2005

Specification of Functional Form in Models of Population Migration

Brian Cushing

Specification of Functional Form in Models of Population Migration

By

Brian Cushing

RESEARCH PAPER 2005-5

Department of Economics and
Regional Research Institute
West Virginia University
Brian.Cushing@mail.wvu.edu

For presentation at the Southern Regional Science Association Conference
Arlington, VA, April 7-9, 2005

Abstract: A characteristic of the empirical literature on internal population migration is widely varying results and often conflicting conclusions regarding relative importance of explanatory factors. There are a number of possible explanations for these conflicting findings, some of which have received little attention in the literature. This paper focuses on one major specification issue in the context of an aggregate migration model: choice of functional form. The discussion lays out a theoretical basis for choosing functional form. It follows this with a comparison of empirical results for several functional forms. Statistical tests are used to choose the most appropriate functional form.

I. Introduction

The vast empirical literature on internal population migration is characterized by diverse conclusions regarding statistical significance and relative importance of most explanatory variables considered. For almost any explanatory factor, one can find some research showing a statistically significant coefficient with the expected sign, while others yield either a statistically insignificant estimated coefficient, or even a statistically significant coefficient with an unexpected sign. Even those models that have a statistically significant coefficient with the expected sign vary widely with respect to estimates of the variable's relative importance, as indicated by the magnitude of estimated elasticities or standardized beta coefficients. Sometimes the disparities within the literature have become the focus of the literature itself, such as with the literature on "economic" versus "amenity" variables and the "welfare migration" literature.

Several factors might explain the migration literature's conflicting findings. First, researchers often use different series of migration data. Data from the decennial Census of Population and Housing, the Panel Survey of Income Dynamics, the Internal Revenue Service, the National Longitudinal Survey, and other sources are likely to yield somewhat different results due to factors such as different coverage of the population and different lengths of the migration period (e.g., one-year versus five-year). Related to differences in data series is the selected geographical unit of analysis for migration, i.e., region, state, metropolitan, or county. Often the selection of the data source determines this. A third source of differences in results is the selection of the specific point in time of the analysis, such as 1965-70 versus 1975-80 versus 1995-2000 Census data; relationships, such as the willingness to trade-off between economic factors and amenities, may change over time. The level of analysis (individual versus aggregate) also can greatly affect results. Increasingly, studies of U.S. migration rely on microdata with details on individuals. Even after many years, however, there is still uncertainty regarding the most effective way to integrate contextual (place) variables with individual characteristics, and how best to present and interpret the coefficients of these models. Finally, the empirical literature varies widely with respect to the number and types of explanatory variables included in analyses. Many empirical migration models have included only a small number of explanatory variables, especially the models of aggregate migration flows. Omitted variable bias has the potential to substantially alter empirical results and conclusions.

Even aside from the issues above, diverse empirical results are apt to persist due to some less obvious, but important, model specification issues. This paper focuses on one such issue: specification of functional form. As discussed by Cushing and Poot (2004), the aggregate migration literature has generally ignored the issue of functional form specification, despite the work of Goss and Chang (1983) and Greenwood's (1985) call for more work on this issue. From an econometric standpoint, choosing the wrong functional form results in biased coefficient estimates. The bias is analogous to a "wrong variable" or "omitted variable" bias. The magnitude of the bias depends on how closely the chosen form approximates the true functional form. Choice of functional form almost always affects estimates of relative importance, such as elasticities, and often affects the sign and statistical significance of some estimated coefficients.

The remainder of this paper considers specification of functional form and its impact on regression results. The next section considers theoretical aspects of functional form specification, developing practical guidelines for addressing this issue. This is followed by an application to a model of interstate population migration. Section IV summarizes standard procedures for estimating functional form. Section V presents and analyzes empirical results of functional form estimation for the migration model and compares results of the estimated functional form with those of the most commonly used functional forms.

II. Theoretical Aspects of Functional Form Specification

Most aggregate migration models use either the linear or double-log (log-log) functional form. These are generally chosen out of convenience (ease of interpretation of coefficients), without much thought given to the theoretical basis for functional form selection. Ease of interpretation, however, is an inappropriate reason for selecting functional form. Choice of functional form should be guided primarily by theoretical considerations. When theory does not provide adequate guidance, methods such those proposed by Box and Cox (1964) and Box and Tidwell (1962) may be used to estimate functional form. Often, a mixed functional form model will be most appropriate, with the functional relationship of each explanatory variable vis-à-vis the dependent variable considered separately, based on theory.

Relatively few econometrics textbooks devote significant space to specification of functional form. Of those that really consider this issue, most do not provide a strong theoretical basis for choosing among functional forms, and provide few theoretically justified guidelines for selecting functional form. This paper employs three main criteria for selecting among functional forms: (1) theoretical range of variables; (2) changes in relationships as variable values change; and (3) the behavior of Y as X goes toward extremes. The first criterion is easily applied. For example, if a variable can take a negative value, then it is not possible to take a log.[1] The second criterion focuses on changes in the slope and elasticity as the explanatory variable changes. The third considers what happens to Y as X goes toward extremes such as ∞, 0, or $-\infty$. Most variables can be fit to a specific functional form or limited to a choice between just a couple alternatives using these criteria. Table 1 presents key characteristics of some of the major functional forms, which will be used for the analysis in this paper.[2]

Only Goss and Chang (1983) considered functional forms in migration modeling. Their analysis focused on changes in elasticities, but restrictive assumptions dictated their results. They claimed that the elasticity of migration with respect to an explanatory variable should increase as the value of the variable increases. They provided no substantive justification for this blanket statement. Consider distance between origin and destination, which was one of their explanatory variables. Intuition suggests that this elasticity decreases with distance, e.g., if an individual considers a move of 2,000 miles, moving an extra 500 miles would not likely. In addition, Goss and Chang (1983) did not consider the possibility of a mixed functional form. While making a significant advance, the analysis of Goss and Chang (1983) was not sufficiently general. Their

---

[1] For nonnegative data, it is common to add some small amount to zero values so that a logarithmic functional form may be used. This practice is generally considered as appropriate as long as the variable does not have a large number of zero values.

[2] Due to the ambiguity regarding names such as log-linear and exponential functional forms, this study adopts the naming convention used by Studenmund (2001) that refers to the linear form as "level-level," the double-log form as "log-log," etc. The first term refers to the form of the dependent variable and the second term to the form of the explanatory variable. "Level" refers to a variable in linear (untransformed) form, while log refers to a variable in log form.

model was only slightly less restrictive than the model they attempted to correct. Unfortunately, this issue has been off the radar in the migration literature since that time.

III. The Migration Model

For illustrative purposes, I use a relatively simple model and choose data to fit in the context of a large number of migration studies, including Goss and Chang (1983). The model is one of aggregate interstate migration flows, using 1995-2000 migration data from the 2000 *Census of Population and Housing*. The model considers place-to-place migration flows with an "allocation rate" as the dependent variable and is similar to models employed by Greenwood (1969), Wadycki (1974), Kau and Sirmans (1976), Goss and Chang (1983), and Cushing (1989). The dependent variable is an allocation rate of migration (ALLRATE): the percentage of all outmigrants from state i who chose state j as the destination. The allocation rate is easily derived from a gravity-type model of migration. Because it focuses on those who migrated without regard for those who did not migrate, origin characteristics need not be considered [see Cushing (1989)]. Since the model considers aggregate (as opposed to individual) migration, it relies on the assumption that areas with characteristics that are generally associated with higher levels of welfare disproportionately attract migrants. Such characteristics include better economic opportunities and better amenities. High costs of migration may mute this attraction.

The explanatory variables are defined in Table 2. The relevance of most of these variables is self-evident. Potential migrants should be attracted to locations with better economic opportunities, as reflected by greater employment growth (EMPGROW) and lower unemployment rates (UNEMPLOY). Likewise, states with more moderate climates, including milder winters (lower heating degree days) and milder summers (lower cooling degree days) should be more attractive on average. Population size (POP95), population density (DENSITY), the number of persons born in destination state *j* but residing in origin state *i* prior to the migration period, as a percentage of the population of state *i* (RETURN), and the number of persons born in origin state *i* but residing in destination state *j* prior to the migration period, as a percentage of the population of state *i* (MIGSTOCK) capture some aspects of amenities and some of migration costs. Larger populations provide more and better economic, social, and cultural opportunities, ceteris paribus. In addition, large populations increase information flows regarding a potential destination, thus reducing migration costs. Greater population density may indicate better opportunities and more cultural diversity, but also more disamenities such as congestion and pollution. The importance of return migration is well-documented. Returning to family and friends is often very desirable and, due to more and better information, migrants may incur relatively low costs in making such a move. Thus larger values of RETURN should coincide with relatively greater migration from state *i* to state *j*. MIGSTOCK is similar to the migrant stock variable that has been used in many studies since Greenwood (1969). Many believe this variable represents a family, friends, and information effect. Finally, due to greater monetary, psychic, and information costs, greater spatial distance between the origin and destination states (DISTANCE) should yield less migration.

Specification of Functional Form

The remaining issue for the model regards choice of functional form. To address this, I make use of the key characteristics of functional forms presented in Table 1 and select those well-known functional forms that best fit the characteristics of each variable.

That ALLRATE has a theoretical range of zero to 100 suggests a log form for the dependent variable.[3] Goss and Chang (1983) provide some empirical support for the log form. If a log form of the dependent variable is correct, then the log-level, log-log, and log-reciprocal relationships are relevant for the explanatory variables. Note from Table 1 that these forms allow for a choice between increasing, constant, or decreasing elasticity.

For EMPGROW, the most appropriate functional relationship (with ALLRATE) is the ***log-level*** form. Unlike the log-log and level-log, this form allows for negative values and allows for a positive value of ALLRATE even when EMPGROW falls to zero. Unlike the reciprocal and log-reciprocal from, the log-level does not have a discontinuity of effect as EMPGROW goes from positive to negative. It also permits EMPLOY to have an overwhelming effect as it takes on a large negative value (more so than the linear form due to the multiplicative relationship to

[3] The limited range of the dependent variable also suggests that a Logit estimation procedure might be appropriate, with the dependent variable rescaled to be a proportion rather than a percent. While beyond the scope of the present paper, this alternative will also be explored

other variables). The log-level is consistent with the expected form of the dependent variable. (A linear form would be an alternative if the dependent variable were to enter in linear form.) Finally, the increasing slope and increasing elasticity are intuitively appealing, implying that as employment growth increases, the migration response will accelerate, suggestive of boomtown growth (and busts) that we sometimes observe.

Depending on assumptions regarding UNEMPLOY, a number of functional forms are possible. If ALLRATE does not approach infinity (100 percent given the range of the variable) as UNEMPLOY approaches zero, but approaches zero as UNEMPLOY approaches 100 percent, then the log-level form is most appropriate. If the first part of the previous statement is incorrect, then the log-log and log-reciprocal may fit well. The elasticity decreases, is constant, or increases, respectively for the three forms. The characteristics of the ***log-level*** and its decreasing elasticity seem most intuitive. Once it reaches a high enough level, further increases in the unemployment rate are unlikely to elicit any additional migration response – those who still decide to move to the location probably do not care about the unemployment rate or job prospects.

Theoretical range suggests a log form for POP95. As POP95 approaches zero, it makes sense that ALLRATE would also approach zero, which fits with both the log-log and log-reciprocal forms. It is not clear whether a constant elasticity (***log-log***) or a decreasing elasticity (***log-reciprocal***) makes more sense in this case. Some might suggest that POP95 could have a quadratic effect, initially having a positive effect, but eventually a negative effect. While appropriate at some geographic level, it is not likely to hold for a sample of US states, where it is possible to avoid large population concentrations within virtually any state.

For DENSITY, behavior of elasticities may provide the best guidance for functional form. If density had an attractive effect, the level of responsiveness should taper off as density gets large. If density has a negative effect, an increasing elasticity might make sense if disamenities such as congestion increase exponentially as density of development reaches a high level. This elasticity response (for both the positive and negative effect) fits the ***log-reciprocal*** form. For a positive relationship, both the log-reciprocal and the log-log functions are consistent with ALLRATE falling to zero as DENSITY falls to zero.

For both HEATDEG and COOLDEG, the exponential, log linear, and log reciprocal forms are all consistent with the property that as HEATCOOL $\rightarrow \infty$, ALLRATE $\rightarrow 0$. But the latter two forms imply that as HEATCOOL $\rightarrow 0$, ALLRATE $\rightarrow \infty$, which is inappropriate. The ***log-level*** (with a decreasing elasticity since $B < 0$) is the hypothesized form.

Like ALLRATE, RETURN has a theoretical range from 0 to 100. In theory, MIGSTOCK could exceed 100, but in practice its value is likely to be of the same order of magnitude of RETURN and ALLRATE. A value of RETURN = 0 or MIGSTOCK = 0 would signify little contact between the origin and destination pair, in which case one might expect that ALLRATE would equal zero or be very close to zero. Likewise, a high value of these variables would indicate an ongoing stream of migration between the pair of states. Together, these suggest that these variables might have some type of proportional relationship with ALLRATE, which leads us to a ***log-log*** relationship.

As discussed previously, intuition indicates that the elasticity of ALLRATE with respect to DISTANCE probably decreases as distance increases. In addition, ALLRATE does not become explosively high as distance approaches zero, e.g., all migrants do not go to adjacent states, ignoring all other states. The ***log-level*** form fits both of these characteristics.

IV. Estimating Functional Form: Box-Cox/Box-Tidwell Analysis

Box and Cox (1964) and Box and Tidwell (1962) developed procedures that allow the proper functional form to be estimated along with the standard parameters of a regression model. Over the past four decades, these procedures have been extended and refined. [For example, see Zarembka (1974), Savin and White (1978), Lahiri and Egy (1981), Spitzer (1982, 1984), and Seaks and Layson (1983).] In its most general form, the Box-Cox/Box-Tidwell (BCBT) model is

(1) $(Y^{\lambda} - 1) = B_0 + B_1(X_1^{\mu_1} - 1)/\mu_1) + B_2(X_2^{\mu_2} - 1)/\mu_2) + \ldots + B_k(X_k^{\mu_k} - 1)/\mu_k) + \varepsilon$

or

(1a) $Y^* = B_0 + B_1(X_1^*) + B_2(X_2^*) + \ldots + B_k(X_k^*) + \varepsilon$

The procedure basically amounts to finding the values of $\lambda, \mu_1, \ldots, \mu_k$ such that Y* is truly linear in the (X_i^*) in Equation (1a).[4] Using any of a number of procedures [see Spitzer (1982)], estimates can be obtained not only for $B_0, B_1, \ldots, B_k$, but also for $\lambda, \mu_1, \ldots, \mu_k$. A value of λ or μ_i equal to zero implies that the variable enters the model in log form.[5] A value of one implies a linear form. Several combinations of λ and μ_i that yield well known functional forms are shown in Table 3. Note that Equation (1) allows each explanatory variable to have a different functional relationship with the dependent variable. Box-Cox estimation sets values for all μ_i, but estimates λ. Box-Tidwell estimation sets the value of λ, then estimates the μ_i. Full Box-Cox/Box-Tidwell estimation estimates all parameters in equation (1).[6]

Goss and Chang (1983) employed a variation of Equation (I) that constrained all explanatory variables to have the same transformation, i.e., $\mu_1 = \mu_2 = \mu_k$ in Equation (1). This restricts all explanatory variables to have the same general functional relationship with the dependent variable. Their discussion, however, provided no reason, a priori, for the validity of this restriction.

Box and Cox (1964, p. 213) clearly recognized that estimation of functional form should be an aid rather than the final authority in choosing a functional form. Like all statistical work, this is simply an estimation procedure based on a sample of data and its results are subject to error. It

[4] As noted by Zarembka (1974) and others, transformation of the explanatory variables only focuses on additivity of effect (nonlinearities). Transformation of the dependent variable focuses on additivity of effect, normality of the errors, and constant error variance. Zarembka points out that estimation of the additivity is robust to normality considerations, i.e., not biased by the focus on achieving normality, as long as the errors are reasonably symmetric. It is not, however, robust to the constant variance focus. Estimation of λ is biased toward a homoscedastic functional form. Since the dependent variable to be used later is defined in such a way to eliminate the greatest potential source of heteroscedasticity, I assume that this bias is not a problem.

[5] For the BCBT regressions, as well as some others, variable values must be positive. Therefore, a few variables had to be adjusted by adding small numbers that would leave all observations positive.

[6] Some variations restrict λ and all μ_i to have the same value, or set λ, then estimate restrict the μ_i with the restriction that all μ_i have the same value.

should not be followed blindly. The analysis below compares Box-Cox/Box-Tidwell results with results using standard functional relationships, as well as the mixed functional form model discussed in Section III of this paper.

V. Empirical Results

The econometric estimation uses migration flows for the lower 48 states, plus the District of Columbia. The model considers migration from each of the 49 states (including the District) to each of the other 48 states, thus yielding a sample size of 2,352. Given the choice of the explanatory variables, including the time period covered, simultaneity bias should not be a problem. As often happens with completely unrestricted Box-Cox/Box-Tidwell estimations, the estimation would not converge for the full model, though I could obtain convergence for some small subsets of the explanatory variables. In all of these reduced models and in a separate Box-Cox estimation, λ was very close to zero. Given these results and the theoretical basis supporting a log form of the dependent variable, the most general estimation reported restricts $\lambda=0$ (log form) and estimates a Box-Tidwell model.

Empirical results are presented in Tables 4 and 5. Estimated elasticities at the mean are presented for purposes of comparison, along with an indication of the change in the elasticity as the value of the variable increases. Table 4 includes the Box-Tidwell results and three other model variations. The "All" model replicates the Box-Tidwell estimation with the restriction that all μ_i are equal. For the "Closest Form" model, $\mu_1, ..., \mu_9$ from the Box-Tidwell results are rounded to the closest interpretable value (1, 0, or -1), followed by ordinary least squares regression. The "Hypothesized Form" model replicates the functional form derived from the theoretical discussion in Section III. Table 5 presents results from four of the most commonly used functional forms: linear (level-level), double-log (log-log), log-level, and level-log.

With the exception of the density variable in the All, Log-Log, and Level-Log models, all estimated coefficients are statistically significant with expected signs, almost all at the one percent significance level. Interestingly, almost all of the weaker hypothesis tests occur in the two models that restrict all explanatory variables to have constant elasticity relationships with ALLRATE.

The Box-Tidwell estimation suggests that EMPLOY, UNEMPLOY, DENSITY, and HEATDEG should have a log-level relationship with ALLRATE. It also supports a log-log relationship for the remaining variables. These functional relationships are constructed in the Closest Form model. For the most part these "suggested" relationships match those developed back in Section III based on theory, as shown in the Hypothesized Form model. The two models differ with respect to the functional form for DENSITY, DISTANCE, and COOLDEG

Based on likelihood ratio tests, all seven of the restricted functional forms can be rejected in favor of the less restrictive Box-Tidwell model.[7] The log(Likelihood Statistics) indicate that those models imposing a linear form on ALLRATE (linear and level-log) perform much worse than do the other models. Approximating the correct form of the dependent variable appears to be crucial. The log-level model, the only one of the remaining functional forms that does not allow any constant elasticity relationships, also fares noticeably worse than the other choices. The ALL, Closest Form, and perhaps the Hypothesized Form" models fare the best relative the Box-Tidwell model.

Most elasticities vary substantially across models, though some interesting patterns emerge. When entered in level form, the two economic opportunity variables are almost always among the more important variables, based on elasticities. When entered in log form, their impacts are much weaker. For MIGSTOCK and RETURN, the pattern is just the opposite. HEATDEG has the most volatile pattern of elasticities across models. It has either the largest or second largest relative impact in every model except for the Box-Tidwell model, including the three highest elasticities in the eight regression results. The contention of Goss and Chang (1983) that all elasticities should increase with the value of the explanatory variable does not hold up.

[7] The null hypothesis that the restricted model is valid can be rejected if $-2\log(\theta) > \chi^2$ with q degrees of freedom, where $\log(\theta)$ = log(Likelihood Statistic from the Unrestricted Model) - log(Likelihood Statistic from the Restricted Model) and q is the number of restrictions. The Box-Tidwell model is the unrestricted model in this case.

VI. Conclusion

The empirical literature on internal population migration is characterized by widely varying results and often conflicting conclusions regarding relative importance of explanatory factors. While variety can spur learning, the lack of consensus can frustrate those trying to learn from this vast literature and undoubtedly leads to extra time and effort for many of those working in the field. A number of factors might explain the conflicting findings. Most often, researchers employ the linear or double-log forms due to their ease of application and interpretation. Several major conclusions can be drawn from this study.

1) Imposing the same structural relationship for all explanatory variables (with the dependent variable) is generally too restrictive on theoretical and empirical grounds. Mixed functional forms are no more difficult in empirical work and may yield superior results;
2) A good theory of functional relationships in a model is very fruitful, especially when combined with a Box-Cox/Box-Tidwell analysis;
3) Box-Cox/Box-Tidwell type regressions should be used as information to be combined with theory in determining functional form. As a purely quantitative technique, the results cannot be expected to be perfect and may be difficult to interpret and apply;
4) Ignoring considerations of functional form altogether may result in substantial biases in estimated coefficients and elasticities, as well as in the behavior of these parameters as variable values change. An empirical model whose important behavioral characteristics do not reflect reality can seriously impede our understanding of whatever process is being studied.

The empirical results also yield some specific information on functional form for studies of population migration. Combined with the results of Goss and Chang (1983), strong evidence suggests that the log form of the dependent variable is much preferred to the linear form. The study provides at least initial theoretical and empirical evidence for the proper functional form of several key variables used in migration studies, along with estimates of their relative impact and the change in their impact on migration. More importantly, the study provides a basis for thinking about choice of functional form.

References

Box, G. and D. Cox. "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Series B 26 (1964), 211-243.

Box, G. and P. Tidwell. "Transformation of the Independent Variables," *Technometrics* 4 (1962), 531-550.

Cushing, Brian. "Use and Misuse of the Allocation Rate in Models of Population Migration," *Annals of Regional Science* 23 (1989), 51-58.

Cushing, Brian and Jacques Poot. "Crossing Boundaries and Borders: Regional Science Advances in Migration Modelling," *Papers in Regional Science*, 83 (2004), 317-338.

Goss, E. and H. Chang. "Changes in Elasticities of Interstate Migration: Implication of Alternative Functional Forms," *Journal of Regional Science* 23 (1983), 223-232.

Greenwood, M. "An Analysis of the Determinants of Geographic Labor Mobility in the United States," *Review of Economics and Statistics* 51 (1969), 189-194.

Greenwood M. "Human Migration: Theory, Models, and Empirical Studies," *Journal of Regional Science* 25 (1985), 521-544

Kau, James and C.F. Sirmans. "New, Repeat, and Return Migration: A Study of Migrant Types," *Southern Economic Journal*, 43 (1976), 144-1148.

Lahiri, K. and D. Egy. "Joint Estimation for Functional Form and Heteroskedasticity," *Journal of Econometrics* 15 (1981), 299-307.

Savin, N. and K. White. "Estimation and Testing for Functional Form and Autocorrelation: A Simultaneous Approach," *Journal of Econometrics* 8 (1978), 1-12.

Seaks, T. and S. Layson. "Box-Cox Estimation with Standard Econometric Problems," *Review of Economics and Statistics* 65 (1983), 160-164.

Spitzer, J. "A Primer on Box-Cox Estimation," *Review of Economics and Statistics* 64 (1982), 307-313.

Spitzer, J. "Variance Estimates in Models with Box-Cox Transformation: Implications for Estimation and Hypothesis Testing," *Review of Economics and Statistics* 66 (1984), 645-652.

Studenmund, A. *Using Econometrics: A Practical Guide*, 4th edition, Reading, MA: Addison-Wesley, 2001.

Wadycki, Walter. "Alternative Opportunities and Interstate Migration: Some Additional Results," *Review of Economics and Statistics*, 6 (1974), 254-257.

Zarembka, P. "Transformations of Variables in Econometrics," in P. Zarembka (ed.), *Frontiers in Econometrics*. New York: Academic Press, 1974.

Table 1: Key Characteristics of Major Functional Forms

	Level-Level: $Y=A+BX$	Log-Level: $Y=e^{A+BX}$ or $\ln Y=A+BX$
Slope: $\Delta Y/\Delta X$	B	BY
Δslope/ΔX	constant	increasing
Elasticity	B(X/Y)	BX
Δelasticity/ΔX	AB>0; increasing AB<0; decreasing	B>0; increasing B<0; decreasing
Value of Y as		
$X \to 0$	term drops out	term drops out
$X \to \infty$	$B>0; Y \to \infty$ $B<0; Y \to -\infty$	$B>0; Y \to \infty$ $B<0; Y \to 0$
$X \to -\infty$	$B>0; Y \to -\infty$ $B<0; Y \to \infty$	$B>0; Y \to 0$ $B<0; Y \to \infty$

	Level-Log: $Y=A+B\ln X$	Log-Log: $Y=AX^B$ or $\ln Y=A^*+B\ln X$
Slope: $\Delta Y/\Delta X$	B/X	B(Y/X)
Δslope/ΔX	B>0; decreasing B<0; increasing	B<0 or B>1; increasing 0<B<1; decreasing B=1; constant
Elasticity	B/Y	B
Δelasticity/ΔX	decreasing	constant
Value of Y as		
$X \to 0$	$B>0; Y \to -\infty$ $B<0; Y \to \infty$	$B>0; Y \to 0$ $B<0; Y \to \infty$
$X \to \infty$	$B>0; Y \to \infty$ $B<0; Y \to -\infty$	$B>0; Y \to \infty$ $B<0; Y \to 0$
$X \to -\infty$	not applicable	not applicable

Table 1: Key Characteristics of Major Functional Forms (cont.)

	Reciprocal: $Y=A+B/X$	Log-Reciprocal: $Y=e^{A+B/X}$ or $\ln Y=A^*+B/X$
Slope: $\Delta Y/\Delta X$	$-B/X^2$	$-BY/X^2$
Δslope/ΔX	$BX>0$; increasing $BX<0$; decreasing	$B>0$ and $(B/X)>-1$; increasing $B<0$ and $(B/X)<-1$; increasing Otherwise, decreasing
Elasticity	$-B/(YX)$	$-B/X$
Δelasticity/ΔX #	>0 if $B(YX^2-BX)>0$ <0 if $B(YX^2-BX)<0$	$B>0$; increasing $B<0$; decreasing
Value of Y as		
$X \rightarrow 0$	$BX>0$; $Y \rightarrow \infty$ $BX<0$; $Y \rightarrow -\infty$	$BX>0$; $Y \rightarrow \infty$ $BX<0$; $Y \rightarrow 0$
$X \rightarrow \infty$	term drops out	term drops out
$X \rightarrow -\infty$	term drops out	term drops out

Recall that for the reciprocal and log-reciprocal functions that $B < 0$ for a positive relationship between Y and X (and vice versa).

Table 2: Detailed Variable Definitions

A. Dependent Variable

ALLRATE - the number of persons residing in destination state *j* on April 1, 2000 who resided in origin state *i* on April 1, 1995, as a percentage of the number of persons who resided in origin state *i* on April 1, 1995 but in another state on April 1, 2000 (percent) [US Bureau of the Census].

B. Explanatory Variables

EMPGROW percent change in total full-time and part-time employment (place of work), 1990-98 [REIS, US Bureau of Economic Analysis];

UNEMPLOY mean annual average unemployment rate, 1995-98 [US Bureau of Labor Statistics];

POP95 total population (millions), 1995 [US Bureau of the Census];

DENSITY population density (hundreds of persons per square mile), 1995 [US Bureau of the Census];

RETURN percent of the population of origin state i born in destination state j at the time of the previous census (1990);

MIGSTOCK number of persons born in origin state i but residing in destination state j, as a percent of the population of origin state i, at the time of the previous census (1990);

DISTANCE highway mileage between the principal city of origin state i and that of destination state j [Official Table of Distances];

HEATDEG average annual heating degree days (thousands of degree days) - weighted average of cities >100,000 or principal city if no cities >100,000 [NOAA];

COOLDEG average annual cooling degree days (thousands of degree days) - weighted average of cities >100,000 or principal city if no cities >100,000 [NOAA].

Table 3: Common Functional Forms Implied by λ and μ

λ	μ	Functional Form
1	1	Level-Level (Linear)
0	0	Log-Log (Double-Log)
0	1	Log-Level
1	0	Level-Log
1	-1	Reciprocal
0	-1	Log-Reciprocal

Table 4: Empirical Results – Box-Tidwell and Mixed Forms

(Dependent Variable = ALLRATE)

	Box-Tidwell λ=0; μ=(below)			Box-Tidwell: All λ=0; μ=-0.04		Closest Form λ=0; μ=(below)			Hypothesized Form λ=0; μ=(below)		
Variable	μ	Elasticity and Change		Elasticity and Change		μ	Elasticity and Change		μ	Elasticity and Change	
EMPGROW	2.12	0.575	↑	0.156	0	1	0.768	↑	1	0.570	↑
UNEMPLOY	3.30	-0.536	↓	-0.155	0	1	-0.510	↓	1	-0.268	↓
POP95	0.14	0.178	↑	0.047**	0	0	0.193	0	0	0.060	0
DENSITY	5.04	0.000	0	0.001^I	0	1	0.033	↑	-1	0.001*	↓
DISTANCE	-0.19	-0.220	↑	-0.109	0	0	-0.248	0	1	-0.096	↓
HEATDEG	0.75	-0.283	↓	-0.358	0	1	-0.938	↓	1	-0.675	↓
COOLDEG	-0.05	-0.395	0	-0.048**	0	0	-0.389	0	1	-0.217	↓
MIGSTOCK	-0.06	0.300	0	0.470	0	0	0.334	0	0	0.436	0
RETURN	-0.05	0.404	0	0.328	0	0	0.406	0	0	0.406	0
Log Likelihood		-1218.380		-1313.060			-1264.310			- 1393.590	

Unless noted, all estimated coefficients are statistically significant at the one percent level.

**significant at the five percent level; *significant at the ten percent level; ^I statistically insignificant

Table 5: Empirical Results – Common Functional Forms

(Dependent Variable = ALLRATE)

	Level-Level λ=1; μ=1		Log-Log λ=0; μ=0		Log-Level λ=0; μ=1		Level-Log λ=1; μ=0	
Variable	Elasticity and Change		Elasticity and Change		Elasticity and Change		Elasticity and Change	
EMPGROW	0.634	↑	0.210	0	0.867	↑	0.012	↑
UNEMPLOY	-0.766	↓	-0.134	0	-0.964	↓	-0.567	↓
POP95	0.143	↑	0.044**	0	0.434	↑	-0.208	↓
DENSITY	0.033	↑	0.021^I	0	0.023	↑	-0.032^I	↓
DISTANCE	-0.384	↓	-0.104	0	-0.755	↓	-0.384	↓
HEATDEG	-1.059	↓	-0.348	0	-1.775	↓	-1.148	↓
COOLDEG	-0.291	↓	-0.034*	0	-0.596	↓	-0.559	↓
MIGSTOCK	0.237	↑	0.481	0	0.042	↑	0.452	↑
RETURN	0.408	↑	0.345	0	0.214	↑	0.341	↑
Log Likelihood	-4331.120		-1510.830		-2632.230		-5111.840	

Unless noted, all estimated coefficients are statistically significant at the one percent level.

**significant at the five percent level; *significant at the ten percent level; ^I statistically insignificant