_{i,j} = x_{i,j}^{'}\beta_0 + \varepsilon_{i,j}, \quad \varepsilon_{i,j} \Big| x_{i,j} \sim N(0,\sigma_0^2)$$

where

$$y_{i,j} > 0$$

The model here would be the simple OLS model were it not for the restriction. However, the use of OLS in such a model would result in inefficient parameter estimates. In this case, then, truncated regression via a maximum likelihood estimator is the appropriate method, and we adopt this approach here. Hayashi (2003) provides the derivation of the maximum likelihood estimator.

Results of this estimation are provided in the second column of Table 3. Because of the large sample size, it is not surprising that all variables are statistically significant at the 1 percent level. The first finding is that 16 people move to a destination commuting zone with a one percentage point lower unemployment rate, *ceteris paribus*. The second finding is that a one percentage point higher average employment growth rate in a destination county leads to 62 *fewer* movers to that county; this is an especially surprising finding when one considers the basic analysis above. Amenity differences seem to matter, with 73 people moving to a county with a one-point higher amenity index score. The differences in median housing values do not seem economically important, with a $1,000

higher level in the destination county leading to 1.13 additional out-migrants to that county. As expected, adjacency seems to be an important destination characteristic.

Table 3. Out-migration estimation results

	Out-migration		In-migration
Variable	Truncated	Tobit	Truncated
Intercept	-978.40	-309.4	-593.69
unemployment rate difference	-16.85	-4.78	-8.22
growth rate difference	-62.27	3.38	32.33
amenity difference	73.05	18.38	-7.36
Median home value difference	1.13	0.75	0.10
adjacent state? (dummy)	310.49	137.05	364.45
origin population	0.52	0.18	0.09

While the truncated model offers an improvement over OLS, it can be argued that information is lost by restricting the model to only those cases where positive migration flows are observed. But it may be as important to consider the fact that people did not move from one county to another as it is to consider the observed migrations. In general, if people move to "better" labor markets and "better" places to live (in terms of amenities) one would expect very little migration to "inferior counties." Thus, use of a truncated sample could still result in selection bias. The Tobit model allows us to properly consider a "mass" of zeroes, which characterizes the Pennsylvania county * US county migration vector (67 * 3,141). Hayashi (2003) provides the derivation of the maximum likelihood estimator. The basic form of the Tobit model is:

$$y_{i,j} = x_{i,j}^{'}\beta_0 + \varepsilon_{i,j}, \quad \varepsilon_{i,j} \Big| x_{i,j} \sim N(0,\sigma_0^2)$$

$$y_{i,j} = \begin{cases} y_{i,j}^{*} & \text{if } y_{i,j}^{*} > 0 \\ 0 & \text{if } y_{i,j}^{*} = 0 \end{cases}$$

We present the Tobit results in the third column of Table 3. Here, we interpret the parameters as the effect of various characteristics on the likelihood of moving to any particular county. For example, we expect 137 people to move to any county in an adjacent state, *ceteris paribus*.

Overall, the signs are generally consistent with the truncated model. The notable exception is the employment growth rate difference, which is now positive, as predicted by theory. In this specification, (arguably the correct one?) we see that a one-percentage point higher in employment growth rate in a non-Pennsylvania county is associated with 3.4 out-migrants to that county. Related, a one percentage point lower unemployment rate in a non-Pennsylvania county will

lead to nearly 5 out-migrants to that county. Both of these findings suggest that job creation can have important effects on reducing out-migration.

We see that amenity differences remain important, with a one point improvement related to 18 out-migrants. Differences in median house values seem not to have an economically significant effect.

In our third model we estimate one in-migration equation. Because the logic of the Tobit model as applied to our problem would require that we look at a humongous migration vector (3,141 * 3,140) dominated by zeroes, we restrict our model to only those instances where the county-to-county flow into Pennsylvania is positive. This results in 18,385 observations. In the final column of Table 3 we provide the results of the maximum likelihood estimation of the truncated model.

When looking at these results it is important to keep in mind that this reports on instances of an observed flow. What we see here is that in-migration to the state is once again driven by adjacency; with 364 of the in-migrants into a typical county from an adjacent state, *ceteris paribus*. Here, we also see that moves are unemployment improving (8.2 in-migrants from any county with a 1-percentage point higher unemployment rate), and employment growth improving (32 additional in-migrants per 1 percentage point difference in growth rates). For in-migration, the sign on amenities is negative, suggesting that select natural features and weather in the state may not be all that attractive. However, this is holding adjacency constant, and many of the states' in-migrants are expatriates from New York, New Jersey and Maryland seeking open space.

5. Discussion and Conclusions

Mark: These are tentative, and we have not had time to think that clearly about the implications. Any ideas on interpreting what we have or suggestions for new directions are greatly welcome.

Our results suggest a number of interesting things. First, our non-parametric and parametric methods seem to show that the model better describes out-migration then it does in-migration; especially for the labor market variables. Thus, we might conclude that the relative labor market conditions are an important push factor, but maybe not so much of a pull factor. This would be consistent with some of the micro-level labor search literature, which suggests people make migration decisions in a two-step process, first deciding to move then deciding where to move. Our findings would support the notion that people want to leave for someplace better, but that there are numerous better places from which to choose.

Our second findings of note are those regarding amenities, and, more particularly, how those are measured. For example, knowing Pennsylvania, there are a substantial number of moves into the state from New York and New Jersey

that are, arguably, amenity improving (the Pocono Mountains in the northeast part of the state are the destination for a large percentage of the states' total in-migrants). Here, people are willing to commute up to two hours one way to Manhattan in order to live outside of the city. Yet, the amenity measures that we use suggest that these counties are not much different than counties from which they migrated. This indicates that how amenities are defined is an important issue.

Also, given that Florida is the destination for a large number of Pennsylvania retirees, it may be important to better segregate our analysis by age class. As this work proceeds we will analyze subsets of the migration data to the extent that it is possible. One predicted implication is that the type of amenity will matter. For example, retirees may consider sun and low-crime as important amenities, whereas younger families might put a greater value on open-space and good schools. If this is true, then a single measure or index of amenities may not be adequate.

As a last point of departure, we intend to take a closer look at the extent and nature of rural-metro differences. Preliminary analysis shows that there is a relatively higher propensity for metro residents to move out of the state. However, rural residents are more likely overall to move, they just tend to migrate to metro counties in Pennsylvania. Yet people moving into the state are disproportionately likely to move to a rural county.

In a nutshell, in-migrants are moving from adjacent states to rural counties (primarily from New York and New Jersey); out-migrants are leaving the state's metro counties for the south and the west; and rural residents are moving to the states' metro areas.