BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR–90/07

# FORECASTING STATE–TO–STATE MIGRATION RATES

by

Edward W. Frees
Visiting Principal Researcher

U.S. Bureau of the Census and

University of Wisconsin – Madison

Statistical Research Division
Bureau of the Census
Room 3000, FOB #4
Washington, DC 20233 U.S.A.



Recommended by: Nash J. Monsour

Report completed: April, 1990

Report issued: April 17, 1990

# FORECASTING STATE-TO-STATE MIGRATION RATES

by

Edward W. Frees

U. S. Bureau of the Census and

University of Wisconsin - Madison

School of Business and Department of Statistics

1155 Observatory Drive

Madison, Wisconsin 53706

April, 1990

Keywords and Phrases: Population forecasting, diagnostic tests, longitudinal data, autoregressive models.

# Forecasting State-to-State Migration Rates

Edward W. Frees, U. S. Bureau of the Census and
University of Wisconsin

## ABSTRACT

Models of internal migration can be found in the demographic, geographic and economic literatures. Unlike most of these models which are based on cross-sectional analysis of migration rates, the approach here is to incorporate longitudinal as well as cross-sectional aspects. Empirical modeling using this approach is possible using a data base recently revised and updated by the U. S. Bureau of the Census. The model building approach presented in this paper relies heavily on diagnostic and graphical methods. An interesting methodological aspect of the model building is the presence of rampant cross-sectional heteroscedasticity, moderate contemporaneous correlation and mild longitudinal autocorrelation in the transformed data. An important conclusion is that, based on the available data, differenced rates appear stationary. A corollary of this conclusion is that the most recent migration rate takes on an important role in short-term forecasting.

# Forecasting State-to-State Migration Rates

## 1. INTRODUCTION

Governments, corporations and individuals rely on projections of the population for a wide variety of planning purposes. At the state level, projections are made not only by the Federal government, but also by nearly every state (U. S. Bureau of the Census, 1988). While projections at the national level enjoy a desirable level of stability, and hence reliability, projections for smaller geographic regions are much more volatile over time (Long, 1977, Long and Wetrogan, 1986). Much of the variability at the state level can be attributed to migration (Hajj, 1975, Ter Heide, 1963). Following a basic demographic accounting method (Shyrock et al, 1976), the population at the end of the year ($P_1$) can be thought of as the population at the beginning of the year ($P_0$), plus births (B), minus deaths (D), plus net in-migrants (NM), i.e,

$$P_1 = P_0 + B - D + NM. \tag{1.1}$$

Migration can be further decomposed into external, or international, migration and internal migration. At the subnational level, birth, death and international migration processes are less volatile than internal migration. This paper is about modeling internal migration, or more specifically, modeling migration from state-to-state. Although only a piece of a larger puzzle, internal migration has been the subject of extensive discussion; an early survey article by Greenwood (1975a) contains over 250 references.

Models of migration are also useful in understanding the wide variety of geographic, economic and demographic factors which affect migration patterns. At the individual level, many factors can influence the decision to relocate. These so-called 'life-cycle' considerations may include completion of schooling, entry into the labor force, change of marital status, birth and aging of children, retirement, cf., Greenwood (1981). At a national level, factors such as general economic conditions, advent of war, and changing demographic composition of the populace influence migration patterns. At a regional level, relocation of employer, level of public assistance benefits, and quality of life are examples of many variables which may influence the decision to migrate. Models which help to explain these sources of variability can be useful in making public policy decisions which, in effect, may alter the very data on which the models are based. It is possible that forces other than economic, geographic and demographic have a considerable impact on levels of internal migration. For example, Long (1983) documents the considerable influence of college and military populations on levels of internal migration.

There are a variety of quantities one could examine to model and project migration at the state level. The most direct quantity is net migration, defined to be number of in-migrants minus out-migrants. Alternatively, one could examine both "gross" in-migrants and out-migrants. A yet more detailed source of data is to examine the out-migrants on a destination-specific basis. Of these three types of quantities, destination-specific out-migration provide the greatest amount of information and are examined in this paper. For each quantity, if numbers of people are considered, these quantities are termed "flows." The corresponding rates are defined to be the flows divided by the origin-specific population. The migration variable used depends on the purpose at hand. The focus of this paper is to forecast migration rates to be used in short-term (5 to 6 years) population projections. Thus, as argued by Long and Wetrogan (1985), because of the presence of other processes used in updating (1), rates are of greater interest than flows. Specifically, the rates available for modeling and forecasting are of the form $R_{o,d,t}$, where 'o' is for state of origin, 'd' is for state of destination and 't' is for time. The index 'o' ranges from 1 to 51 which includes the 50 states in the Union plus the District of Columbia. The index 'd' also ranges from 1 to 51 but since intra-state moves are not counted, o $\neq$ d. Thus, for each t, there are (51 x 50 =) 2,550 cells. The index 't' ranges from 1 to 13 corresponding to migration years 1975 to 1987. This data set was created from Internal Revenue Service (IRS) matched administrative records and is further discussed in Appendix A. To get an idea of recent state-to-state migration, in Figure 1a is a map of net migration by state. These rates were computed by matching 1986 with 1987 IRS returns. To get an idea of recent patterns in level of migration, in Figure 1b is a time series plot of migrants as a percentage of population over time.

Plot of 1987 Net Migration Rates by State

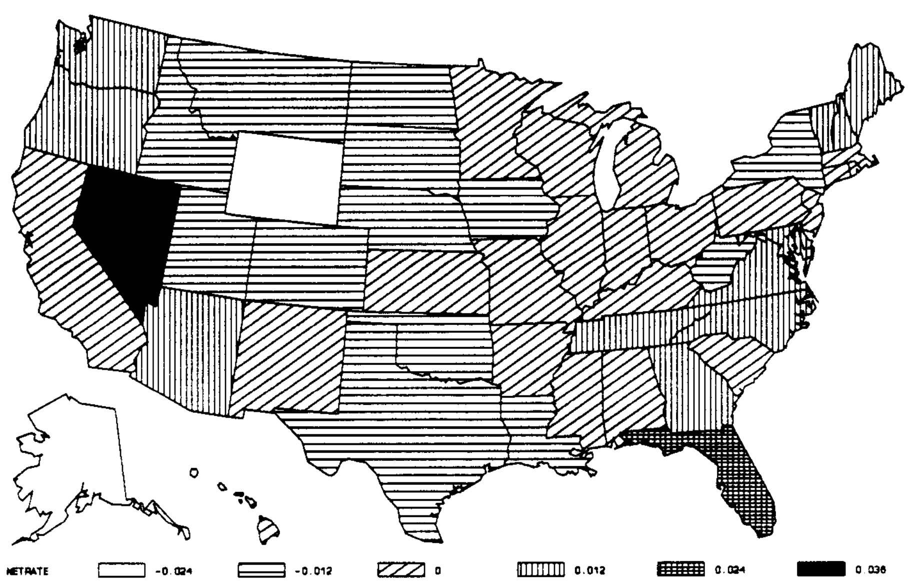



*FIGURE 1A*

Time Series Plot of Migration Rates

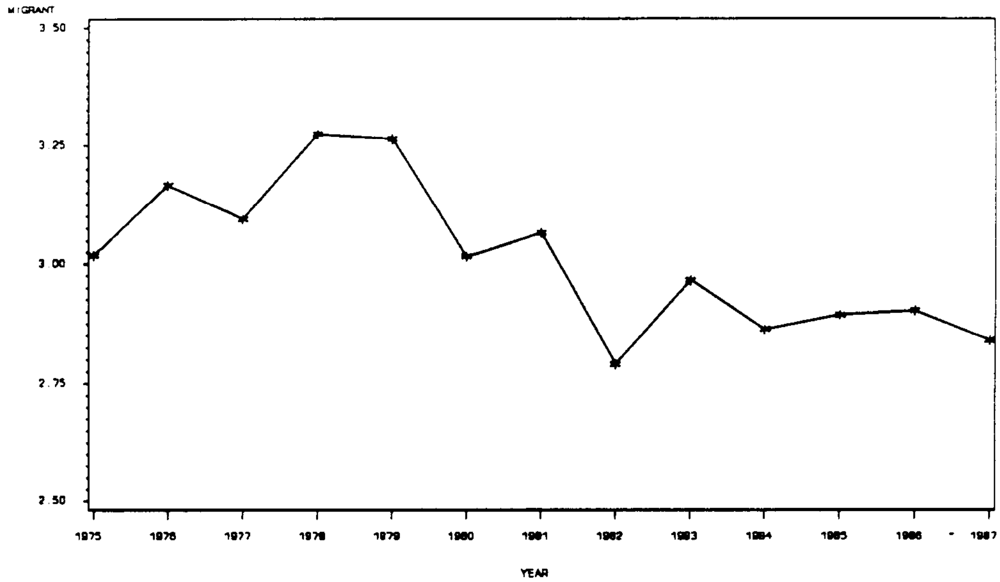



*FIGURE 1B. Migration rate is number of migrants as a percentage of population.*

Models of migration can be found in the demographic, geographic and economic literatures. Some good sources which contrasts these different approaches can be found in the reports of 1977 and 1982 conferences, co-sponsored by the American Statistical Association (ASA) and the National Science Foundation (NSF), in Erickson and Engels (1977) and Isserman (1986). Other sources include monographs by Greenwood (1981) and Rogers and Willekins (1986) which contain discussions from the economic and demographic perspectives, respectively. As an example of these models, consider the so-called "gravity" model (cf., Greenwood, 1975a),

$$M_{od} = c\ P_o\ P_d\ /\ D^a_{od}\ \ (I_d\ /I_o)^b\ (E_d\ /\ E_o)^f\ e_{od} \tag{1.2}$$

for migration from the $o^{th}$ to $d^{th}$ state. Here, P is state population, I is state income, E is state (un)employment, D is distance between population centroids, a, b, c and f are parameters to be estimated, and $e_{od}$ is the multiplicative error term. This model can be easily converted to the linear model via the logarithmic transform. Model (1.2) and other models share the characteristic that they have been estimated using cross-sectional analysis of either migration (cf., Plane and Rogerson, 1985) or changes in migration (cf., Plane, 1987) in lieu of following each state-to-state cell through time. As noted in the 1977 and 1982 ASA/NSF conference reports, no real attempts have been made to forecast migration cells using time series methodology due to the lack of available data, not the lack of researchers' awareness of the desirability of following each state-to-state cell through time. As noted in the 1977 and 1982 ASA/NSF conference reports, no real attempts have been made to forecast migration cells using time series methodology due to the lack of available data, not the lack of researchers' awareness of the desirability of this approach.

The main purpose of this paper is develop a model that incorporates a time series, as well as cross-sectional, approach to modeling and forecasting state-to-state migration rates. Techniques from the longitudinal, or pooled cross-sectional time series, data literature are used. For introductory material from an econometric and a biostatistics viewpoint, see Dielman (1983), Judge et al (Chapters 12 and 13, 1985), Hsiao (1986), Rao (1987) and Ware (1985). From a modeling viewpoint, an interesting aspect of this exercise is the rampant cross-sectional heteroscedasticity, moderate contemporaneous correlation, and mild longitudinal autocorrelation that is present in the transformed data. Because of the lack of experience in analyzing state migration over time, statistical criteria are used to judge the desirability of models in lieu of theory from the underlying functional fields; demographic, economic and geographic. Indeed, it is hoped that this research provides a foundation for future work by researchers from these different functional fields. Thus, an extensive portion of the paper, Section 2, is devoted to an exploratory graphical analysis which presumably is basic to any subsequent model development.

In Section 3, some forecasting models using only current and previous rates are introduced. It is useful to build a model for forecasting rates using only current and past rates for at least three reasons. First, the process of model building often reveals interesting features of the data. Second, the model obtained is a useful starting point in constructing more complex models using additional information in the form of explanatory variables. Third, in many situations it is not clear that the additional explanatory variables used in the model building will be a reliable source for future applications of the forecasting model and, hence, it is desirable to have simpler alternatives available. These models are evaluated using the in-sample diagnostic devices introduced in Section 2 and some out-of-sample validation measures. In Section 4, the forecasting implications of these models are discussed.

## 2. EXPLORATORY GRAPHICAL ANALYSIS

In this section, graphical and diagnostic techniques are used to explore the structure of the data. On diagnostic methods for regression data, some basic references are Cook and Weisberg (1982) and Belsley, Kuh and Welsh (1980). Another good source is Carroll and Ruppert (1988) whose treatment focuses on estimating the structure of the variance. This turns out to be an important feature of this data set. As noted by Ware (1985), little attention has been given to diagnostic methods for longitudinal data.

To begin the graphical analysis of the destination-specific out-migration rates $R_{o,d,t}$, recall that there are 51 x 50 x 13 = 33,150 observations. Because of the magnitude of the full data set, meaningful graphs of only subsets can be presented. First consider the rates for a particular state of origin. In Figure 2a is a multiple time series plot for the origin state of Wisconsin where each series represents a particular state of destination. For example, the two series at the top of Figure 2a are out-migration rates to Illinois and Minnesota, respectively. Here, time refers to the migration year. From Figure 2a, I note that the variability of the series seems to increase with the level, typical of rate data and what would be expected under a Binomial or Poisson distribution for the underlying counts. Figure 2a, and similar graphs for other states in Appendix C, indicate that the mean for each series does not depend on time.

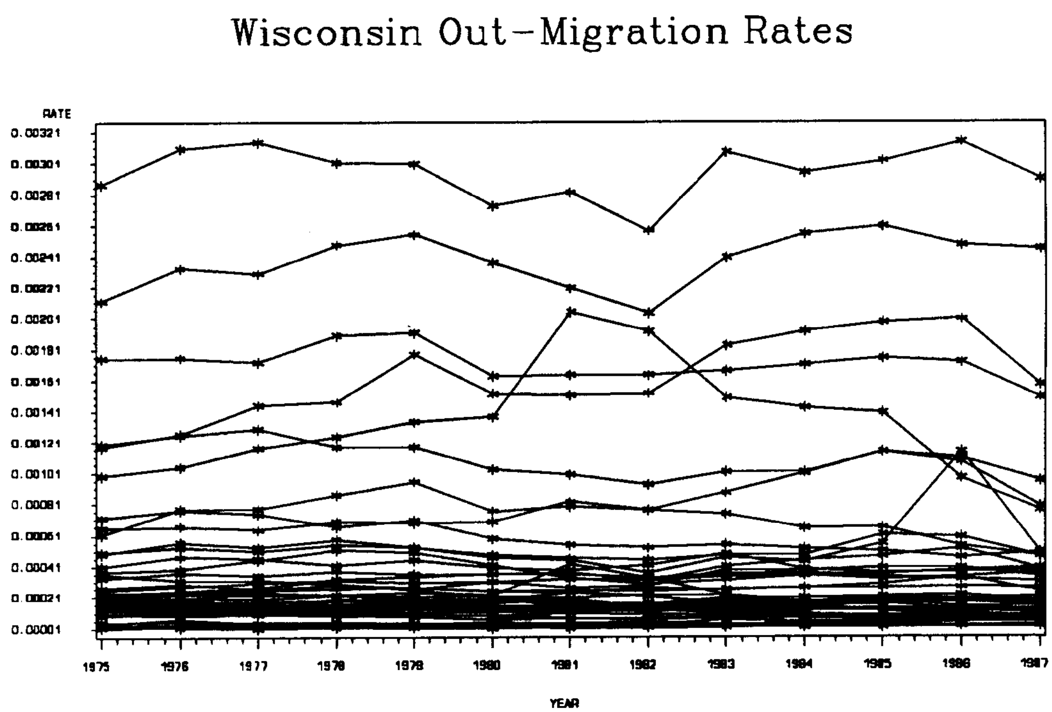



*FIGURE 2A.* *Time series plot from 1975 to 1987, inclusive.*

In analyzing variance functions for heteroscedastic and autocorrelated data, transformations are an important tool. Within the Box-Cox family of monotonic power transforms (cf., Cook and Weisberg, 1982, p. 60-61), the square root and logarithmic transforms are especially useful for rate data. It is possible to use the folded-power transforms since migration rates are bounded above by one. However, since the largest rate is less than 5% there is little advantage in considering this latter family of transformations. In Figure 2b is the same set of rates as in Figure 2a except plotted on the logarithmic scale. One advantage of the logarithmic transform is that changes in logged data can be interpreted as percentage changes of the untransformed data. From Figure 2b, note that the destination-states that have low mean levels experience higher variability than destination-states with high mean levels. I interpret this to mean that destination-states with low average levels experience higher percentage changes than destination-states with high average levels. Graphs for the square root transform and for other states can be found in Appendix C. These graphs also show that the use of simple transforms to approximate normality, as suggested by Binomial or Poisson distributions, do not seem to be a reasonable model for this data.

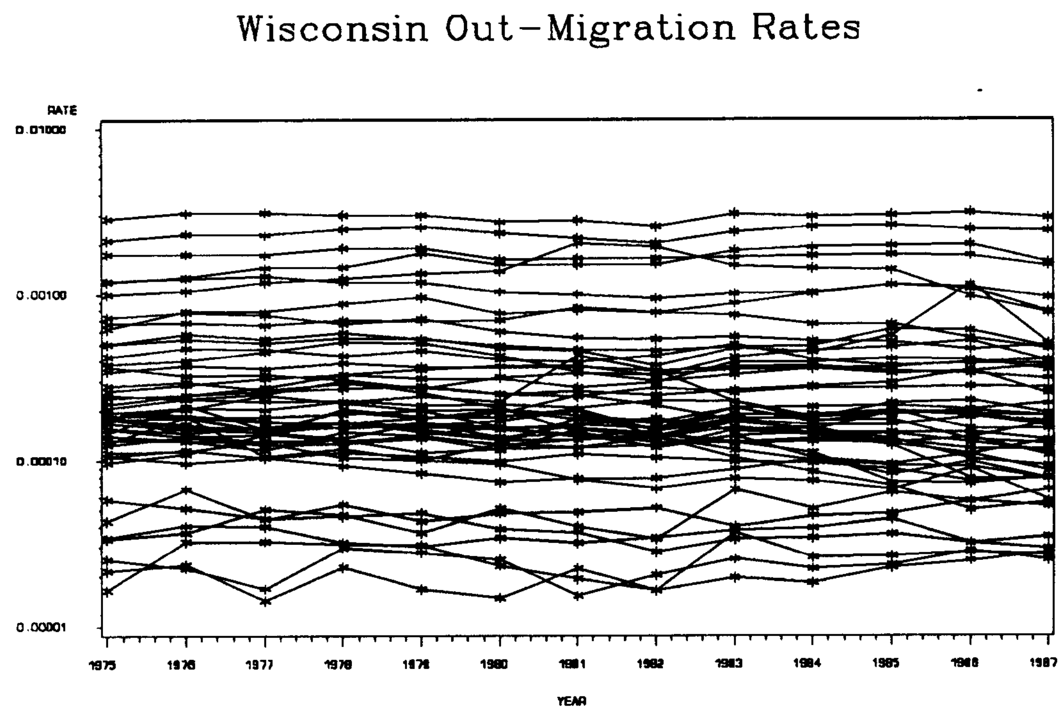



*FIGURE 2B.* *Data is plotted on logarithmic scale.*

To further investigate this relationship between average level and variability, consider a naive model,

$$Y_{o,d,t} = \mu_{o,d} + \sigma_{o,d}\, e_{o,d,t}\ , \tag{2.1}$$

where Y is the rate, or a transformed version, $\mu_{o,d}$ and $\sigma_{o,d}$ are parameters to be estimated, and {e} are i.i.d. mean zero, variance one error terms. Without assuming any functional relationships among the different mean and variance levels, use Y = R initially and let $\hat{\mu}_{o,d}$ and $\hat{\sigma}_{o,d}$ be the usual minimum variance unbiased estimates of $\mu_{o,d}$ and $\sigma^2_{o,d}$ , respectively. Call these time series means and variances since the averaging is done over time. In Figure 3 is a plot of $\hat{\sigma}_{o,d}$ versus $\hat{\mu}_{o,d}$ , indicating that the variability does increase as a function of the mean. Plots of $\hat{\sigma}_{o,d}$ versus state populations in Appendix C indicate some relationships but not as strong as would be suggested by Binomial and Poisson models.

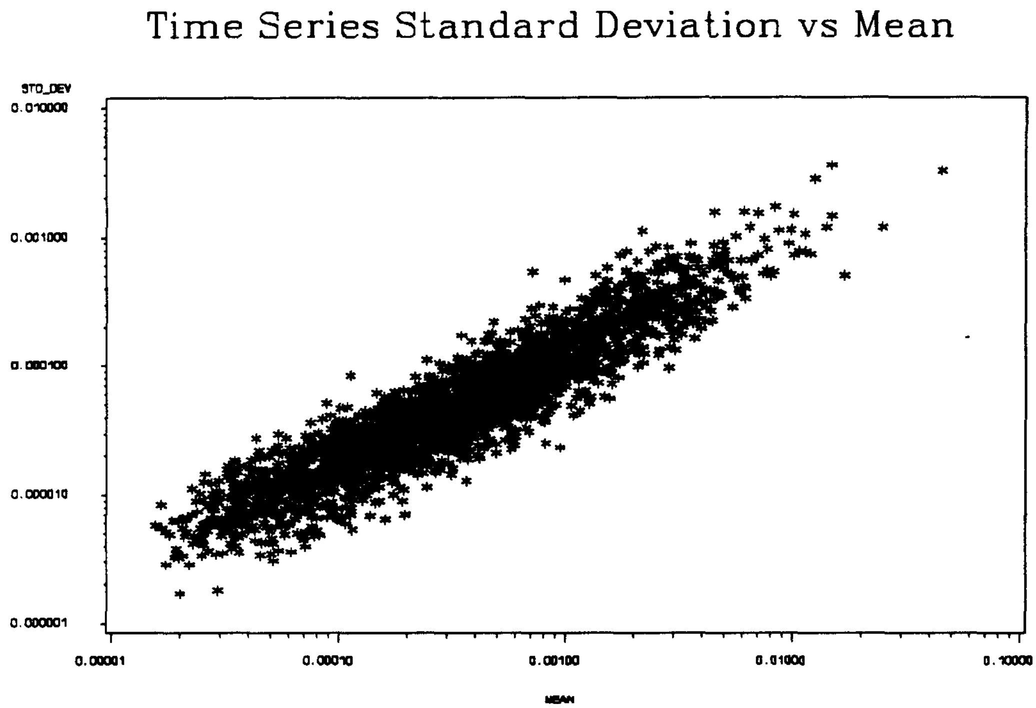



*FIGURE 3. Estimates are plotted on logarithmic scale.*

Diagnostic checks of the basic model (2.1) can be made using the standardized residuals,

$$\hat{e}_{o,d,t} = (Y_{o,d,t} - \hat{\mu}_{o,d}) / \hat{\sigma}_{o,d}. \tag{2.2}$$

Plots of the residuals indicate fewer, but still some, discernible patterns. For the origin state of Wisconsin in Figure 4a, few patterns are readily apparent. For the corresponding plot for the state of New York in Figure 4b, most series are positively autocorrelated and are related to one another. To further investigate the autocorrelation aspect, lag 1 residual autocorrelation coefficients $\hat{\rho}_{o,d}$ were computed for each o,d. A plot of these coefficients, in Figure 5, reveal some autocorrelation but no widespread patterns. Autocorrelation, and concomitant potential nonstationarity, is addressed further in Section 3.

Time Series Plot of Wisconsin Residuals

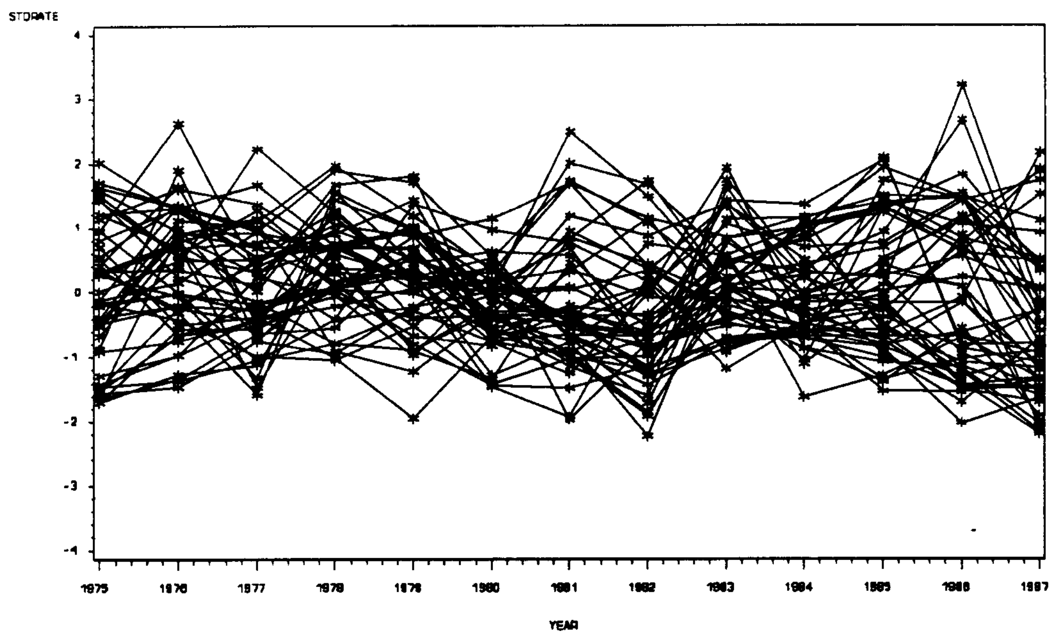



*FIGURE 4A*

Time Series Plot of New York Residuals

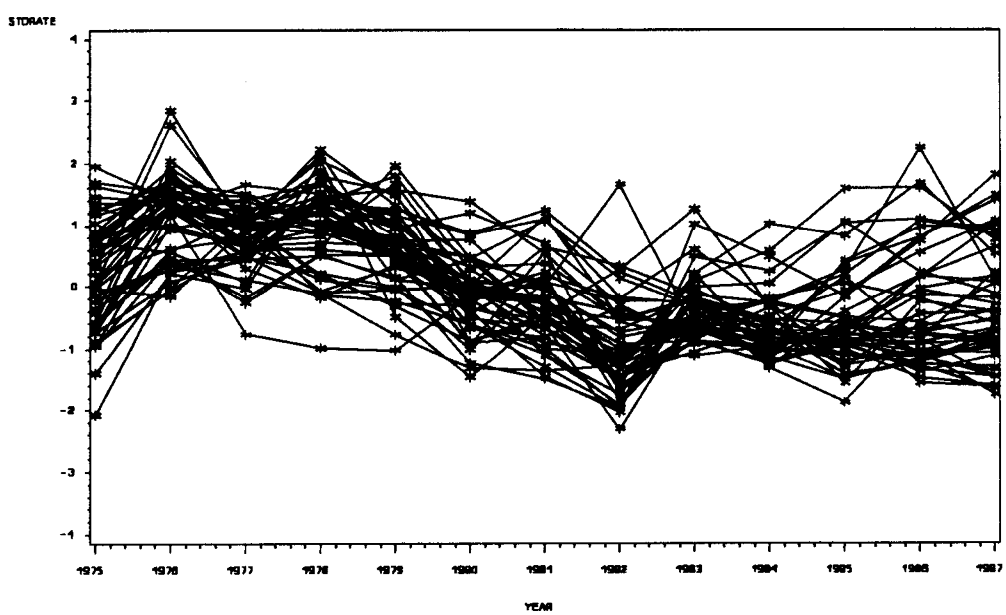



*FIGURE 4B*

Plot of Lag 1 Autocorrelation vs Population

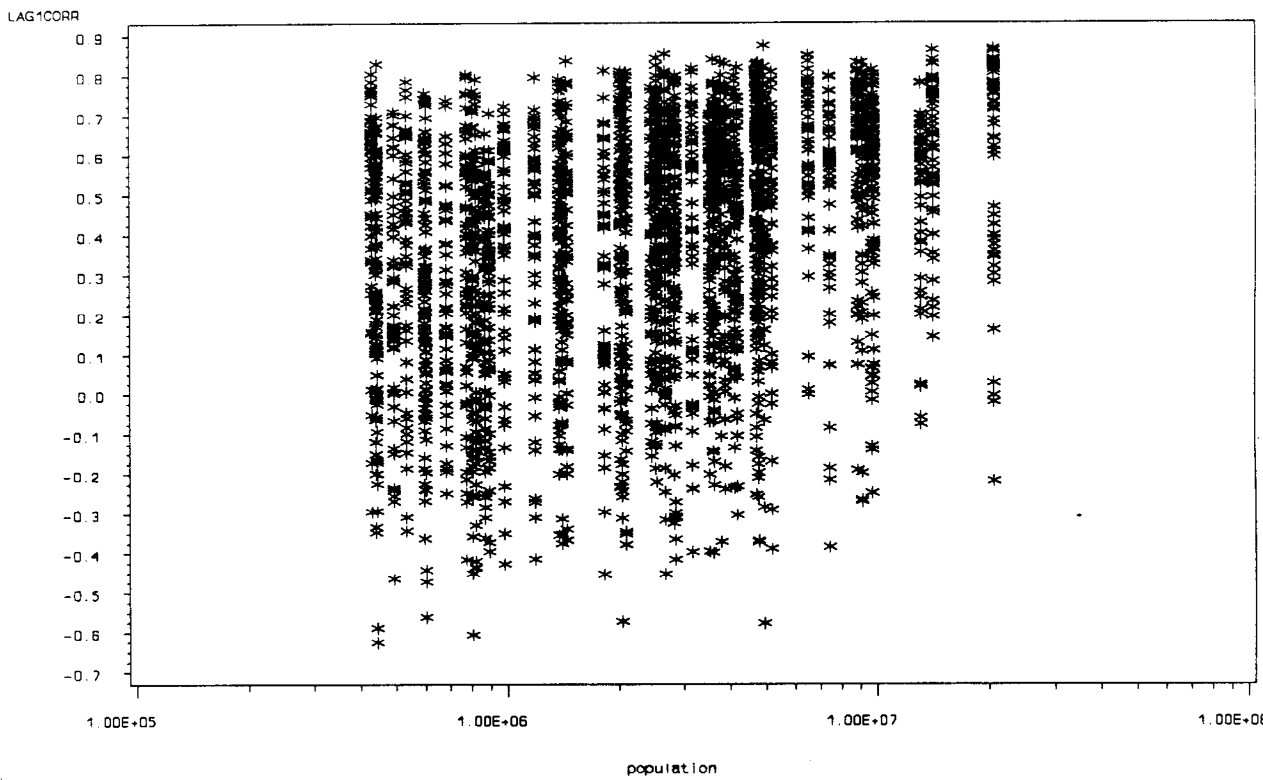



*FIGURE 5*

Residuals are also useful in providing insights into the choice of the appropriate transform. In Appendix C are plots of the residuals using rates, the square roots and logarithms for various selected states. These time series plots indicate that the choice of transformation seems to have little effect. This observation is more striking in Figure 6, where a plot of residuals from the logarithmic transform is on the vertical axis as compared to the similar residuals from the untransformed data on the horizontal axis. Because of the size of the full data set, only 1987 rates were plotted. This lack of effect is well-known to applied time series analysts where it is often observed that, if the variability of a time series is small compared to the mean, a transformation will have little effect (cf., Roberts, 1988).

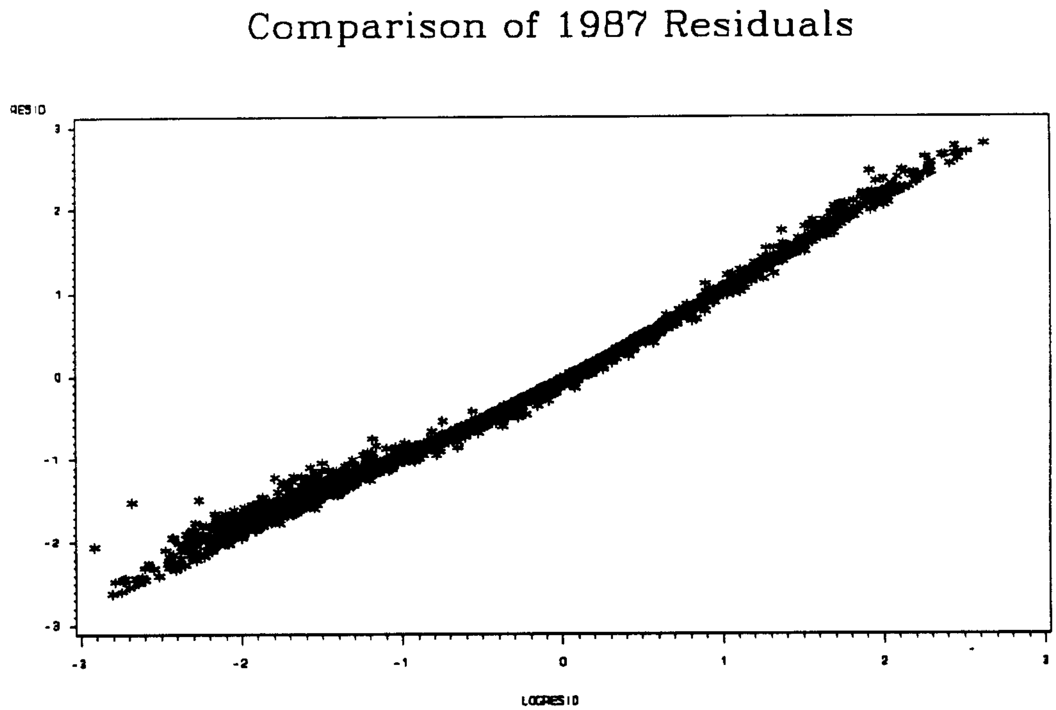



*FIGURE 6. Standardized residuals compared to standardized residuals from the logarithmic model.*

Following checks of heteroscedasticity and autocorrelation, the third dimension of residual checking is for contemporaneous correlation. In the migration literature, contemporaneous correlation can be viewed as one model specification for interdependencies among states, cf., Greenwood (1975b) for an alternative specification. These correlation parameters are unidentifiable in regression models without replication or ordering of observations. However, for longitudinal data, these correlations are regularly considered and, indeed, their presence is the reason for the optimality of the so-called "seemingly unrelated regression" (SUR) models when compared to ordinary least squares. Since there are 2,550 equations of the form in (2.1), there are in principle 2550 x (2550 - 1) / 2 = 3,249,975 contemporaneous correlation parameters to be estimated. Because this number of parameters is clearly not supported by the data set, the model building strategy is to build sets of equation having error terms without contemporaneous correlation but that includes parameters, or other features of the model, that account for observing contemporaneous correlation in the dependent variable. To check that the residuals of this model are not contemporaneously correlated, I use a diagnostic statistic in lieu of a graphical technique due to the potential number of correlations. When using statistics to check hypotheses, we are susceptible to nonlinearities in the data that may be obvious in a graph. However, computational compromises must be made.

A nonparametric, and distribution-free, measure of association between the $i^{th}$ and $j^{th}$ flows is $r_{ij}$, Spearman's rank correlation coefficient. More specifically, define $r_{ij} = s_{ij} / (s_{ii}\ s_{jj})^{1/2}$ , where $s_{ij} = (T-1)^{-1} \sum_t (R_{i,t} - (T+1)/2)\ (R_{j,t} - (T+1)/2)$ and $\{R_{i,1}, ..., R_{i,T}\}$ are the ranks of $\{\hat{e}_{i,1}, ..., \hat{e}_{i,T}\}$. In this application, it turns out that the positive contemporaneous correlations prevail, and thus I consider the statistic

$$R_{AVE} = \binom{M}{2}^{-1} \sum_{i<j} r_{ij}. \tag{2.3}$$

Here, M is the number of flows. The limiting distribution is known, that is, $(T-1)\ ((M-1)\ R_{AVE} + 1) = FR \rightarrow_D \chi^2_{(T-1)}$ as $M \rightarrow \infty$ where FR is Friedman's statistic, cf., Hettmansperger (1984, pp. 196, 210). To account for the fact that negative correlations may offset positive correlations, I also use

$$R^2_{AVE} = \binom{M}{2}^{-1} \sum_{i<j} r^2_{ij}. \tag{2.4}$$

It is easy to check that $E\ R^2_{AVE} = 1/(T-1)$ under the null hypothesis of no correlation. Further, the limiting distribution has been recently established in a companion paper, Frees (1990).

While $R^2_{AVE}$ and $R_{AVE}$ are statistics which summarize the entire correlation matrix, certain subsets may be of particular interest. For example, in this application, contemporaneous correlations having common states of origin or destination are important. Define $r_{d_1,d_2}(o)$ to be the Spearman's rank correlation coefficient between the flow having state of origin 'o' and state of destination '$d_1$' and the flow having state of origin 'o' and state of destination '$d_2$'. I use the statistics

$$R_{AVE,o} = (51 \cdot 50 \cdot 49/2)^{-1} \sum_{o,d_1,d_2} r_{d_1,d_2}(o), \tag{2.5}$$

and

$$R^2_{AVE,o} = (51 \cdot 50 \cdot 49/2)^{-1} \sum_{o,d_1,d_2} r_{d_1,d_2}(o)^2 \tag{2.6}$$

where $R_{AVE,d}$ and $R^2_{AVE,d}$ are defined similarly. Here, $\sum_{o,d_1,d_2}$ means the sum over $\{o, d_1, d_2\}$ all distinct with $d_1 < d_2$. For example, for the residuals in (2.2), it turns out that $R^2_{AVE} = 19.2\%$, $R_{AVE} = 11.7\%$, $R^2_{AVE,o} = 21.9\%$, $R_{AVE,o} = 21.5\%$, $R^2_{AVE,d} = 28.4\%$ and $R_{AVE,d} = 37.1\%$. I interpret these statistics as indicating that flows exhibit a great deal of contemporaneous correlation. Moreover, those flows having a common state of destination exhibit an even stronger relationship. It is precisely these issues that are addressed in subsequent sections.

## 3. MODELS OF OUT-MIGRATION RATES

For purposes of judging between alternative models, the first 11 time periods, 1975-1985 rates, are used for construction of the model. The last two time periods, 1986 and 1987 rates, are used for out-of-sample validation. Since the primary purpose of the model building is forecasting, when there is a conflict between in-sample and out-of-sample measures in judging alternative models, out-of-sample criteria will be used. (It is interesting to note the distinction that is drawn between the concepts of forecasts and projections in the demographic literature, cf., Keyfitz, 1972).

There are several candidates available for the choice of measure to be used in summarizing the out-of-sample performance. For each year, there are 2,550 forecasts to be compared to the held-out values. The usual least squares theory leads to minimizing the prediction error sum of squares. Carroll and Ruppert (1988, p. 61) remark that accounting for heteroscedasticity in the data often has a more dramatic effect on prediction intervals than point estimates. This suggests examining the performance of in-sample based prediction intervals as compared to actual out-of-sample performance. However, from a demographic perspective, the most important criteria are either the number, or percentage, of migrant forecast errors (cf., Keyfitz, 1972). More specifically, let $M_{o,d,t}$ be the actual number of migrants and $\hat{M}_{o,d,t}$ be the forecast number of migrants. I use the forecast error criterion

$$FE_t = \Sigma_{o,d} \mid \hat{M}_{o,d,t} - M_{o,d,t} \mid / \Sigma_{o,d} \; M_{o,d,t} \,. \tag{3.1}$$

Here, $\Sigma_{o,d}$ means the sum over distinct pairs {o,d}. An alternative criterion, employed by Isserman et al (1985), is (3.1) with gross migration replaced by net migration. With the choice of criterion in (3.1), the role of estimating variance parameters is smaller than would be the case if forecast intervals were the primary goal of the modeling procedure.

The choice of the forecast error criterion has important ramifications in identifying the structure of the model. Thinking of migrants as a rate times population of origin, using FE in (3.1) indicates that rate forecast errors for states with larger populations tend to dominate those of states with smaller populations. An important example can be seen in the structural identification of the temporal correlation of the data. From figure 5, note that more populated states tend to have higher lag 1 autocorrelations. While the largest of the lag 1 autocorrelations coefficients was less than 0.8 in absolute value, the median was approximately 0.33. In interpreting these coefficients, recall that they are bounded by one in absolute value and that the textbook standard error for each coefficient is $1/(T)^{1/2} = 13^{-1/2} \approx 0.28$. While it is difficult to say for a series of only 13 points, the need to difference the data to accommodate potential nonstationarity should not be ruled out. Another rule of thumb used to judge whether the data should be differenced is to compare the standard deviation of the differenced data with that of the original data (cf, Roberts, 1988). Of the 2,550 flows, it turned out that only 1,026 had smaller standard deviations when differenced. The average of the 2,550 standard deviations for the original data was smaller than corresponding average for the differenced data by 3%. However, when this average was weighted by population size, the weighted average for differenced data was smaller by 11%. Similar results were attained using logged data. This weighting was suggested by the forecast error criterion in (3.1). Note that in (3.1) absolute values are used whereas in the above rule of thumb the square root of a sum of squares, the standard deviation, was used. However, the primary raison d'être for any rule of the thumb is it's usefulness in selecting a desirable model. As described below, it turns out that the differenced data models were the most successful with respect to the criterion FE in (3.1).

## 3.1 SOME BASIC MODELS

To compute parameter estimates of each model described below, I used the generalized least-squares estimation technique. This was due to the overall size of the data set and the small number of observations available for each flow. For the autoregressive models, the first time point was used only as an explanatory variable, not as a dependent varible.

The first set of interesting models were as follows. Each model defined in this subsection is presented using the original and logged data. Define Models 1a and 1b by (2.1) using rates and differenced rates, respectively. This turned out to be somewhat overparametrized and thus define Models 2a and 2b by

$$Y_{o,d,t} = \sigma_{o,d}\, e_{o,d,t} \tag{3.2}$$

for differences and second differences of rates, respectively. Omitting the 2,550 mean parameters $\mu_{o,d}$ simplified the model considerably. Note that the variance parameters, $\sigma_{o,d}$, are not part of the linear forecasting equation and would not be expected to play an important role in the reduction of $FE_t$. In deciding to use rates or differenced rates, the AR(1) model can be viewed as a simple compromise. Thus, define Models 3a and 3b using

$$Y_{o,d,t} = \alpha_{o,d} + \beta_{o,d}\, Y_{o,d,t-1} + \sigma_{o,d}\, e_{o,d,t} \tag{3.3}$$

for rates and differenced rates, respectively. Again, this model turns out to be overparameterized and thus by dropping the intercept terms from (3.3), define Models 4a and 4b by

$$Y_{o,d,t} = \beta_{o,d} Y_{o,d,t-1} + \sigma_{o,d} e_{o,d,t} \tag{3.4}$$

for the original and differenced rates, respectively. Table 1 summarizes the performance of Models 1-4 with respect to the forecast error criterion in (3.1). Forecasts using the transfored data were transformed back to the original scale to compute the forecast error criterion. From Table 1, note that using logged or the original data again seems to matter little, as noted in Figure 6. As noted above, leaving out the intercept parameters significantly improved the performance, both in going from Models 1 to 2 and in going from Models 3 to 4. In all cases, the forecast error was higher for t=13 as compared to t=12, as one would expect. Of the models 1-4, the best models are Model 2a and 4b, that is, models of differenced data without the intercept term.

TABLE 1. FORECAST ERROR IN PERCENT

| | | Original Data | | Logged Data | |
|---|---|---|---|---|---|
| MODEL | DATA | $FE_{12}$ | $FE_{13}$ | $FE_{12}$ | $FE_{13}$ |
| 1a | no difference | 17.28 | 19.54 | 16.83 | 18.92 |
| 1b | $1^{st}$ difference | 8.70 | 13.07 | 8.63 | 13.01 |
| 2a | $1^{st}$ difference | 8.39 | 11.96 | 8.39 | 11.96 |
| 2b | $2^{nd}$ difference | 9.69 | 13.21 | 9.82 | 13.38 |
| 3a | no difference | 11.06 | 14.46 | 10.87 | 13.86 |
| 3b | $1^{st}$ difference | 9.44 | 13.41 | 9.23 | 13.06 |
| 4a | no difference | 8.65 | 12.72 | 8.63 | 13.04 |
| 4b | $1^{st}$ difference | 8.69 | 12.02 | 8.63 | 12.01 |

Models 1-4 encompass many of the alternatives that have been proposed for forecasting internal migration rates. Using model 1a with the original rates yields a forecast equal to the time series average of the flows. This model was used in 1989 as the basis for the "Series B" projections provided by the U. S. Census Bureau to members of the Federal-State Cooperative Program for Population Projections (FSCPP) (Signe Wetrogan, personal communication). Using model 2a with original rates yields a forecast equal to the most recent rate, the traditional time-invariant Markov assumption in demography. Other models can be interpreted as yielding a forecast equal to the most recent rate plus a trend factor. Using the logarithmic transform essentially means forecasting percentage changes. Thus, even with trend factors, the forecasts are constrained to lie in the 0 to 1 range for rates.

Many of the out-of-sample results in Table 1 could have been anticipated by in-sample performance. By in-sample performance, I mean checking residuals for contemporaneous correlation, autocorrelation and "average size." To measure contemporaneous correlation, use the $R_{AVE}$ and $R^2_{AVE}$ statistics defined in (2.3) and (2.4), respectively. Lag one residual autocorrelations were computed for each flow and it turned out that the distribution of the 2,550 coefficients was reasonably thin-tailed and symmetric. Thus, I only report the summary statistics, AR1MN and AR1SD, for the mean and standard deviation of this "distribution," respectively. This is done merely to summarize the results, not to insist nor suggest that a random coefficients model is appropriate, although this is certainly a possibility. I use $\hat{\sigma}_{o,d}$ to measure the "average size" of a residual in each flow and SUMSD = $\Sigma_{o,d}\ \hat{\sigma}_{o,d}$ as a summary measure. In accordance with the out-of-sample criterion in (3.1), it also seems reasonable to weight residuals by population size. Thus, I define the weighted average WGTSD = $\Sigma_{o,d}\ (P_o / 1000)\ \hat{\sigma}_{o,d} / 2{,}550$ where $P_o$ is the most recent in-sample origin population. The results of these in-sample measures for Models 1-4 are reported in Table 2.

TABLE 2. IN-SAMPLE PERFORMANCE OF MODELS 1-4

| MODEL | DATA | $R^2_{AVE}$ | Exp | $R_{AVE}$ | AR1MN | AR1SD | SUMSD | WGTSD |
|---|---|---|---|---|---|---|---|---|
| 1a | no difference | 19.40 | 10.00 | 12.15 | .319 | .309 | .265 | .309 |
| 1b | 1st difference | 13.47 | 11.11 | 10.93 | -.253 | .273 | .243 | .259 |
| 2a | 1st difference | 13.47 | 11.11 | 10.93 | -.253 | .273 | .243 | .259 |
| 2b | 2nd difference | 19.10 | 12.50 | 14.41 | -.495 | .208 | .382 | .399 |
| 3a | no difference | 13.20 | 11.11 | 11.19 | -.029 | .205 | .193 | .212 |
| 3b | 1st difference | 17.06 | 12.50 | 12.96 | -.418 | .207 | .269 | .294 |
| 4a | no difference | 20.47 | 11.11 | 13.12 | -.253 | .273 | .228 | .244 |
| 4b | 1st difference | 14.11 | 12.50 | 9.36 | -.112 | .167 | .211 | .229 |
| 1a | no difference - log | 19.40 | 10.00 | 12.15 | .320 | .310 | 436 | 609 |
| 1b | 1st difference - log | 13.60 | 11.11 | 11.22 | -.251 | .274 | 441 | 532 |
| 2a | 1st difference - log | 13.60 | 11.11 | 11.22 | -.251 | .274 | 441 | 532 |
| 2b | 2nd difference - log | 19.20 | 12.50 | 14.32 | -.498 | .209 | 714 | 844 |
| 3a | no difference - log | 13.26 | 11.11 | 11.39 | -.030 | .202 | 331 | 422 |
| 3b | 1st difference - log | 17.16 | 12.50 | 12.84 | -.420 | .209 | 469 | 599 |
| 4a | no difference - log | 20.47 | 11.11 | 13.12 | -.252 | .274 | 418 | 505 |
| 4b | 1st difference - log | 14.13 | 12.50 | 9.51 | -.115 | .166 | 373 | 461 |

Legend: Exp is the expected value of $R^2_{AVE}$. Under the DATA column, "- log" refers to data in logarithms.

Several conclusions emerge from Table 2. As noted above, there is little difference in models using the original rates and models using logarithmic rates. Because of the desirable property of yielding bounded forecasts noted above, I henceforth report only models using logarithmic rates. In examining the average size and autoregressive properties of the residuals, Models 3a and 4b seem to have the best performance. Simple differencing, going from model 1a to 1b, does not seem to be sufficient to remove temporal correlation effects. The AR(1) model with no differencing (Model 3a) has the best in-sample performance but has poor out-of-sample performance. From the viewpoint of the amount of contemporaneous correlation in the residuals, all Models 1 - 4 are inadequate. Recall that in interpreting the $R^2_{AVE}$ statistic that it should be compared to 1/(T-1) which varies by the amount of data available for fitting the model. My overall conclusion is that Models 3a and 4b had the best in-sample performance with Model 2a being a close third. Note, however, that there are 5,100 linear parameters in Model 3a. Thus, it is not surprising that it did not perform well on the out-of-sample criterion above in Table 1. This suggests the desirability of parsimonious models which are further discussed below.

## 3.2 SOME ALTERNATIVE AUTOREGRESSIVE MODELS

In model 4b, there are still 2,550 linear parameters. Thus, it is natural to inquire as to whether the performance could be improved by restricting the number of autoregressive coefficients. I consider

$$Y_{o,d,t} = \beta_o Y_{o,d,t-1} + \sigma_{o,d} e_{o,d,t} , \tag{3.5}$$

$$Y_{o,d,t} = \beta_d Y_{o,d,t-1} + \sigma_{o,d} e_{o,d,t} , \tag{3.6}$$

and

$$Y_{o,d,t} = \beta Y_{o,d,t-1} + \sigma_{o,d} e_{o,d,t} . \tag{3.7}$$

Models 5 - 7 can be interpreted as intermediate versions of Models 2 and 4. Due to the number of parameters, Models 5 - 7 offer more flexibility than Model 2 and are more parsimonious than Model 4. It seems reasonable to posit that the flows are related in some sense and, through Models 5 - 7, I investigate whether the flows share common parameters.

The in-sample and out-of-sample performance of Models 5 - 7 can be found in Tables 3 and 4, respectively. Although not reported here, the autoregressive coefficients turned out to be close to one for the models without differencing. Hence, each of these models performed similarly to Model 2a, both on an in and out-of-sample basis. Thus, I henceforth only discuss models of the differenced rates. In examining Table 3, one can see that models with more parameters provided a better fit, as expected. While model 4 performed the best in terms of accounting for the autocorrelation, Models 5, 6 and 7 are close seconds. In examining Table 4, all models performed similarly in terms of forecasting. At this point, Model 7b seems to be the choice based on the principle of parsimony.

Several variants of the autoregressive modeling and fitting scheme were investigated. Intercept terms were included in Models 5 - 7. The resulting in and out of sample performance of these models turned out to be slightly inferior. In the case of homoscedastic errors, it is well-known that certain biases arise when using least squares estimation techniques to fit autoregressive models. These biases can be particularly important in longitudinal data, cf., Hsiao (1986, p. 73). Generalized least square versions of alternative unbiased estimators, cf., Hsiao (1986, p. 75) were used to fit Models 5 - 7 with no improvement over the generalized least squares fitted models reported here. I also experimented with some shrinkage forecasts, as discussed in Garcia-Ferrer et al (1987) without any real success.

TABLE 3. IN-SAMPLE PERFORMANCE OF MODELS 2a, 4b and 5-7

| MODEL | DATA | $R^2_{AVE}$ | Exp | $R_{AVE}$ | AR1MN | AR1SD | SUMSD | WGTSD |
|---|---|---|---|---|---|---|---|---|
| 2a | 1st difference - log | 13.60 | 11.11 | 11.22 | -.251 | .274 | 441 | 532 |
| 4b | 1st difference - log | 14.13 | 12.50 | 9.51 | -.115 | .166 | 373 | 461 |
| 5a | no difference - log | 13.60 | 11.11 | 10.95 | -.251 | .274 | 441 | 533 |
| 5b | 1st difference - log | 14.16 | 12.50 | 10.13 | -.136 | .284 | 418 | 516 |
| 6a | no difference - log | 13.60 | 11.11 | 10.96 | -.251 | .274 | 441 | 533 |
| 6b | 1st difference - log | 14.18 | 12.50 | 10.25 | -.139 | .270 | 417 | 509 |
| 7a | no difference - log | 13.60 | 11.11 | 10.95 | -.251 | .274 | 441 | 533 |
| 7b | 1st difference - log | 14.14 | 12.50 | 10.17 | -.136 | .289 | 421 | 515 |

TABLE 4. FORECAST ERROR IN PERCENT

| MODEL | DATA | $FE_{12}$ | $FE_{13}$ |
|---|---|---|---|
| 2a | 1st difference - log | 8.39 | 11.96 |
| 4b | 1st difference - log | 8.63 | 12.01 |
| 5a | no difference - log | 8.48 | 11.94 |
| 5b | 1st difference - log | 8.57 | 11.97 |
| 6a | no difference - log | 8.36 | 12.05 |
| 6b | 1st difference - log | 8.64 | 11.97 |
| 7a | no difference - log | 8.43 | 11.85 |
| 7b | 1st difference - log | 8.70 | 12.03 |

## 3.3 TIME - VARYING COEFFICIENTS

From the high contemporaneous correlation coefficients in Table 3, it is evident that while the autoregressive models have addressed the autocorrelation aspect of the data, they have contributed little to our understanding of the contemporaneous correlation aspect. Again, the principle is that even though I am primarily interested in the point forecast criterion in (3.1), understanding the variance structure will presumably lead to more efficient estimates. The standard device, seemingly unrelated regression estimates, is not available even for the reduced Models 5 and 6 because of the small number of observations available for each flow.

An alternative is to assume that there are parameters common to the flows and that vary through time. Specifically, consider Model 8,

$$Y_{o,d,t} = \alpha_t + \sigma_{o,d}\, e_{o,d,t} \tag{3.8}$$

where $\{\alpha_t\}$ are parameters to be estimated. In this section, Y represents the first differences of the logarithm of the rates. Similar to the discussion in Section 3.2, it is also useful to consider Models 9 and 10,

$$Y_{o,d,t} = \alpha_{o,t} + \sigma_{o,d}\, e_{o,d,t} \tag{3.9}$$

and

$$Y_{o,d,t} = \alpha_{d,t} + \sigma_{o,d}\, e_{o,d,t}\ , \tag{3.10}$$

where $\{\alpha_{o,t}\}$ and $\{\alpha_{d,t}\}$ are parameters to be estimated. The detailed in-sample performance of Models 8 - 10 is in Table 5 where, for the reader's convenience, some of the summary statistics for Models 2a, 4b and 7b are restated. From Table 5, it is evident that the time-varying coefficients in Model 8 have reduced the contemporaneous correlation as measured through $R_{AVE}$ and $R^2_{AVE}$ compared to the models introduced in Section 3.1 and 3.2. Further, the in-sample fit, as measured through SUMSD and WGTSD, is better than Models 2a and 7b. It is not surprising that the in-sample fit of Model 4b is superior since it has 2,550 linear parameters. Model 8 does not, however, account for all aspects of contemporaneous correlation. As foreshadowed in Section 2, I use the summary statistics $R^2_{AVE,o}$, $R_{AVE,o}$, $R^2_{AVE,d}$ and $R_{AVE,d}$ to investigate correlation aspects of flows having common origin or common destination. To accommodate these correlation aspects, Model 9 is an extension of Model 8 in the sense that each flow with the same state of origin shares a common time-varying coefficient, and similarly for Model 10. From the summary statistics in Table 5, Model 10 provides the best in-sample performance based on the contemporaneous correlation statistics and the in-sample fit statistics.

TABLE 5. IN-SAMPLE PERFORMANCE OF MODELS 2a, 4b, 7b and 8-10

| MODEL | $R^2_{AVE}$ | $R_{AVE}$ | $R^2_{AVE,o}$ | $R_{AVE,o}$ | $R^2_{AVE,d}$ | $R_{AVE,d}$ | AR1MN | AR1SD | SUMSD | WGTSD |
|---|---|---|---|---|---|---|---|---|---|---|
| 2a | 13.60 | 11.22 | 15.49 | 16.58 | 18.22 | 23.02 | -.251 | .274 | 441 | 532 |
| 4b | 14.13 | 9.58 | 16.04 | 15.27 | 18.42 | 22.12 | -.115 | .166 | 373 | 461 |
| 7b | 14.14 | 10.07 | 15.85 | 15.53 | 19.47 | 24.25 | -.136 | .289 | 421 | 515 |
| 8 | 11.78 | 0.08 | 12.48 | 5.62 | 14.74 | 13.89 | -.205 | .291 | 417 | 486 |
| 9 | 11.83 | 0.19 | 12.32 | -0.09 | 15.22 | 15.05 | -.221 | .291 | 404 | 463 |
| 10 | 12.11 | 0.03 | 13.01 | 7.45 | 12.58 | -0.98 | -.283 | .262 | 386 | 435 |

Unfortunately, Models 8 - 10 can not be used directly to forecast future flows since the time-varying coefficients are unidentifiable for future values of t. One way to circumvent this problem is to assume that the coefficients are random. Thus, the above can be viewed as estimates of realizations of an exogeneous process. The model I entertain is the simplest possible: I assume that the time-varying coefficients follow a white noise process. Specifically, I use $\alpha_t = \alpha + u_t$, where $\alpha$ is a fixed parameter to be estimated and $\{u_t\}$ is an i.i.d. process with variance $\sigma^2(\alpha)$ that is independent of $\{e_{o,d,t}\}$. Similar assumptions were made to extend Models 9 and 10 to the random coefficients case. Assuming normally distributed random variables, Models 8 - 10 were re-fit using maximum likelihood estimation. To distinguish between generalized least squares and maximum likelihood, let 'Model 8a' be the fixed coefficients model in equation (3.8) and 'Model 8b' be the corresponding random coefficients model, and similarly for Models 9 and 10. From these estimates, the resulting forecasting performance is detailed in Table 6. Based on these statistics, the choice appears to be Model 8b.

TABLE 6. FORECAST ERROR IN PERCENT

| MODEL | $FE_{12}$ | $FE_{13}$ |
|---|---|---|
| 2a | 8.39 | 11.94 |
| 4b | 8.63 | 12.00 |
| 7b | 8.67 | 11.93 |
| 8b | 8.53 | 11.64 |
| 9b | 8.68 | 12.00 |
| 10b | 8.42 | 12.09 |

## 4. FORECAST EVALUATION AND CONCLUDING REMARKS

From the exploratory Section 2 and the modeling Section 3, Models 2a, 7b and 8b emerge as desirable candidate models. In this section, I interpret and evaluate the forecasts from these models.

First, note that these are models of destination-specific out-migration rates and that the forecasts of these models are rates. Hence, to evaluate the forecast error criterion in (3.1), a base year population is required. I consistently use the most recent population available. In cases where this population is known, as in $FE_{12}$ and $FE_{13}$, this reduces to evaluating a one-step forecast error. In cases where the population is not known, as in the multi-step projections below, there are clearly other alternatives one might consider. One could sequentially update the state population based on forecasts of internal migration alone. Alternatively, this update could be based on forecasts of internal and external migration, fertility and mortality. Because this paper is concerned with short-term forecasts and because the level of net migration is relatively low for most states (see Figures 8a and 8b below), future populations were held constant in computing forecast errors. While this is appropriate for evaluating forecast performance, it does result in one uncomfortable fact. Forecasts in a given year of net migration flows do not sum to zero, as they are constrained to by definition. While this could be accommodated for with a scaling factor, since population is an exogenous factor in the modeling, this modification was not performed.

In Figure 7 is a time series plot of actual 1975 - 1987 migration rates with forecasts from Models 2a, 7b and 8b. Here, "migration rate" is as in Figure 1b, that is, the number of migrants as a percentage of the population. For each model, forecasts of rates were made and then the population definition above was used to compute the forecast migration rate. From Figure 7, we see that forecasts from Models 2a and 7b are similar when compared to those of Model 8b. For Model 8b we see a downward drift because it turned out that the estimate of the mean of the distribution of $\{\alpha_t\}$ was negative. This does not seem to be an unreasonable extrapolation based on the time series plot in Figure 7. The interpretation is of yet more interest. In Model 8b, it is easy to check that long-term forecasts tend to zero internal migration. This is in contrast to the case of Models 2a and 7b, where long-term forecasts are equal to the most recent rate or tend to a constant level which is close to the most recent rate.

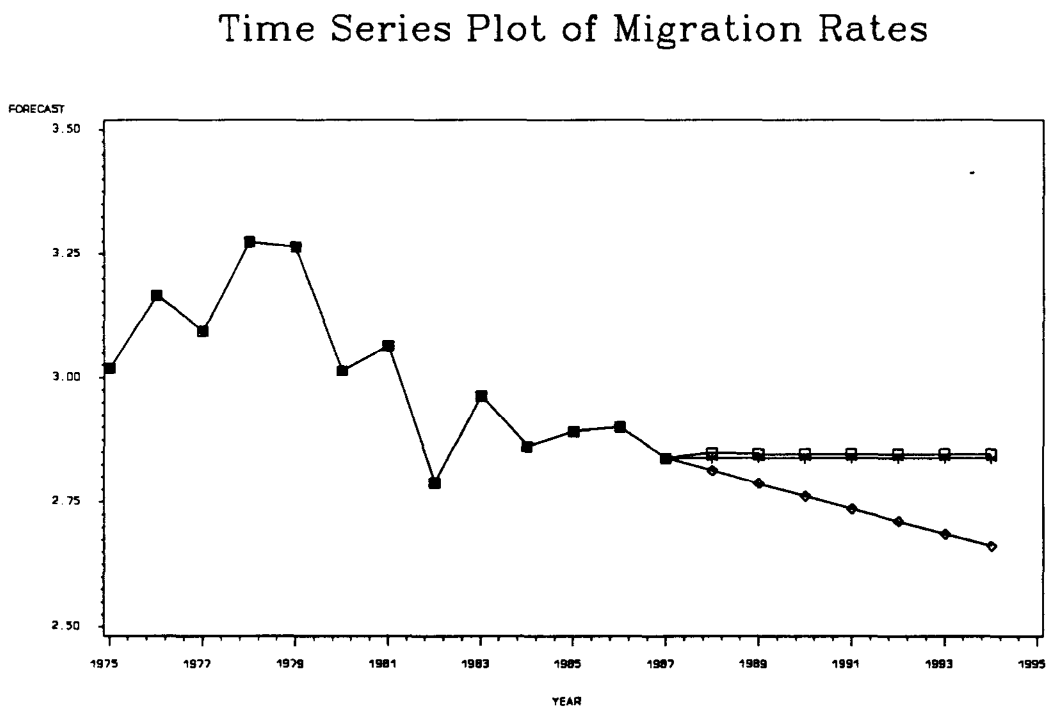



*FIGURE 7. 1975 - 1987 are actual rates, 1988 - 1994 are forecasts. The upper, middle and lower lines are from Models 7b, 2a and 8b, respectively.*

How do these alternative models affect population projections of states in the short-term? In Figures 8a and 8b are time series plots of the net migration rate by state. Years 1975 to 1987 are actual rates and, for years 1988 to 1994, the forecasts for Model 7b are in Figure 8a and the forecasts for Model 8b are in Figure 8b. Forecasts for Model 2a were nearly identical to those of Model 7b and are not included here. Indeed, despite the dramatic difference in long-term forecasts described in the paragraph above, short-term forecasts for Models 7b and 8b were close. When examining all 50 states, the largest difference was in 1994 which was less than 0.5%.

## Net Migration Rates by State – Model 7b

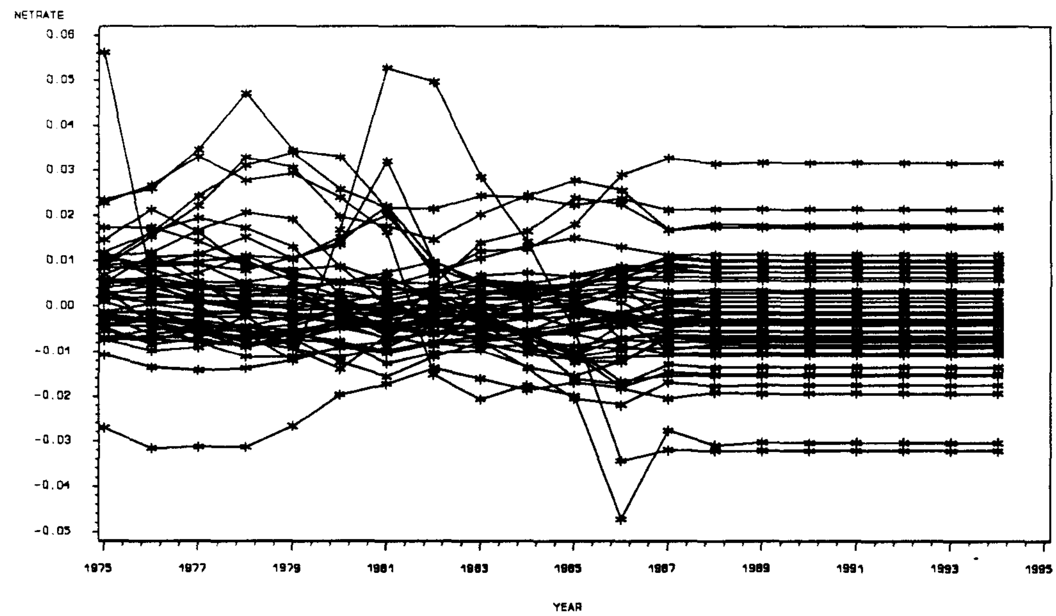



*FIGURE 8A.* *1975 - 1987 are actual rates, 1988 - 1994 are forecasts. Net rate is annual migration change divided by initial state population.*

## Net Migration Rates by State – Model 8b

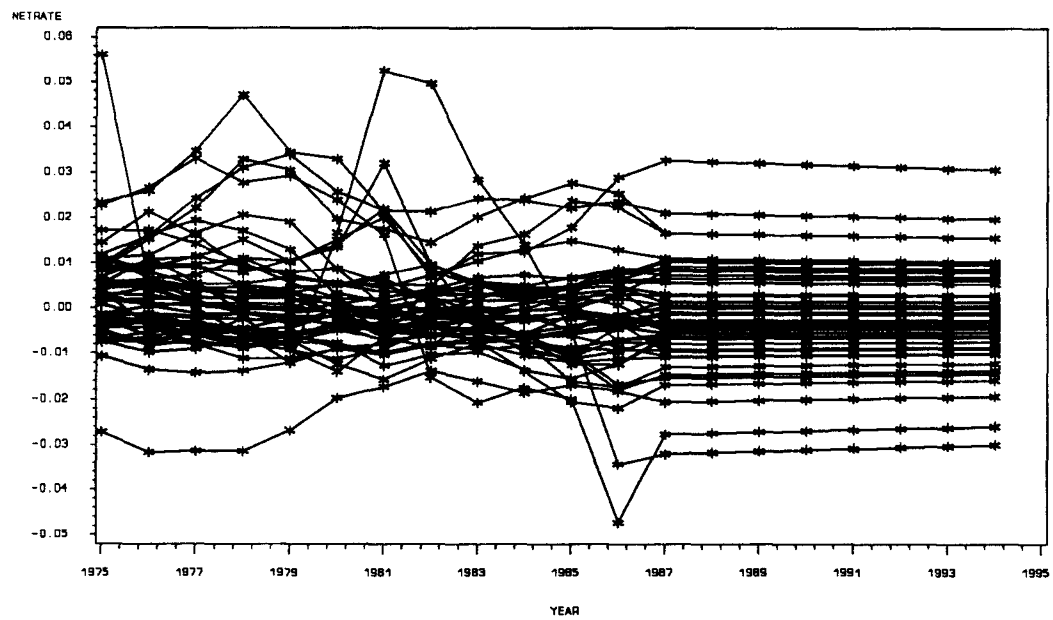



*FIGURE 8B.* *1975 - 1987 are actual rates, 1988 - 1994 are forecasts. Net rate is annual migration change divided by initial state population.*

Although I have stressed a point estimate criterion for model selection, a desirable feature of stochastic models of migration rates is that forecast intervals can easily be generated. As an example, in Figure 9 are actual and forecast net migration rates for the state of Wisconsin. The point forecasts are generated using Model 7b, that is, the integrated autoregressive model. The upper and lower bands represent approximate 95% prediction intervals, assuming Gaussian errors. Note that the prediction interval bandwidth increases as the forecast lead time increases. I consider this to be a desirable attribute of models of differenced rates, signifying our decreasing ability to reliably forecast long-term migration rates. These bands were calculated assuming the errors in Model 7b are uncorrelated through time but not excluding contemporaneous correlation. More specifically, I assume only that Corr ( $e_{o_1,d_1,t_1}$ ,$e_{o_1,d_1,t_1}$ ) = 0 for all $t_1 \neq t_2$ where each of $o_1$, $d_1$, $o_2$, $d_2$, ranges from 1 to 51. This model allows features of autocorrelation and contemporaneous correlation in the errors. Both of these features were evident in the exploratory analysis of Section 2. I also calculated prediction intervals assuming no contemporaneous correlation, i.e., Model 7b. Interestingly, these bands were only about half as wide as those allowing for contemporaneous correlation. The details of the calculation of the forecast intervals are in Appendix B.

Wisconsin Net Migration Rates

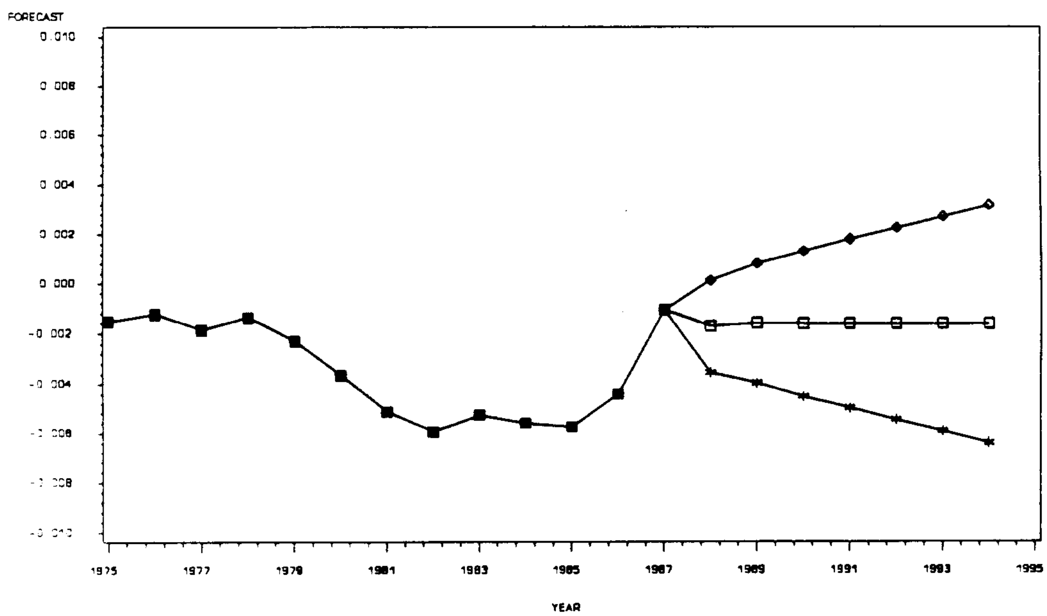



*FIGURE 9. Point forecasts are from Model 7B. The middle line represents the point forecast, and the upper and lower lines yield an approximate 95% confidence band.*

For short-term forecasts of rates, at this level of detail, there is no demographic, economic or geographic theory that dominates reasoning in selecting a model. Statistical criteria, such as graphical aids, diagnostic statistics, and in and out of sample summary measures, are used in this paper to explore the data and identify an appropriate model. Demographic considerations are especially relevant when interpreting long run forecasts of the models. Qualitative characteristics of long-term forecasts are important to consider in model selection since the Census Bureau uses the data to make population projections into the year 2010. Although the statistical model does not support such long-term projections, it is convenient to have a model that is consistent with demographic theory. Perhaps the most important conclusion of this study is that changes in rates are more stable than rates themselves and hence are more suitable for modeling and forecasting. Hence, confidence bands for forecasts increase as time increases in lieu of approaching an asymptotic level and, under the random walk and autoregressive models, the most recent rate plays an important role in long as well as short-term forecasts. Under the random coefficients models, forecasts for the long-term are zero internal migration, a demographic model often used as a benchmark to compare several projections. Somewhat surprisingly, the statistical criteria was not sensitive to the transformation of rates and hence, a logarithmic transformation was used. Although unimportant for short-term forecasts, this transformation had the desirable effect of constraining long-term forecasts to lie between zero.and one. Coupled with the declining forecasts under Model 8b, it is what produces an ultimate forecast of zero internal migration.

## REFERENCES

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

Breusch, T. S. and Pagan, A. R. (1980), "The Lagrange multiplier test and its applications to model specification in econometrics," *Review of Economic Studies* 47, 239-253.

Carroll, R. J. and Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman-Hall.

Cook, R. D. and Weisberg, S. (1982), *Residuals and Influence in Regression*, Chapman-Hall.

Dielman, T. E. (1983), "Pooled cross-sectional and time series data: A survey of current methodology," *American Statistician* 37, 111-122.

Engels, R., and Healy, M. K. (1981), "Measuring interstate migration flows: An origin-destination network through Internal Revenue Service Records," *Environment and Planning A* 13, 1345-1360.

Ericksen, E. P. and Engels, R. A., Editors, (1977), *Report of the Conference on Economic and Demographic Methods for Projecting Population*, American Statistical Association, Washington, D. C.

Frees, E. W. (1990), "Assessing contemporaneous correlations in longitudinal data," submitted to *Econometric Theory.*

Garcia-Ferrer, A., Highfield, R. A., Palm, F. and Zellner, A. (1987), "Macroeconomic forecasting using pooled international data," *Journal of Business and Economic Statistics* 5, 53-67.

Greenwood, M. J. (1975a), "Research on internal migration in the United States: A survey," *Journal of Economic Literature* 13, 397-433.

Greenwood, M. J. (1975b), "A simultaneous equations model of urban growth and migration," *Journal of the American Statistical Association* 70, 797-810.

Greenwood, M. J. (1981), *Migration and Economic Growth in the United States: National, Regional and Metropolitan Perspectives*, Academic Press, New York.

Hajj, H. (1975), "Internal migration models and population forecasts," *Journal of the Urban Planning and Development Division. Proceedings of the American Society of Civil Engineers* 101, UP2 (Nov.), 201-215.

Hettmansperger, T. P. (1984), *Statistical Inference Based on Ranks*, Wiley, New York.

Hsiao, C. (1986), *Panel Data*, Cambridge University Press, New York.

Isserman, A. M., Editor, (1986), *Population Change and the Economy: Social Science Theory and Methods*, Kluwer-Nijhoff, Boston, MA.

Isserman, A. M., Plane, D. A. and McMillen, D. B. (1982), "Internal migration in the United States: An evaluation of federal data," *Review of Public Data Use* 10, 285-311.

Isserman, A. M., Plane, D. A., Rogerson, P. A. and Beaumont, P. M. (1985), "Forecasting interstate migration with limited data: A demographic-economic approach," *Journal of the American Statistical Association* 80, 277-285.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T. C. (1985), *The Theory and Practice of Econometrics*, Chapter 13, Wiley, New York.

Keyfitz, N. (1972), "On future population," *Journal of the American Statistical Association* 67, 347-363.

Laird, N. M. and Ware, J. H. (1982), "Random-effects models for longitudinal data," *Biometrics* 38, 963-974.

Long, J. F. (1977), "Prospects for a composite demographic-economic model of migration for subnational population projections," *Report of the Conference on Economic and Demographic Methods for Projecting Population*, pp. 153-169. American Statistical Association, Washington, D. C.

Long, J. F. (1983), "The effects of college and military populations on models of interstate migration," *Socio-Economic Planning Sciences* 17, No. 5-6, pp. 281-290.

Long, J. F. and Wetrogan, S. I. (1985), "Creating annual state-to-state migration flows with demographic detail by merging census, survey, and administrative record tabulations," Paper presented at the annual meeting of the Population Association of America, San Francisco, CA, April 2-5.

Long, J. F. and Wetrogan, S. I. (1986), "The stability of interstate migration flows: The relative utility of demographic, human capital, and ecological perspectives," Paper presented at the annual meeting of the Population Association of America, Boston, MA, March 28-30.

Plane, D. A. (1987), "The geographical components of change in a migration system," *Geographical Analysis* 19, 283-299.

Plane, D. A. and Rogerson, P. A. (1985), "Economic-demographic models for forecasting interregional migration," *Environment and Planning A*, 17, 185-198.

Plane, D. A. and Rogerson, P. A. (1986), "Dynamic flow modeling with interregional dependency effects: An application to structural change in the US migration system," *Demography* 23, 91-104.

Plane, D. A., Rogerson, P. A. and Rosen, A. (1984), "The cross-regional variation of in-migration and out-migration," *Geographical Analysis* 16, 162-175.

Rao, C. R. (1987), "Prediction of future observations in growth curve models," *Statistical Science* 2, 434-471.

Roberts, H. V. (1988), *Data Analysis for Managers*, The Scientific Press.

Rogers, A. and Willekens, F. J. (1986), *Migration and Settlement, 1986*, D. Reidel Publishing Co., Dordrecht, Netherlands.

Rogerson, P. A. (1984), "New directions in the modelling of interregional migration," *Economic Geography* 60, 111-121.

Siegel, J. S. and Hamilton, C. H. (1952), "Some considerations in the use of the residual method of estimating net migration," *Journal of the American Statistical Association*, 475-500.

Shyrock, H. S., Siegel, J. and Associates (1976), *The Methods and Materials of Demography*, (Ed., E. G. Stockwell), Academic Press, New York.

Ter Heide, H. (1963), "Migration models and their significance for population forecasts," *The Milbank Memorial Fund Quarterly* 41, No. 1, 56-76.

U. S. Bureau of the Census (1988), "Projections of the populations of states, by age, sex and race: 1988 to 2010," Current Population Reports, Series P-25, No. 1017. U. S. Government Printing Office, Washington, D. C.

Ware, J. H. (1985), "Linear models for the analysis of longitudinal studies," *American Statistician* 39, 95-101.

# APPENDIX A. DESCRIPTION OF THE DATA

For researchers from functional fields who are interested in making causal inferences about the data, it is important to consider the manner in which the data was created. As noted in the introduction, the data was created from IRS matched administrative records. The data set has been recently revised and updated and provides enough detail to offer researchers an opportunity to investigate empirically many substantive issues that could only be speculated about before. However, like most large data sets there are biases which may be important for the application at hand. To understand these potential biases, in this Appendix I provide an overview of the creation of the data set. Further accounts can be found in Engels and Healy (1981) and Isserman, Plane and Rogerson (1982) . These two articles discuss the quality of the IRS data set as compared to other sources of information on internal migration. Specifically, these alternative sources are estimates from the Census Bureau's Decennial Census and Current Population Survey and from the Social Security Administration's Continuous Work History Sample. The data set considered in this paper is new in the sense that many of the gaps in time noted by the above researchers have been filled in. Further, the data set is longer (now 13 years) and thus permits an in-depth examination of the temporal patterns in the data.

To get an idea of the magnitude of the administrative data processing task, there were approximately 97 million returns representing 214 million persons in the 1985 Tax year. This data was forwarded to the Census Bureau from the IRS on 132 computer tapes. The returns are due at the IRS office on April 15 following the tax year. In a typical year, the Census Bureau receives information up to and including the $39^{th}$ week following April. This represents about 95% of the returns and about 88% of the population. The 1986 migration rate is based on a match of the 1985 Tax year return to the 1986 Tax year return. The 1986 Tax year return actually represents an address in the first quarter of 1987 for most filers. Thus, the Census Bureau receives the information to compute the 1986 rate in about November of 1987 and actually makes the computation in early 1988. After the state-to-state flows are computed by the Census Bureau, the summarized data is forwarded to the IRS where it is available to the public.

Returns are matched based on social security numbers of the primary filer. Residence is identified by mailing address listed on the return. Often filers use their tax preparer's mailing address or college students use their parents' mailing address. This is a potential source of bias which is thought to be minor when considering migration at the state level. Of course, it could become more important at the county or metropolitan level.

Because returns are on a household in lieu of an individual basis, it is difficult to retain demographic, i.e., age, sex and race, information on filers. Indeed, because migration data is based on IRS returns, disaggregating the data may hurt more than help. For example, many elderly are legally poor and hence are not required to file returns. It is estimated that only 30% of the elderly (65 and over) file returns. Similar problems of population coverage are known to occur for various age-sex-race cohorts. For state population projections, the argument is that these segments of the population are small compared to the total population and that their migration patterns may not differ that much from the overall population, especially since international migration, e.g., Puerto Rico, is not considered here. However, for other investigations this may be a crucial point.

The final caveat concerns data collection procedures which change over time. To a certain extent, one would like these procedures to be consistent over time even if there exists certain biases. However, as is the typical case, procedures do change and this should influence interpretations of the results of any modeling efforts. For example, beginning with the 1987 Tax year, it is no longer possible for a person to file a tax return and still be claimed as an exemption on another person's return without notifying the IRS. In the 1980 Tax year, there were an estimated 2.1 million duplicate exemptions, primarily children who had enough income to be required to file a return. For an investigation using age as covariate information, the change in the handling of duplicate exemptions could represent an important source of bias.

## APPENDIX B. INTEGRATED AUTOREGRESSIVE MODEL FORECAST INTERVALS

In this Appendix, I develop the formulae for the forecast intervals used in Figure 9. The integrated autoregressive model with contemporaneous correlations is assumed. Thus, $Y_{o,d,t} = \beta\, Y_{o,d,t-1} + \sigma_{o,d}\, e_{o,d,t}$ where Y is the difference of the logged rates, $Y_{o,d,t} = \log(R_{o,d,t}) - \log(R_{o,d,t-1})$. Let i denote the forecast lead time, T denote the latest time available and $E_T$ be the expectation conditional on data available up to and including time T. Some useful preliminary calculations are as follows. Recursively substitute into the model definition to get

$$Y_{o,d,T+i} = \beta^i\, Y_{o,d,T} + \sigma_{o,d}\, (\, \Sigma_{j=1}^{i}\ \beta^{i-j}\, e_{o,d,T+j}\, ). \tag{B.1}$$

Now, with $E_T\, Y_{o,d,T+i} = \beta^i\, Y_{o,d,T}$, we have

$$E_T \log(R_{o,d,T+i}) = \log(R_{o,d,T}) + \ Y_{o,d,T}\ \Sigma_{j=1}^{i}\ \beta^j\ = \log(R_{o,d,T}) + \ Y_{o,d,T}\ \beta(1-\beta^i)/(1-\beta).$$

From this equation, the point forecast of the destination-specific outmigration rate $R_{o,d,T+i}$ is defined as

$$F_{o,d,i} = R_{o,d,T}\, (R_{o,d,T}/R_{o,d,T-1})^{\beta(1-\beta^i)/(1-\beta)}\ .$$

Define $P_o$ to be the population of the $o^{th}$ state at time T. Recall the assumption that forecast populations are assumed to remain constant for future years, an easily modified assumption. Let 'w' be the index for the state under consideration, for example, in my numbering scheme w=50 for the state of Wisconsin. Then the i-step point forecast for Wisconsin net migration rate is

$$NMF_{w,i} = (\, \Sigma_{o\neq w}\, P_o\, F_{o,w,i}\ - P_w\, \Sigma_{d\neq w}\ F_{w,d,i}\, )\, /\, P_w. \tag{B.2}$$

Here, $\Sigma_{o\neq w}$ means the sum of o over {1, ..., 51} but $o\neq w$. Now, the forecast error for the destination-specific outmigration point forecast $F_{o,d,i}$ is $FE_{o,d,i} = R_{o,d,T+i} - F_{o,d,i}$. Similarly to (B.1), after some algebra, we have

$$\Sigma_{j=1}^{i}\, Y_{o,d,T+j} = (Y_{o,d,T})\beta(1-\beta^i)/(1-\beta) + \sigma_{o,d}\, (\, \Sigma_{j=1}^{i}\ \ e_{o,d,T+j}\, (1-\beta^{i+1-j})\, /(1-\beta)).$$

Thus, with $R_{o,d,T+i}\ = R_{o,d,T}\, \exp(\, \Sigma_{j=1}^{i}\, Y_{o,d,T+j}\, )$, we have

$$R_{o,d,T+i}\ = F_{o,d,i}\, \exp\{\sigma_{o,d}\, (\, \Sigma_{j=1}^{i}\ \ e_{o,d,T+j}\, (1-\beta^{i+1-j})\, /(1-\beta))\ \}.$$

This yields

$$FE_{o,d,i} = F_{o,d,i} \left(\exp\{\sigma_{o,d} \left( \Sigma_{j=1}^{i}\ e_{o,d,T+j}\ (1-\beta^{i+1-j})\ /(1-\beta)\right) \} - 1\right). \tag{B.3}$$

As in (B.2), we are now in a position to define the i-step net migration forecast error for the $w^{th}$ state,

$$NMFE_{w,i} = \left( \Sigma_{o\neq w}\ P_o\ FE_{o,w,i}\ -\ P_w\ \Sigma_{d\neq w}\ FE_{w,d,i} \right) / P_w. \tag{B.4}$$

To compute the bias term $E_T\ NMFE_{w,i}$, let G(t) be the moment generating function of the i.i.d. sequence $\{e_{o,d,t}\}$. Now,

$$E_T\ FE_{o,d,i} = F_{o,d,i}\ \{ \Pi_{j=1}^{i}\ G(\ \sigma_{o,d}(1-\beta^j)/(1-\beta))\ -\ 1\} \tag{B.5}$$

and thus,

$$\begin{aligned} E_T\ NMFE_{w,i} = (\Sigma_{o\neq w}\ (P_o/P_w)\ F_{o,w,i}\ \{ \Pi_{j=1}^{i}\ G(\ \sigma_{o,w}(1-\beta^j)/(1-\beta))\ -\ 1\} \\ -\ \Sigma_{d\neq w}\ F_{w,d,i}\ \{ \Pi_{j=1}^{i}\ G(\ \sigma_{w,d}(1-\beta^j)/(1-\beta))\ -\ 1\}) . \end{aligned} \tag{B.6}$$

Now, to compute $Var_T\ NMFE_{w,i} = E_T\ (NMFE_{w,i})^2 - (E_T\ NMFE_{w,i})^2$, we have

$$Var_T\ FE_{o,d,i} = (F_{o,d,i})^2\ \{ \Pi_{j=1}^{i}\ G(\ 2\sigma_{o,d}(1-\beta^j)/(1-\beta))\ -\ \Pi_{j=1}^{i}\ G^2(\ \sigma_{o,d}(1-\beta^j)/(1-\beta))\ \} . \tag{B.7}$$

To approximate covariances, I use the approximation that correlations are stable under transformations, that is, for random variables $X_1$ and $X_2$, Corr $(\exp(X_1), \exp(X_2)) \approx$ Corr $(X_1, X_2)$. Thus, since

$$Corr(\sigma_{o_1,d_1}\{ \Sigma_{j=1}^{i}\ e_{o_1,d_1,T+j}\ (1-\beta^{i+1-j})\}, \sigma_{o_2,d_2}\{ \Sigma_{j=1}^{i}\ e_{o_2,d_2,T+j}\ (1-\beta^{i+1-j})\}) = Corr(\ e_{o_1,d_1,t}, e_{o_2,d_2,t}\ )$$

this, and (B.3), yields the approximation

$$Cov_T\ (\ FE_{o_1,d_1,i}, FE_{o_2,d_2,i}\ ) \approx Corr(\ e_{o_1,d_1,t}, e_{o_2,d_2,t}\ )\ (Var_T(FE_{o_1,d_1,i})\ Var_T(FE_{o_2,d_2,i}))^{1/2}.$$

Thus, with (B.2),

$$\mathrm{Var}_T\ \mathrm{NMFE}_{w,i} = \mathrm{Var}_T\ (\Sigma_{o\neq w}\ (P_o/P_w)\ F_{o,w,i}) \tag{B.8}$$
$$+\ \mathrm{Var}_T\ (\Sigma_{d\neq w}\ F_{w,d,i}) - 2\ \mathrm{Cov}_T(\Sigma_{o\neq w}\ (P_o/P_w)\ F_{o,w,i},\ \Sigma_{d\neq w}\ F_{w,d,i})$$

$$= \Sigma_{o_1\neq w}\ \Sigma_{o_2\neq w}\ (P_{o_1}\ P_{o_2}\ /P_w^2)\ \mathrm{Corr}(\ e_{o_1,w,t},\ e_{o_2,w,t})\ (\mathrm{Var}_T(\mathrm{FE}_{o_1,w,i})\ \mathrm{Var}_T(\mathrm{FE}_{o_2,w,i}))^{1/2}$$

$$+\ \Sigma_{d_1\neq w}\ \Sigma_{d_2\neq w}\ \mathrm{Corr}(\ e_{w,d_1,t},\ e_{w,d_2,t})\ (\mathrm{Var}_T(\mathrm{FE}_{w,d_1,i})\ \mathrm{Var}_T(\mathrm{FE}_{w,d_2,i}))^{1/2}$$

$$-\ 2\ \Sigma_{o\neq w}\ \Sigma_{d\neq w}\ (P_o\ /\ P_w)\ \mathrm{Corr}(\ e_{o,w,t},\ e_{w,d,t})\ (\mathrm{Var}_T(\mathrm{FE}_{o,w,i})\ \mathrm{Var}_T(\mathrm{FE}_{w,d,i}))^{1/2}.$$

In the case of Gaussian errors, $G(t) = \exp(t^2/2)$. With $h(\beta,i) = \Sigma_{j=1}^{i}\ (1-\beta^j)^2\ /\ (1-\beta)^2$, from (B.7), we have

$$\mathrm{Var}_T\ \mathrm{FE}_{o,d,i} = (F_{o,d,i})^2\ \{\exp(\ 2\sigma_{o,d}^2\ h(\beta,i))\ -\ \exp(\ \sigma_{o,d}^2\ h(\beta,i))\ \}. \tag{B.9}$$

The approximate 95% confidence bands in Figure 9 were computed using $\mathrm{NMF}_{w,i} \pm 2\ (\mathrm{Var}_T\ \mathrm{NMFE}_{w,i})^{1/2}$, after using (B.9) in (B.8). The parameters $\beta$, $\{\sigma_{o,d}\}$ and $\{\mathrm{Corr}(\ e_{o_1,d_1,t},\ e_{o_2,d_2,t})\}$ were replaced by corresponding estimates. The rationale behind this symmetric interval is that the basic statistic, $\mathrm{NMFE}_{w,i}$, is the sum over 100 weakly dependent random variables. Hence, a central limit theorem argument can be used to justfy that the distribution of this sum can be approximated by a normal distribution. Note that this argument is true regardless of the form of the moment generated function, G, that is used. Certainly, it would be possible to approximate the distribution of this sum by alternative methods. A more pessimistic view would be to interpret the plus or minus 2 standard error bounds as a 75% confidence band using a Chebyshev type argument. An alternative point forecast I investigated was $\mathrm{NMF}_{w,i} - E_T\ \mathrm{NMFE}_{w,i}$, a conditionally unbiased estimator of $(\ \Sigma_{o\neq w}\ P_o\ R_{o,w,T+i}\ -\ P_w\ \Sigma_{d\neq w}\ R_{w,d,T+i}\ )\ /\ P_w$. For this application, the correction term $E_T\ \mathrm{NMFE}_{w,i}$ was negligible compared to the point estimate $\mathrm{NMF}_{w,i}$ and its corresponding standard error.

# APPENDIX C

## Exploratory Graphical Analysis

## on

## Forecasting State-to-State Migration Rates

Figures 1. Plots of Summary Migration Rates by Year

Figures 2. State Specific Plots of Out-Migration Rates by Year

Figures 3. Plots of 1985 versus 1984 Out-Migration Rates

Figures 4. Plots of Times Series Standard Deviation versus Explanatory Variables

Figures 5. Standardized State Specific Plots of Out-Migration Rates by Year

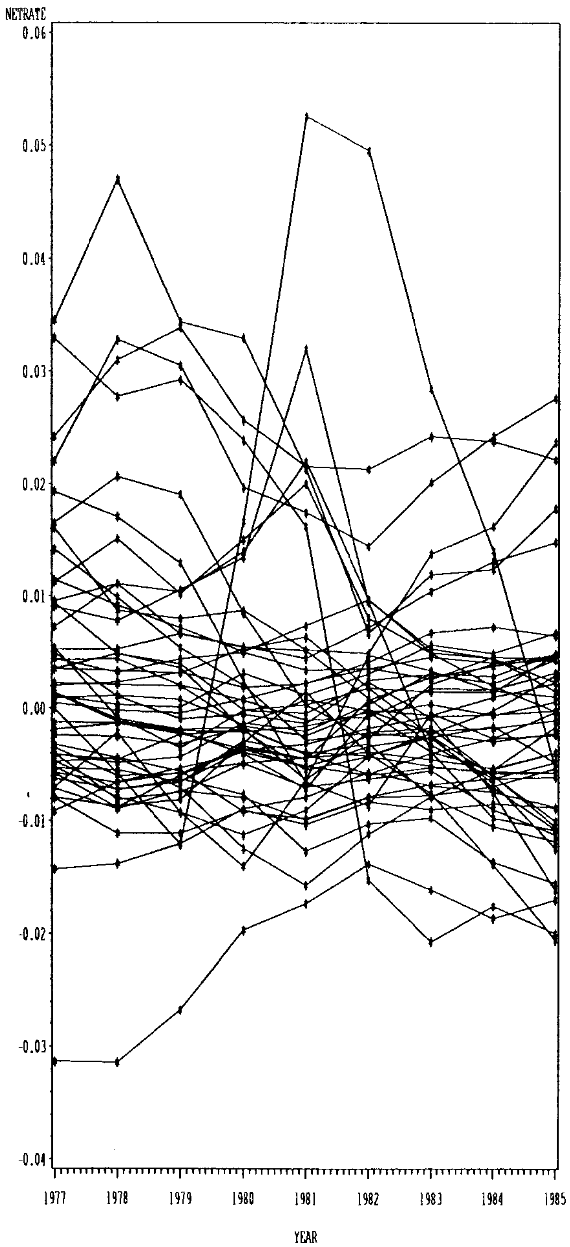
Figures 6. Autocorrelation of Standardized Rates

Figures 7. Plots of Net, In and Out Actual and Forecasts for Model 7b

Figures 8. Plots of Net, In and Out Actual and Forecasts for Model 8b

## Plot of Net Migration Rates by State

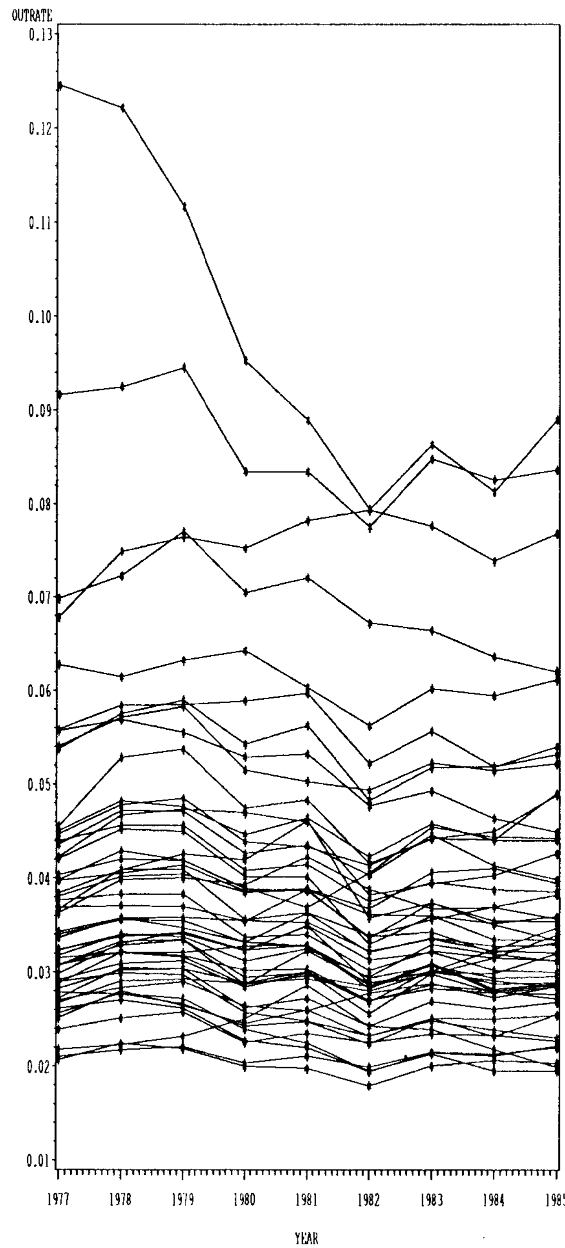
Time is 1977 - 1985, inclusive




Net-Rate is defined as Annual Population Migration Change
Divided by Initial State Population

## Plot of Out-Migration Rates by State

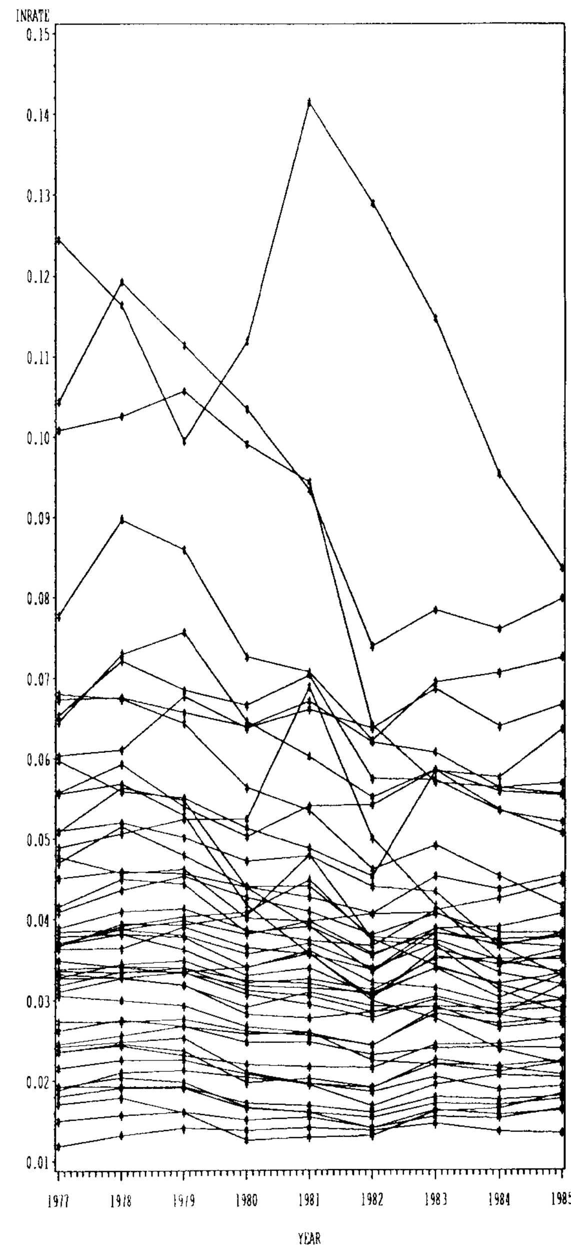
Time is 1977 - 1985, inclusive




Out-Rate is defined as Annual Population Outflow
Divided by Initial State Population

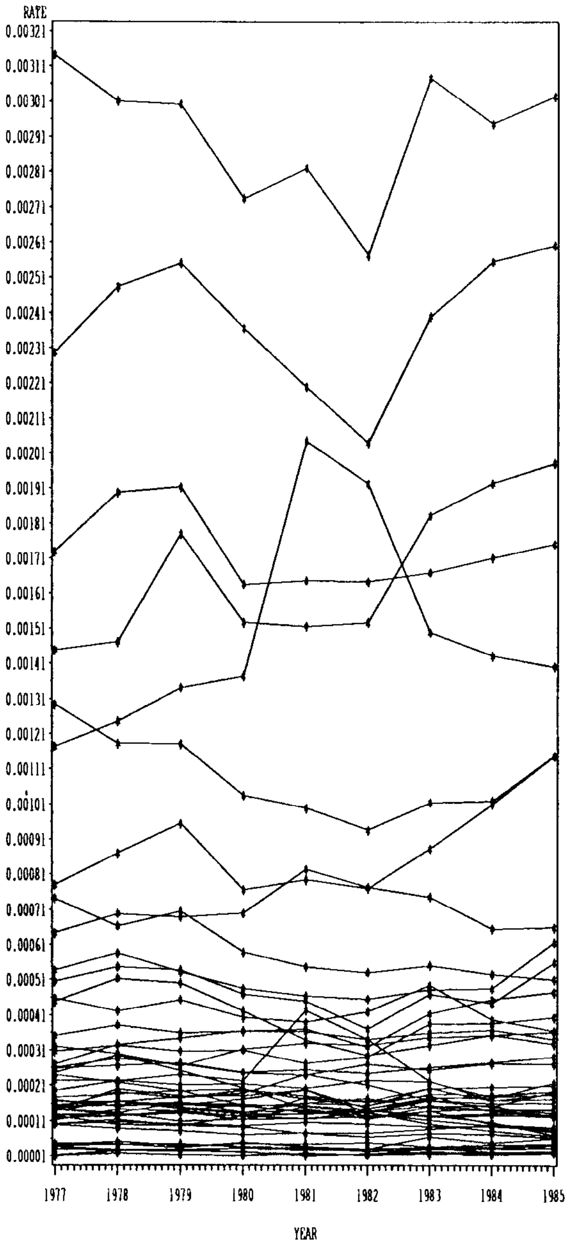
## Plot of In-Migration Rates by State

Time is 1977 - 1985, inclusive



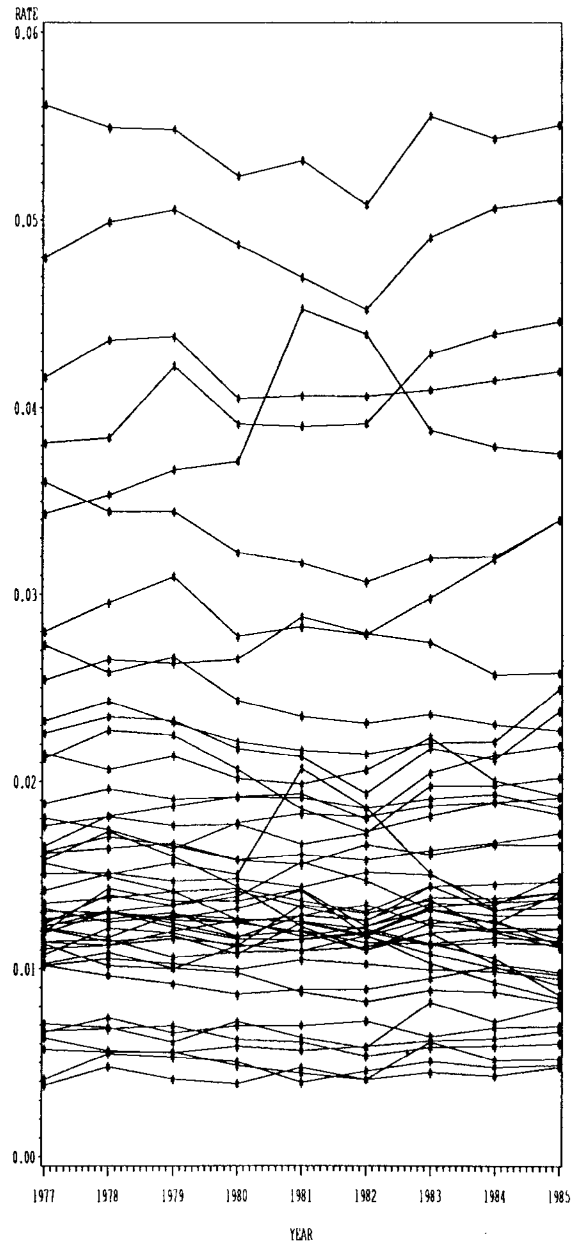

In-Rate is defined as Annual Population Inflow
Divided by Initial State Population

## Plot of Wisconsin Out-Migration Rates
Time is 1977 - 1985, inclusive



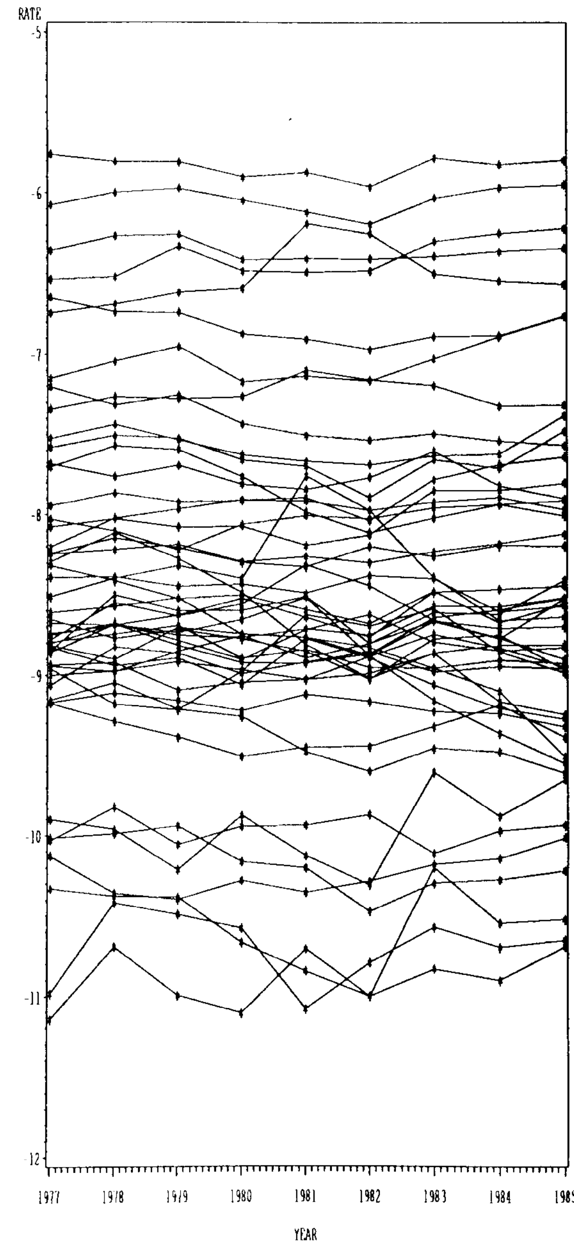

By State of Destination

## Plot of Wisconsin Out-Migration Rates
Time is 1977 - 1985, inclusive

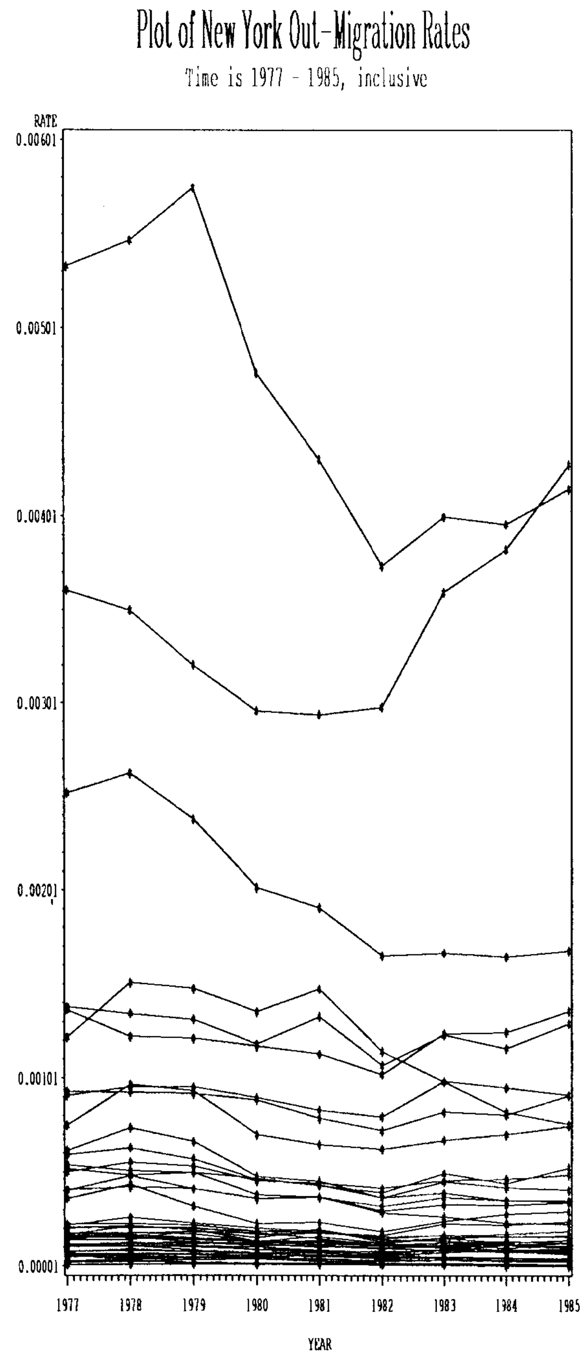



By State of Destination
Data is in Square Roots

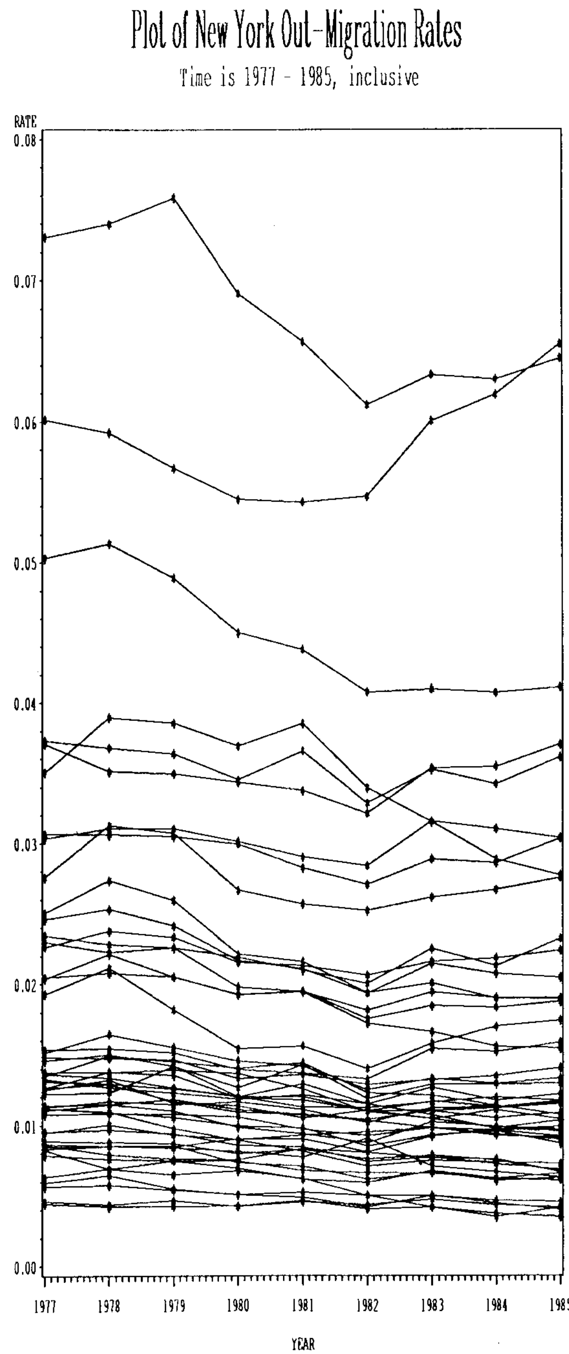
## Plot of Wisconsin Out-Migration Rates
Time is 1977 - 1985, inclusive



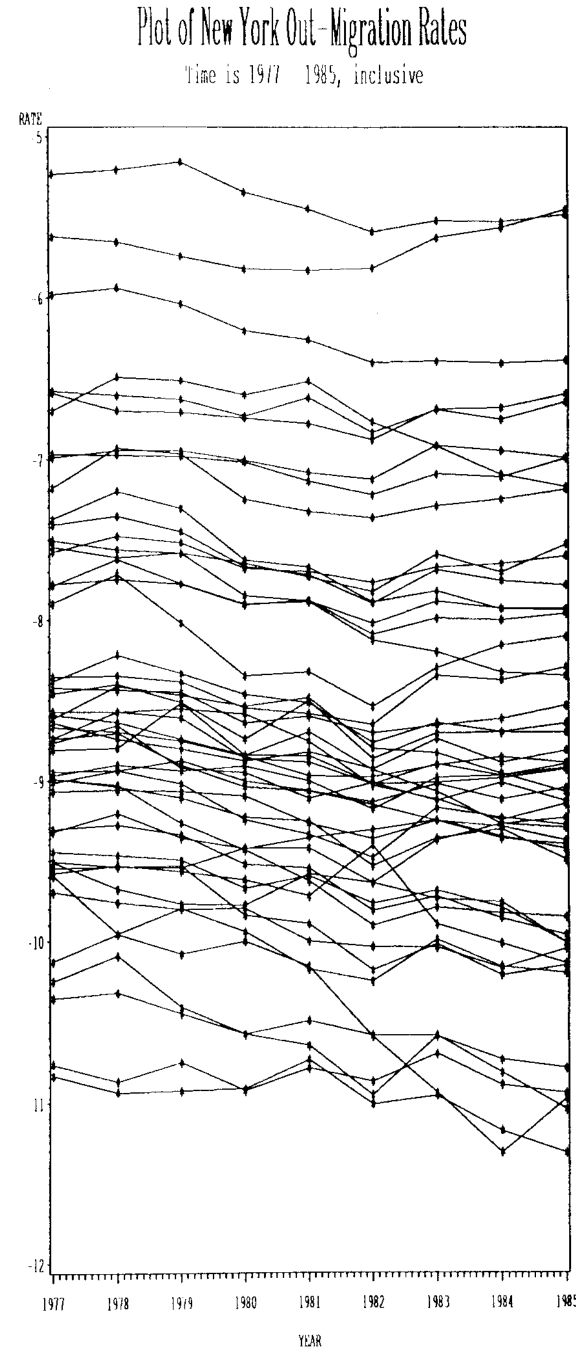

By State of Destination
Data is in Natural Logarithms




By State of Destination



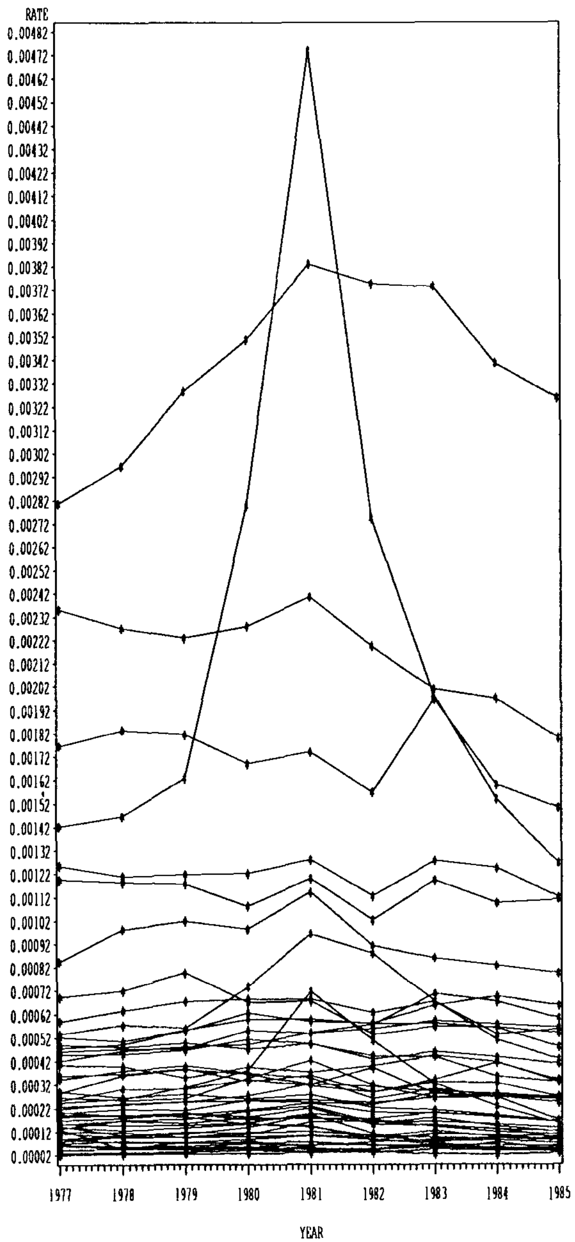

By State of Destination
Data is in Square Roots




By State of Destination
Data is in Natural Logarithms

Plot of Michigan Out-Migration Rates

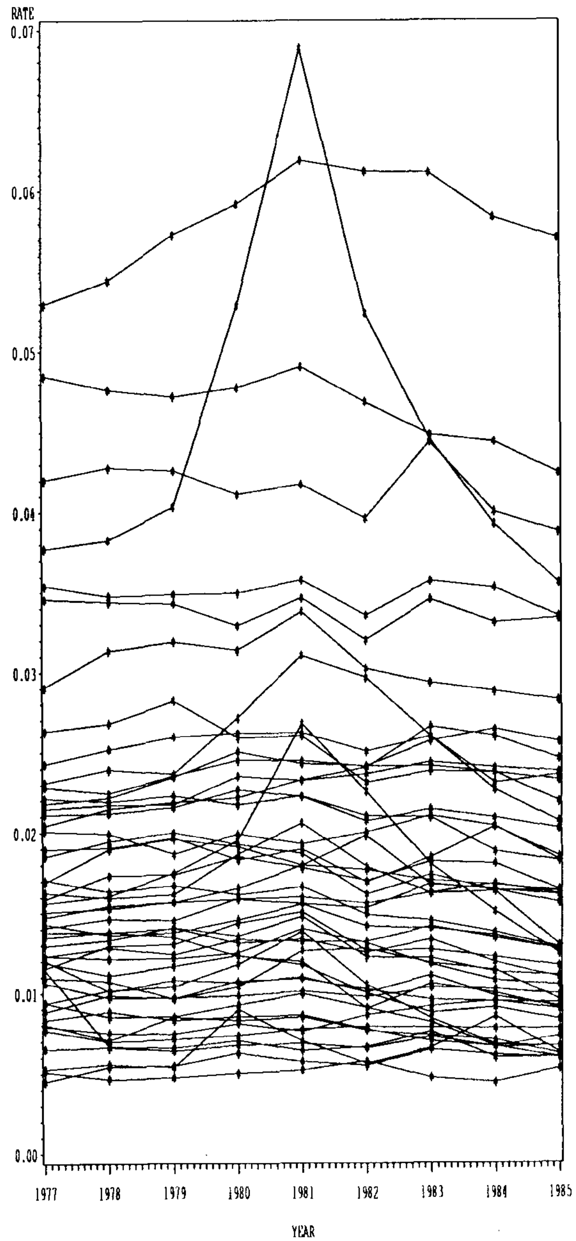
Time is 1977 - 1985, inclusive




By State of Destination

Plot of Michigan Out-Migration Rates

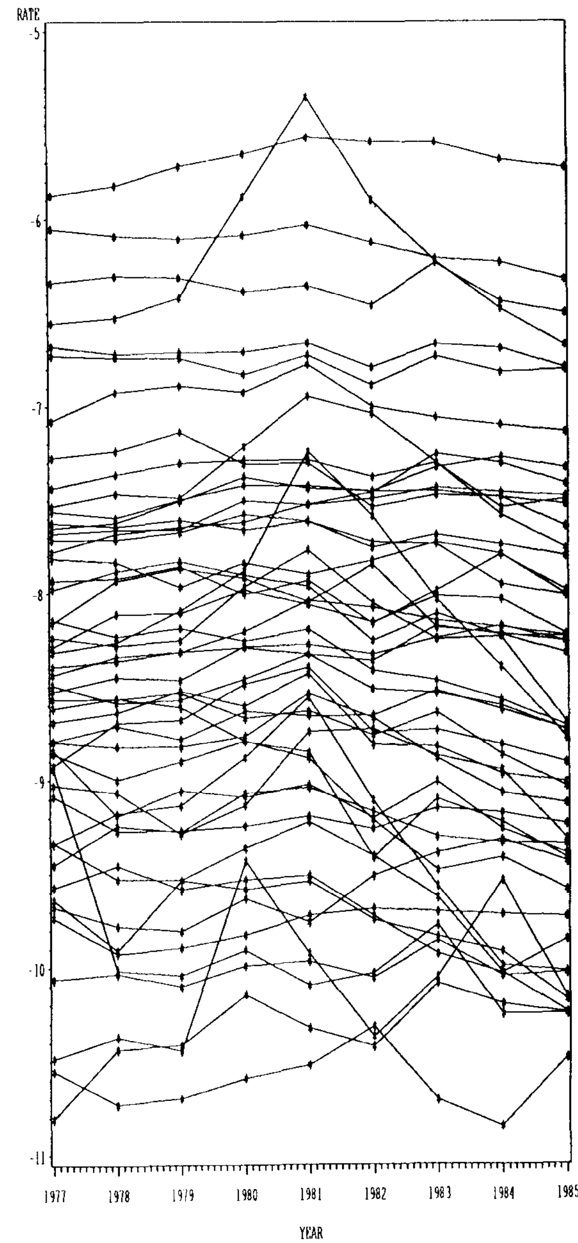
Time is 1977 - 1985, inclusive




By State of Destination
Data is in Square Roots

Plot of Michigan Out-Migration Rates

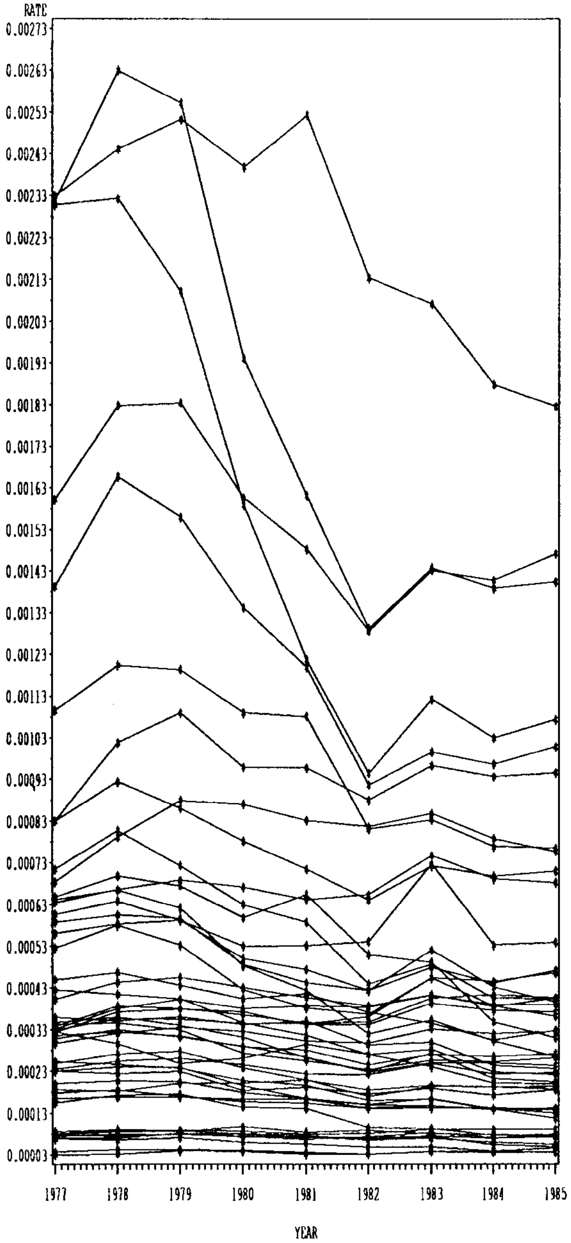
Time is 1977 - 1985, inclusive




By State of Destination
Data is in Natural Logarithms

## Plot of California Out-Migration Rates

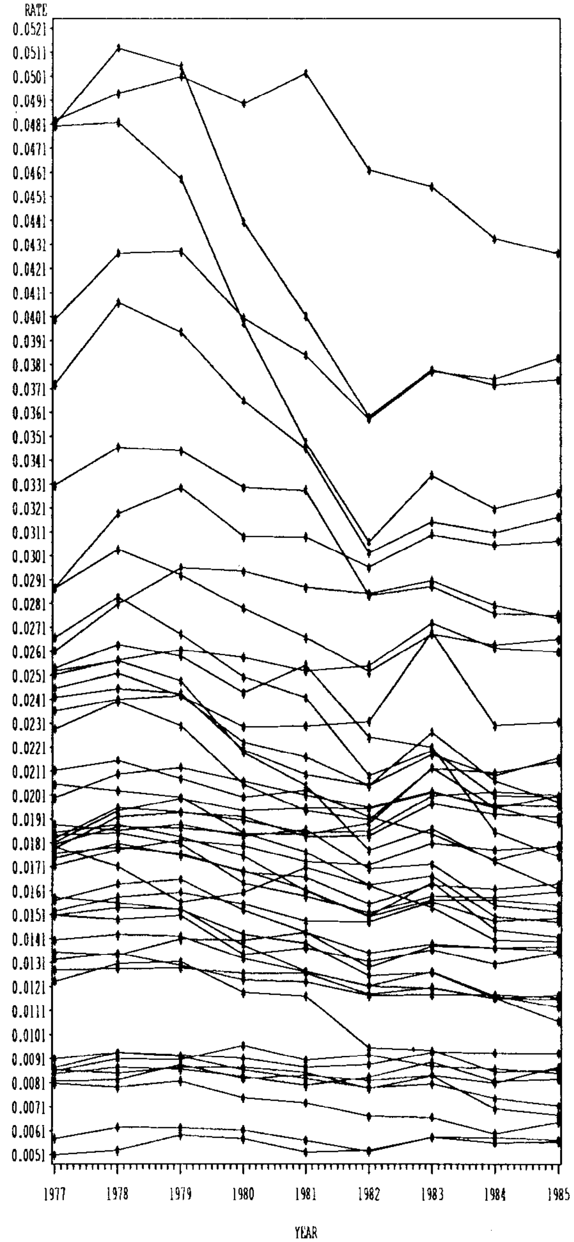
Time is 1977 - 1985, inclusive




By State of Destination

## Plot of California Out-Migration Rates

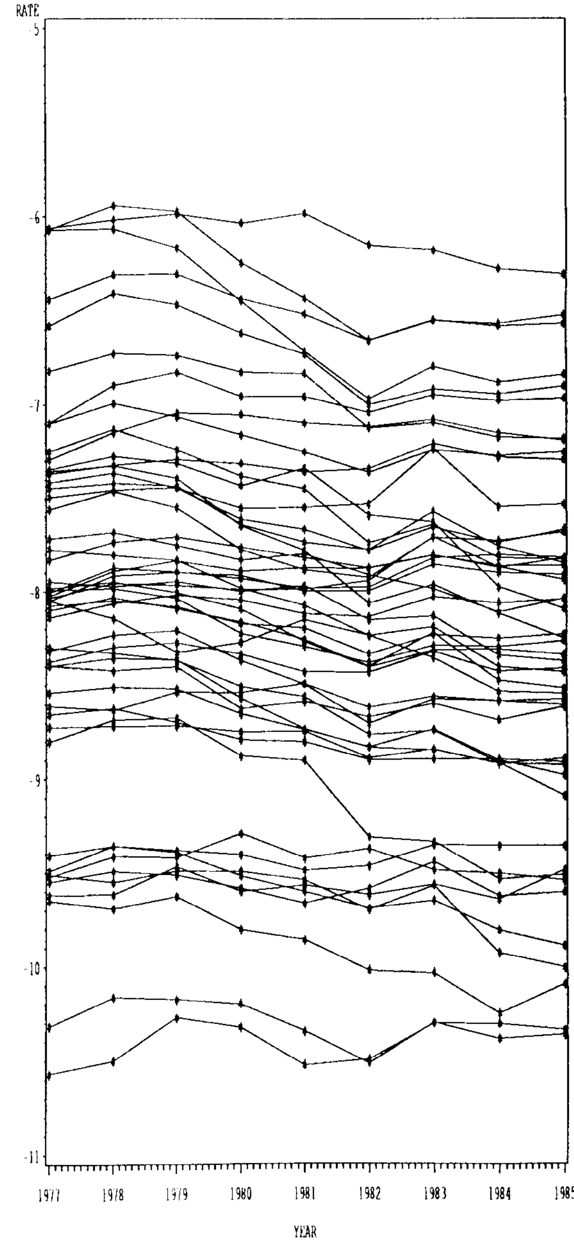
Time is 1977 - 1985, inclusive



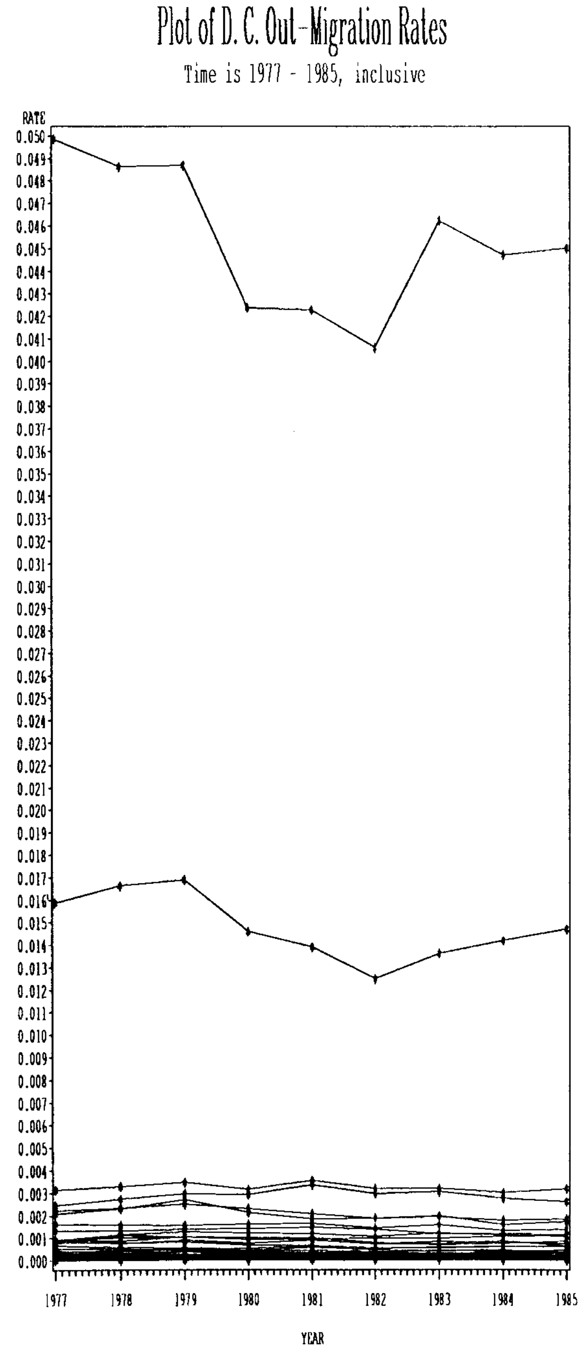

By State of Destination
Data is in Square Roots

## Plot of California Out-Migration Rates

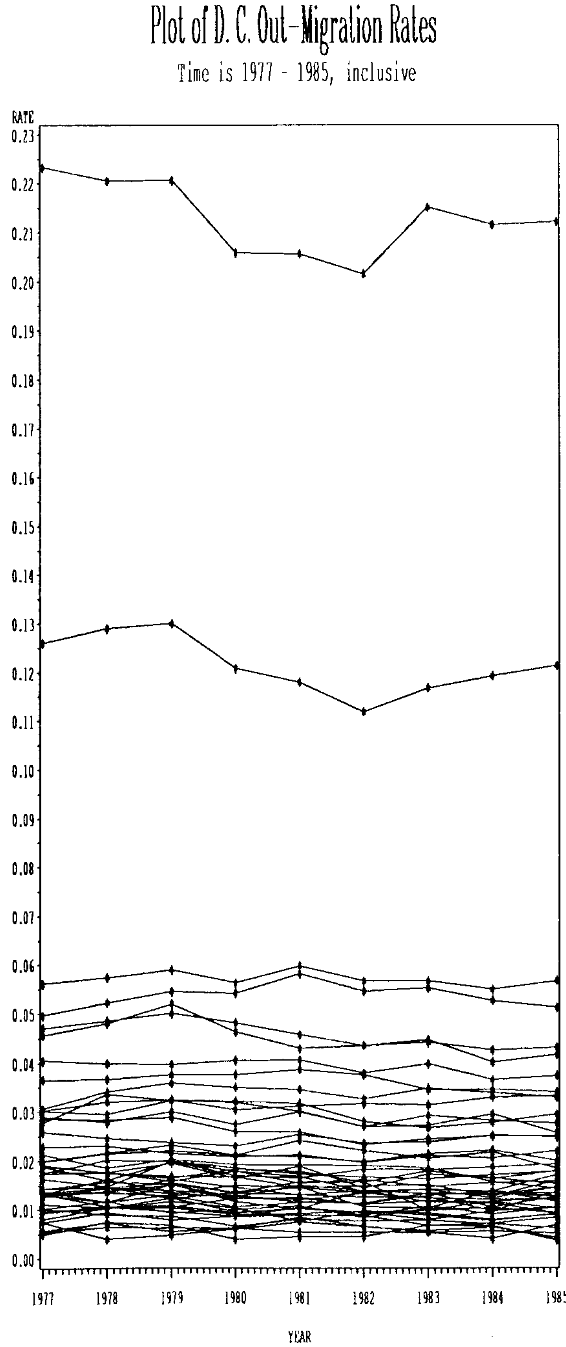
Time is 1977 - 1985, inclusive



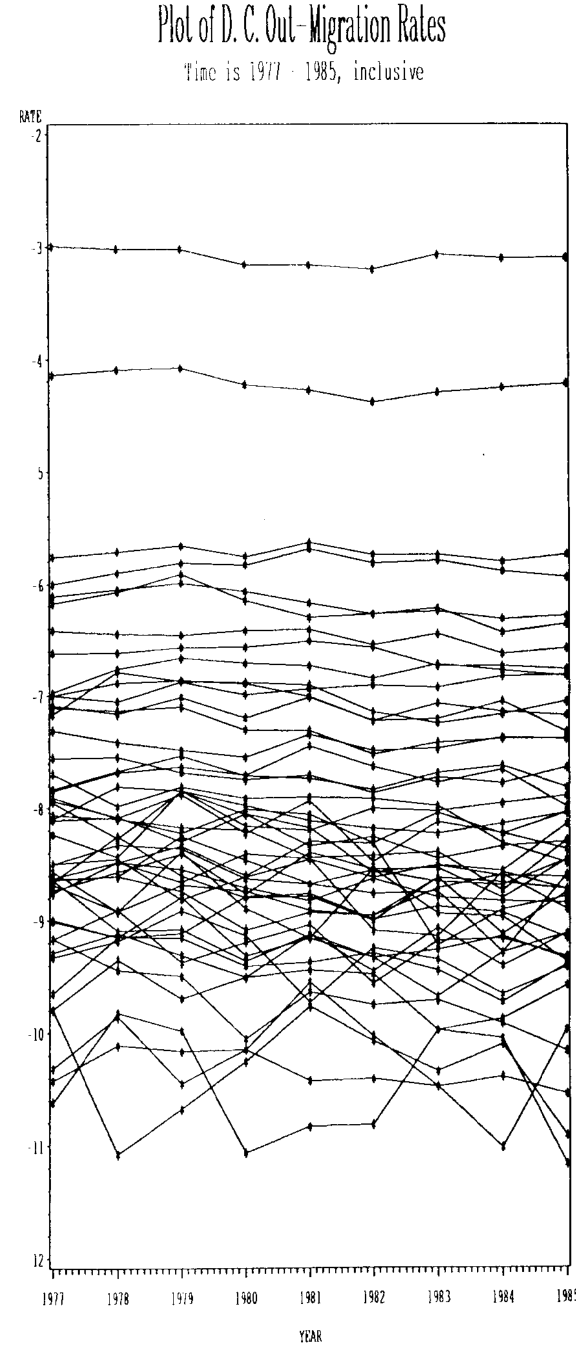

By State of Destination
Data is in Natural logarithms




By State of Destination

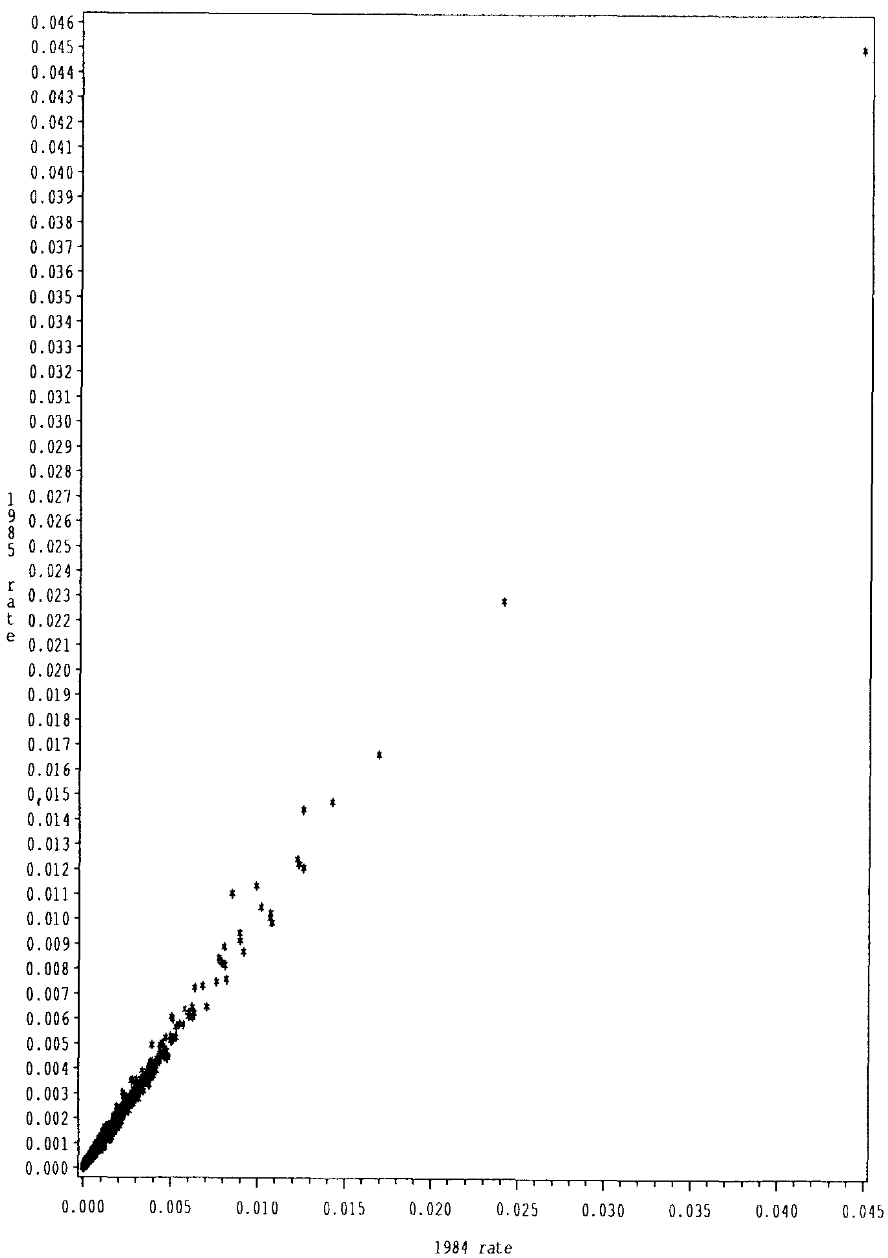



By State of Destination
Data is in Square Roots



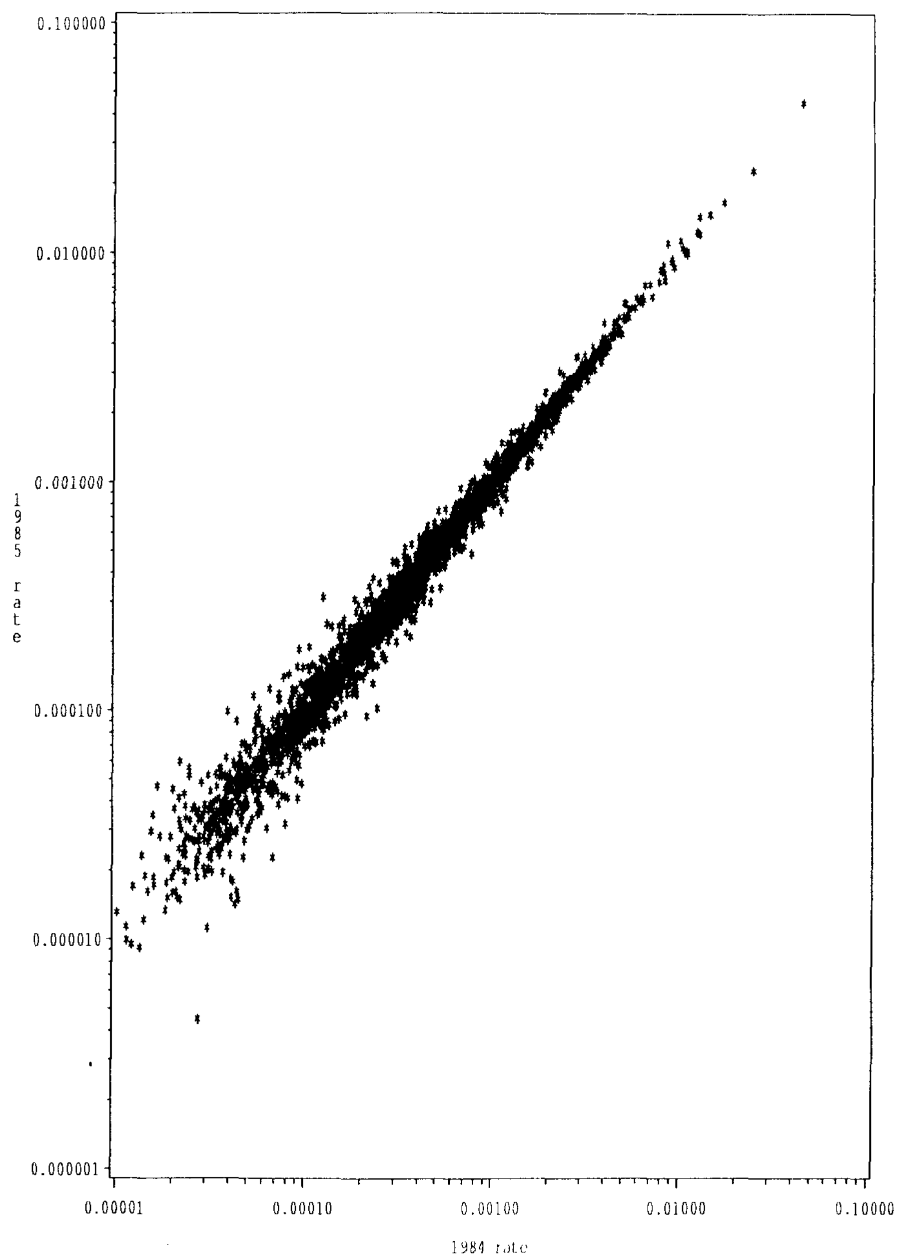

By State of Destination
Data is in Natural Logarithms

1985 vs 1984 Out-Migration Rates




Each point is a Destination-specific Out-Migration Rate

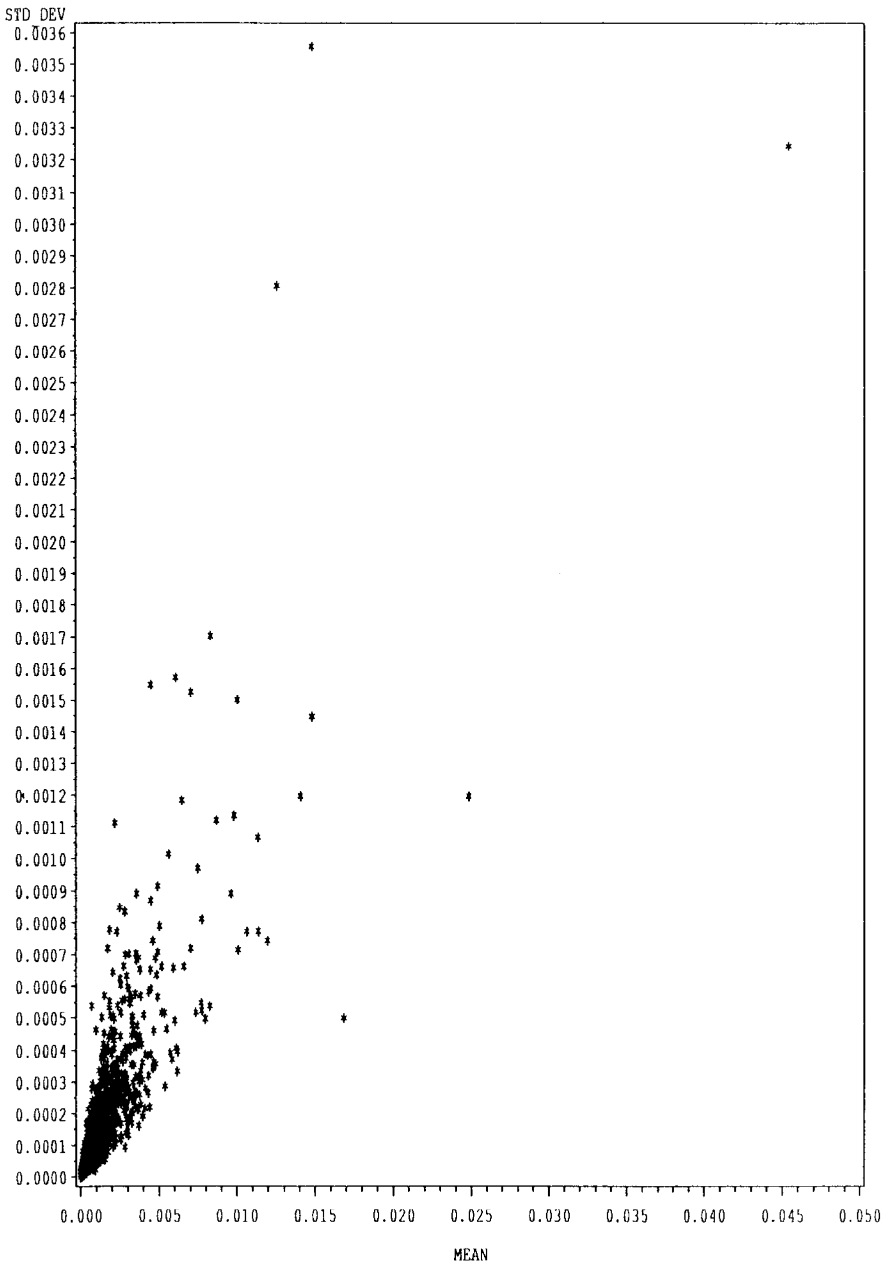
1985 vs 1984 Out-Migration Rates




Each point is a Destination-specific Out-Migration Rate
Data is plotted in Logarithmic Scale

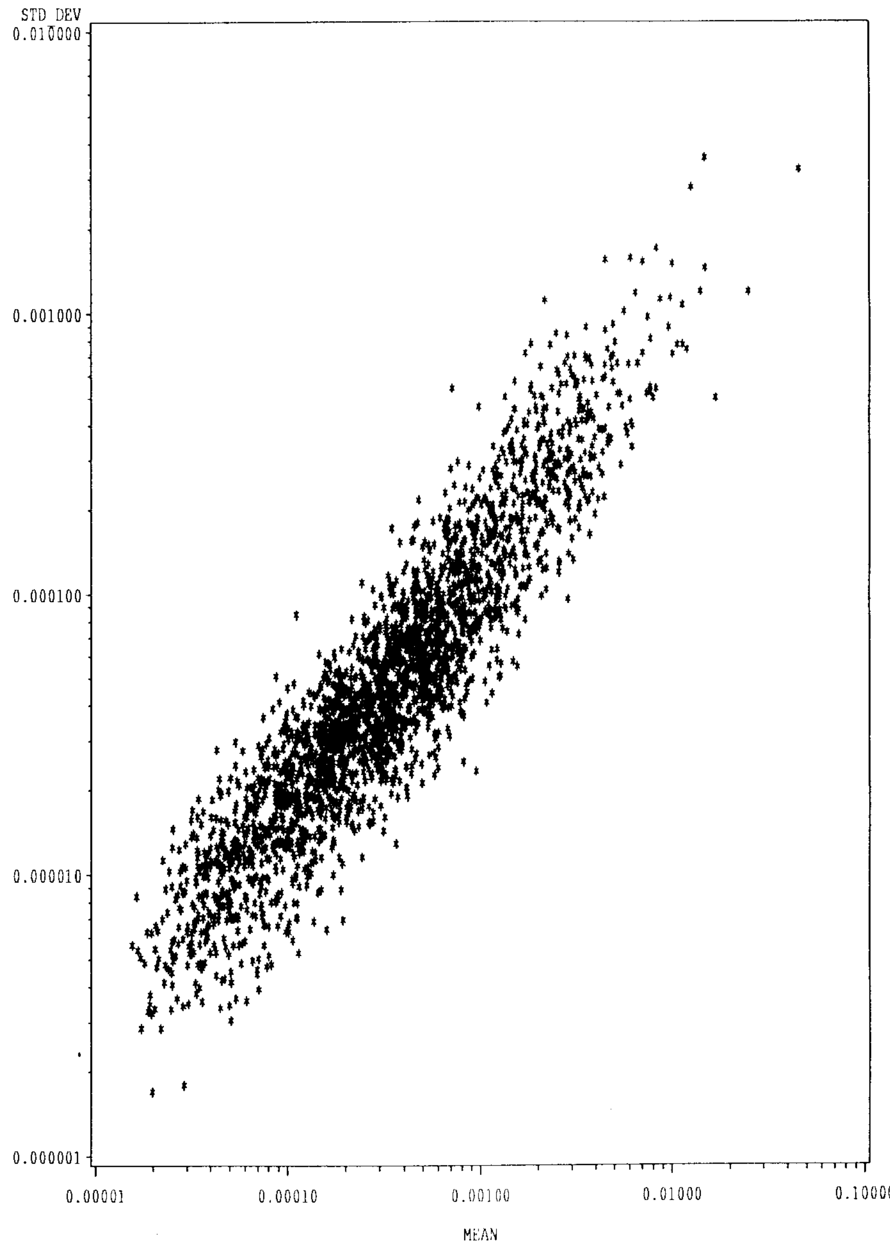
# Time Series Standard Deviation vs Mean

Time is 1977 - 1985, inclusive




# Time Series Standard Deviation vs Mean

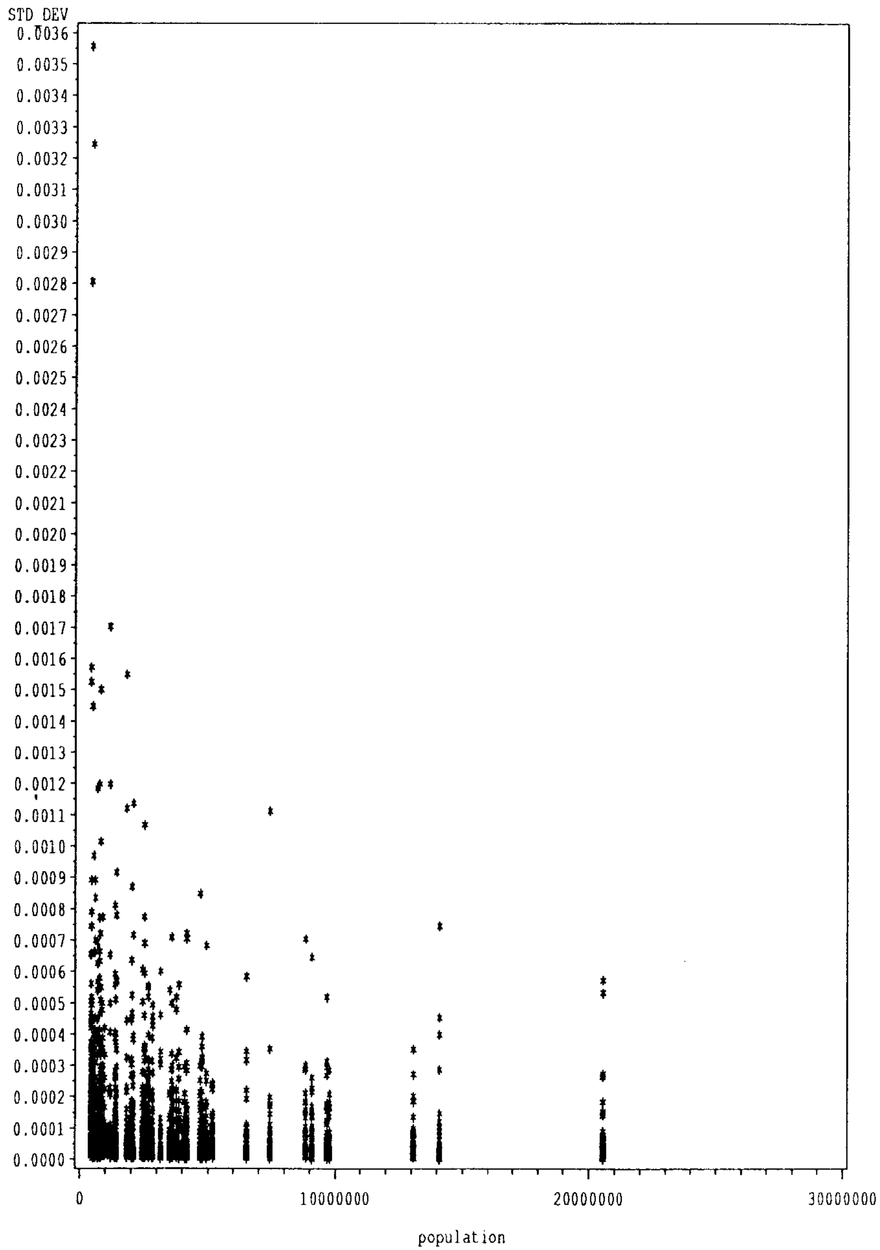
Time is 1977 - 1985, inclusive




Data is plotted in Logarithmic Scale

## Time Series Standard Deviation vs Population

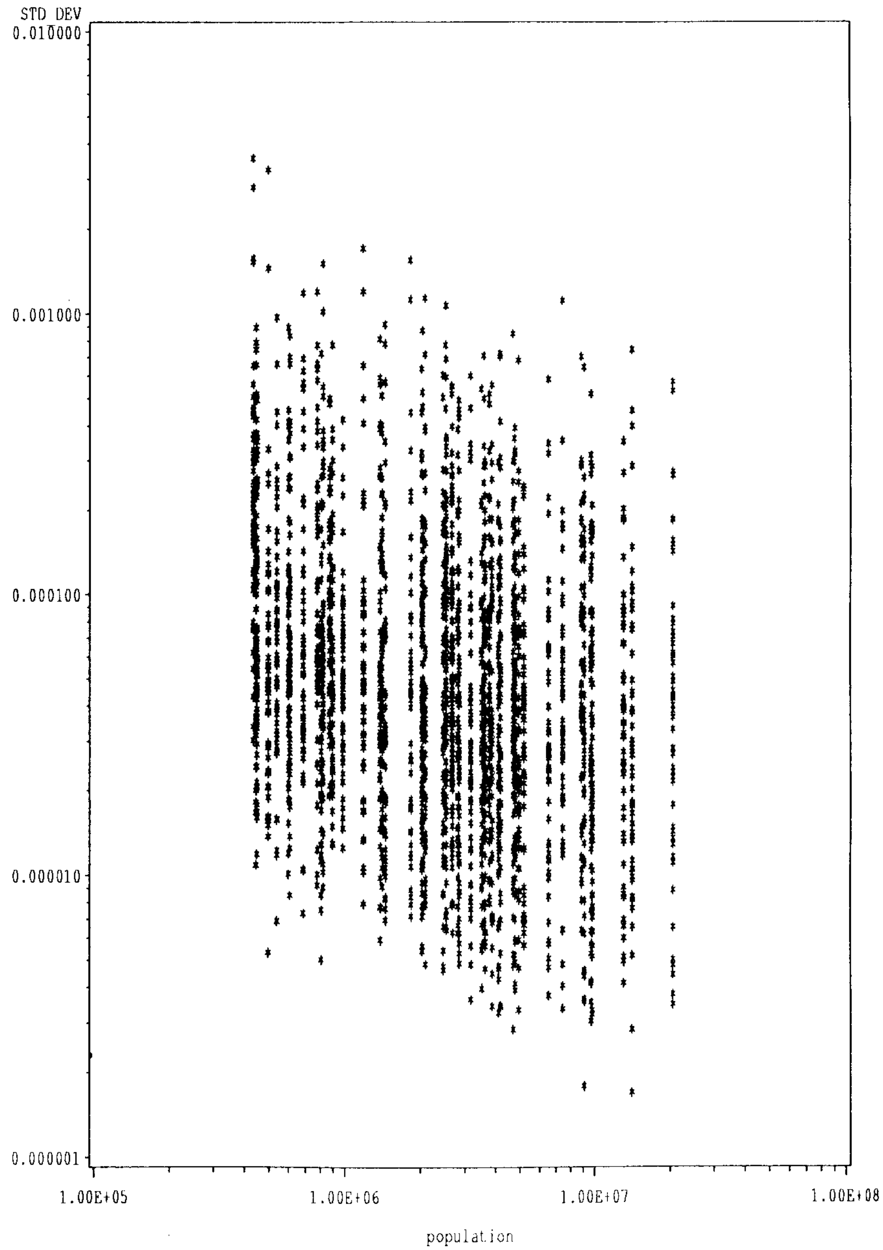
Time is 1977 - 1985, inclusive




Population is Origin-Specific

## Time Series Standard Deviation vs Population

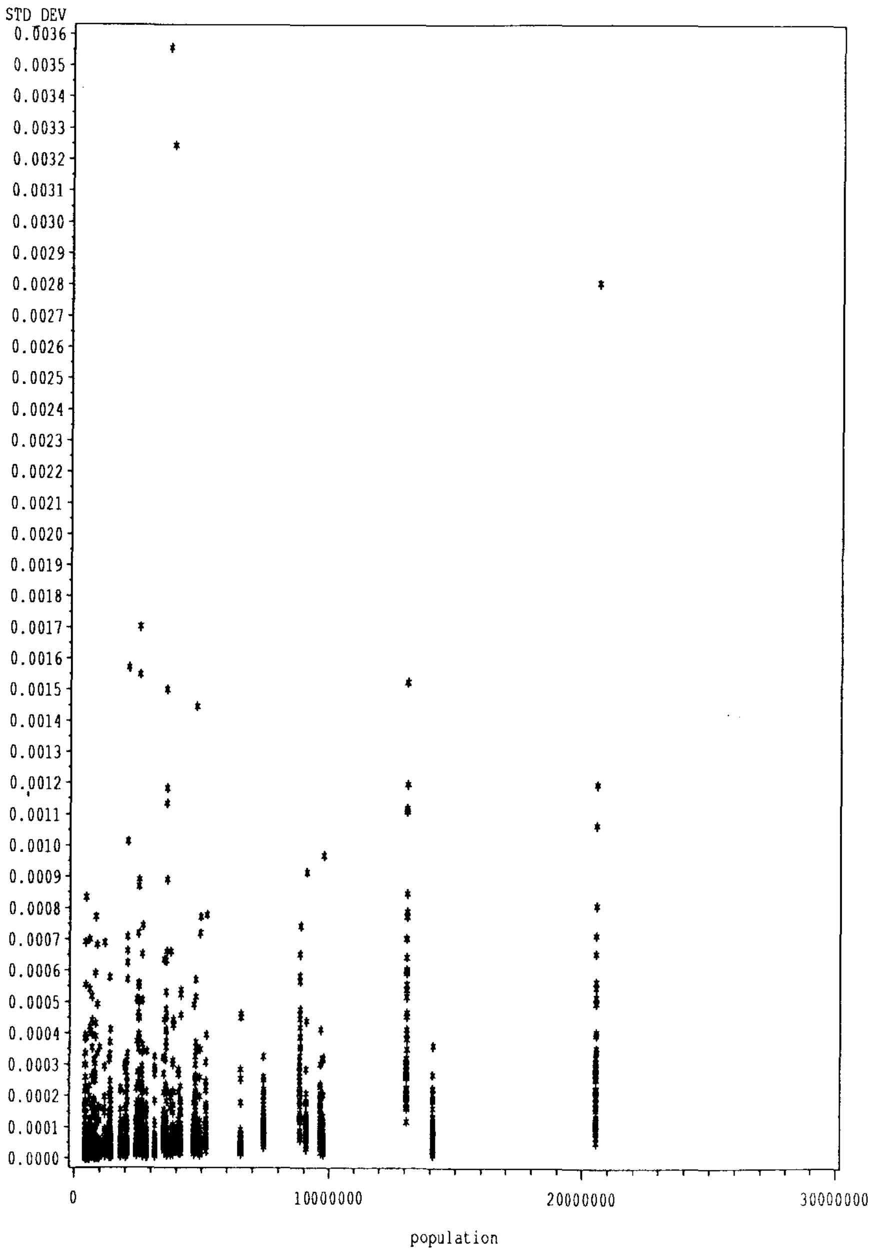
Time is 1977 - 1985, inclusive




Population is Origin-Specific
Data is plotted in Logarithmic Scale

# Time Series Standard Deviation vs Population

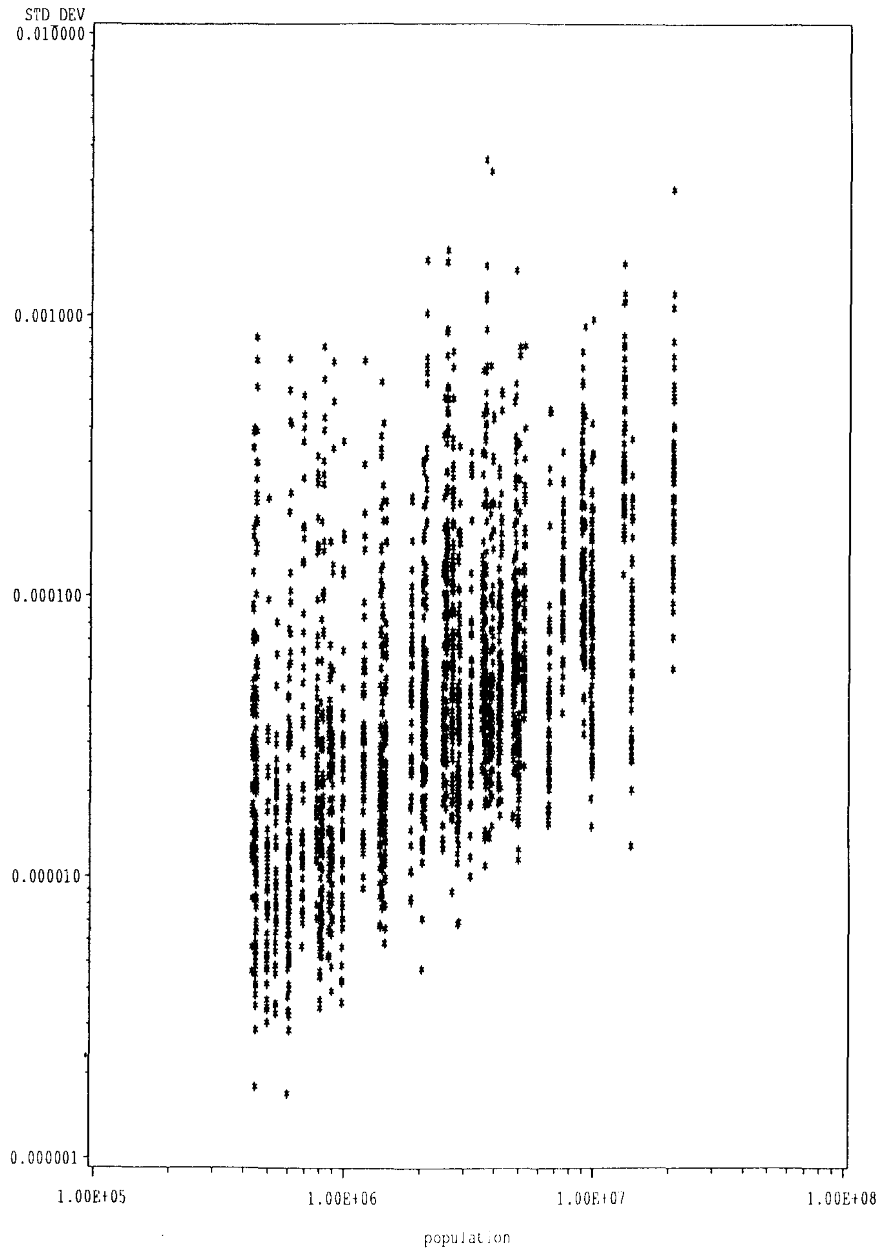
Time is 1977 - 1985, inclusive




Population is Destination-Specific

# Time Series Standard Deviation vs Population

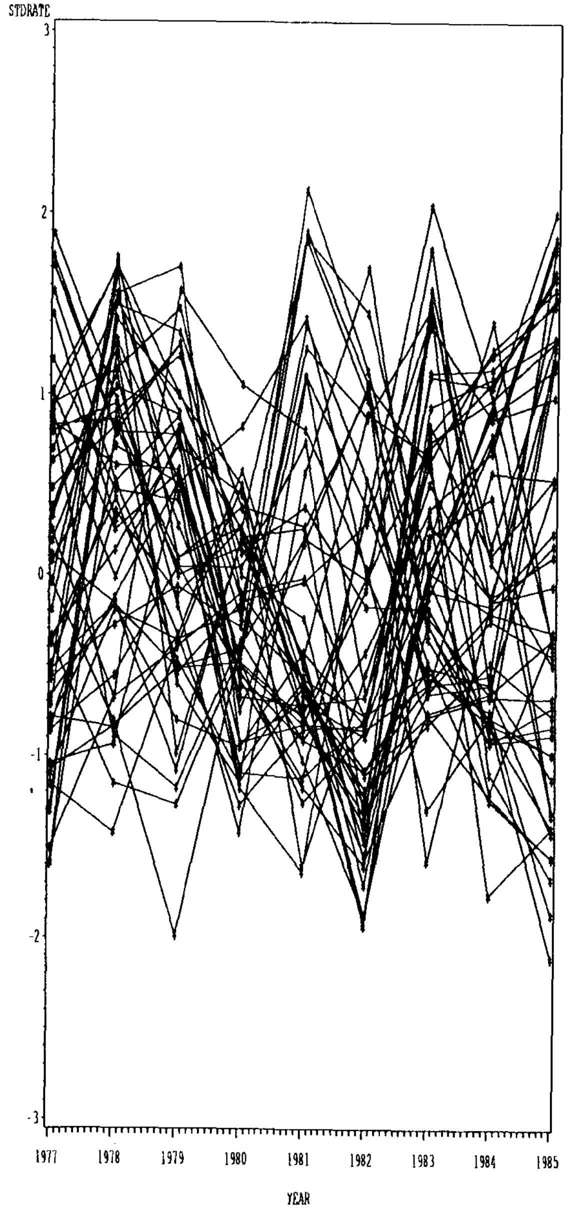
Time is 1977 - 1985, inclusive




Population is Destination-Specific
Data is plotted in logarithmic Scale

Plot of Wisconsin Out-Migration Standardized Rates

Time is 1977 - 1985, inclusive

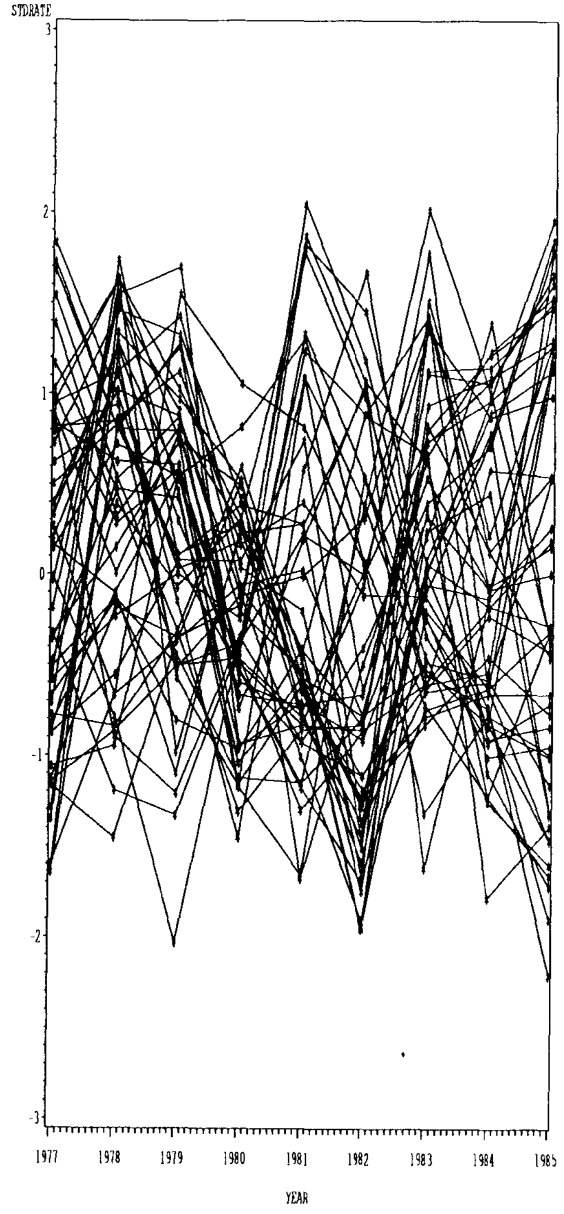



By State of Destination

Standardization by Time Series Mean, SD Filter

Plot of Wisconsin Out-Migration Standardized Rates

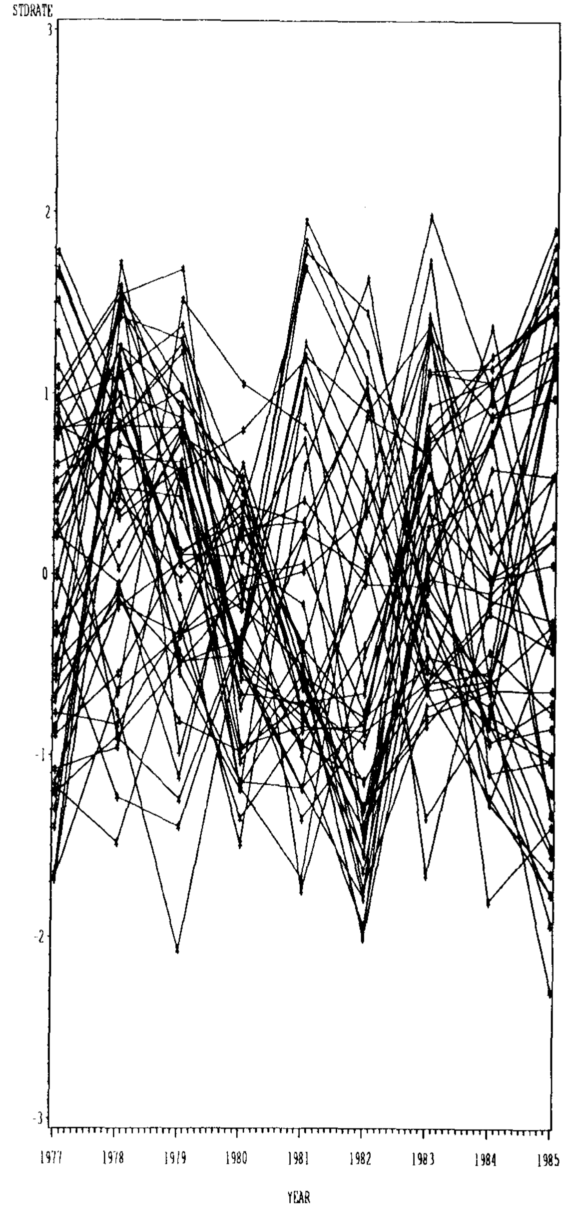
Time is 1977 - 1985, inclusive




By State of Destination
Data is in Square Roots
Standardization by Time Series Mean, SD Filter

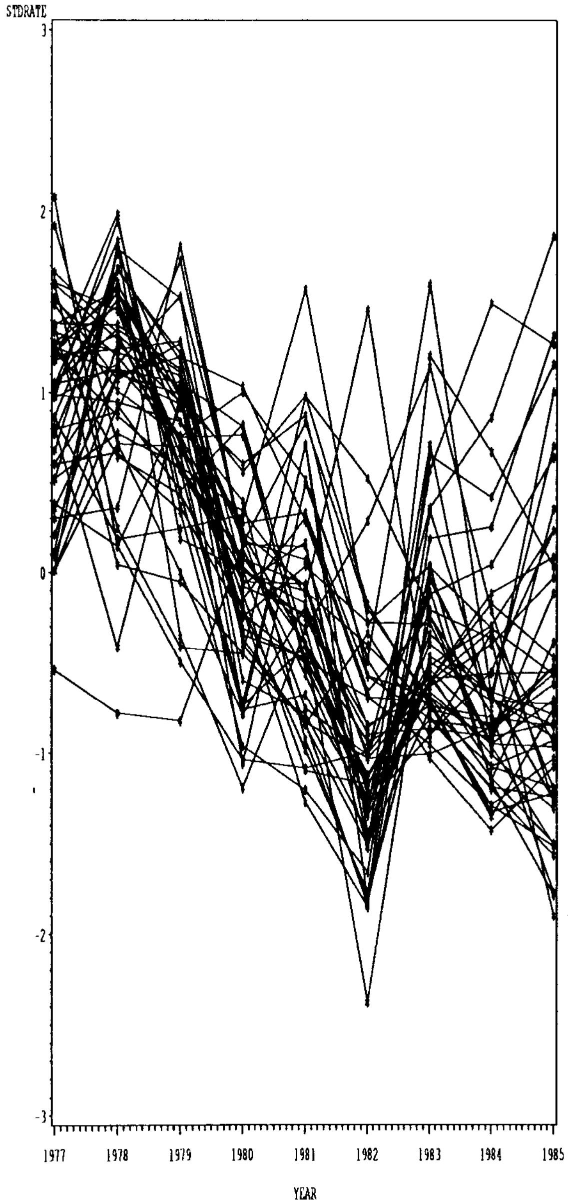
Plot of Wisconsin Out-Migration Standardized Rates

Time is 1977 - 1985, inclusive




By State of Destination
Data is in Natural Logarithms
Standardization by Time Series Mean, SD Filter

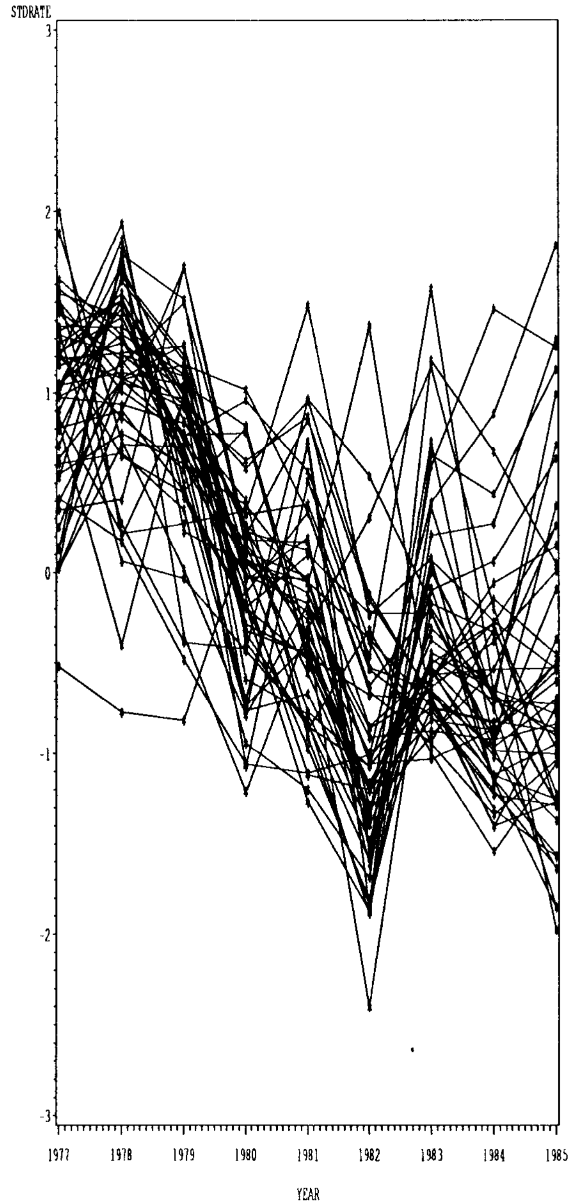
Plot of New York Out-Migration Standardized Rates
Time is 1977 - 1985, inclusive




By State of Destination

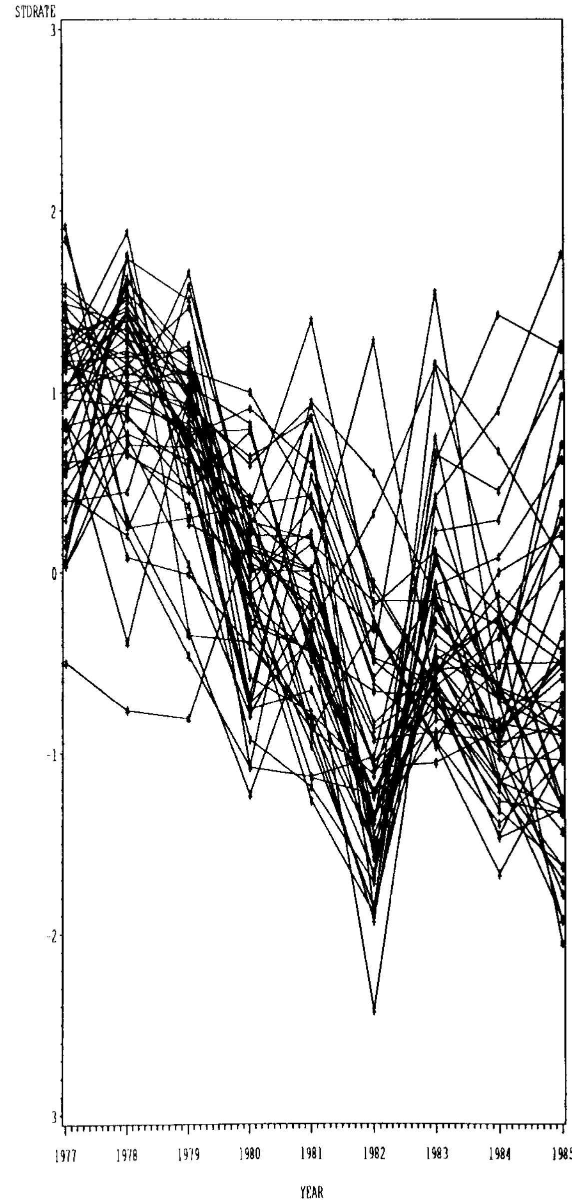
Standardization by Time Series Mean, SD Filter

Plot of New York Out-Migration Standardized Rates
Time is 1977 - 1985, inclusive



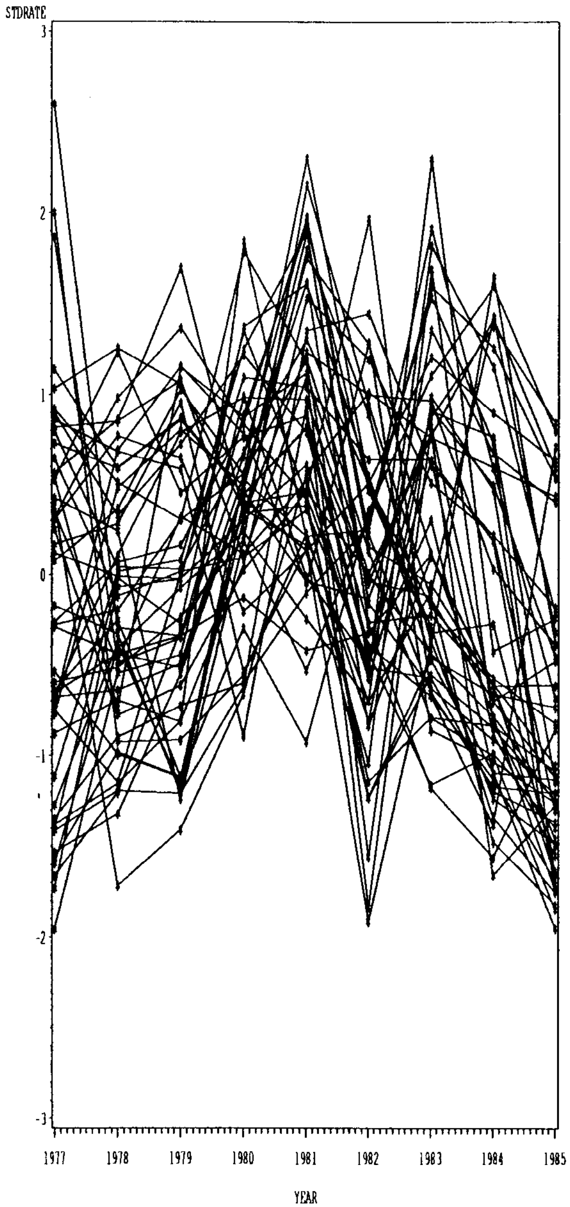

By State of Destination
Data is in Square Roots
Standardization by Time Series Mean, SD Filter

Plot of New York Out-Migration Standardized Rates
Time is 1977 - 1985, inclusive




By State of Destination
Data is in Natural Logarithms
Standardization by Time Series Mean, SD Filter

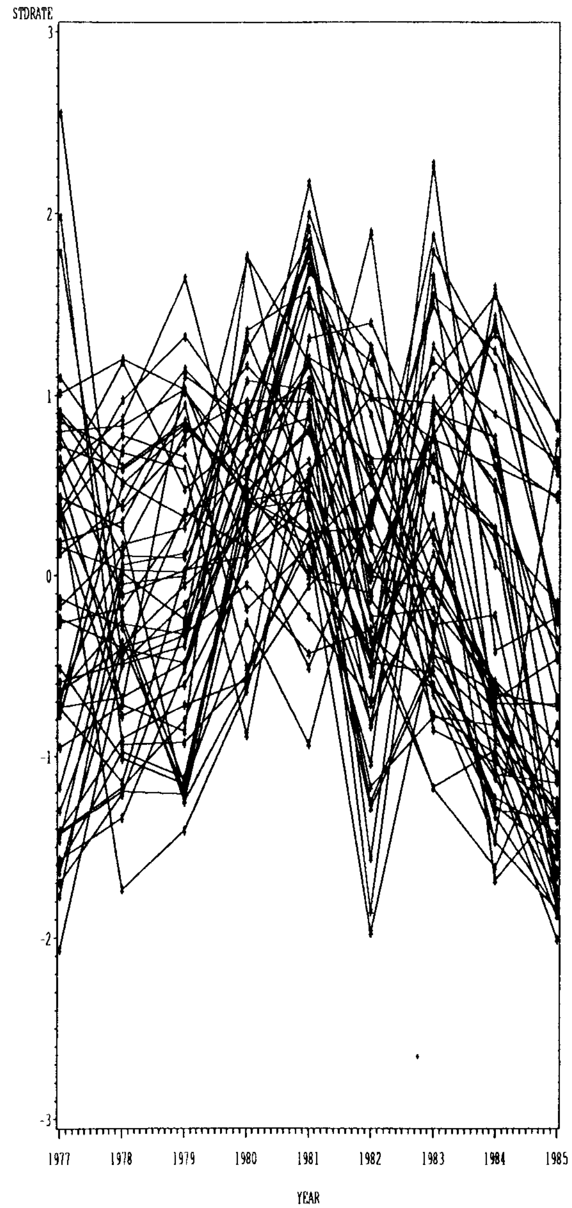
Plot of Michigan Out-Migration Standardized Rates
Time is 1977 - 1985, inclusive




By State of Destination

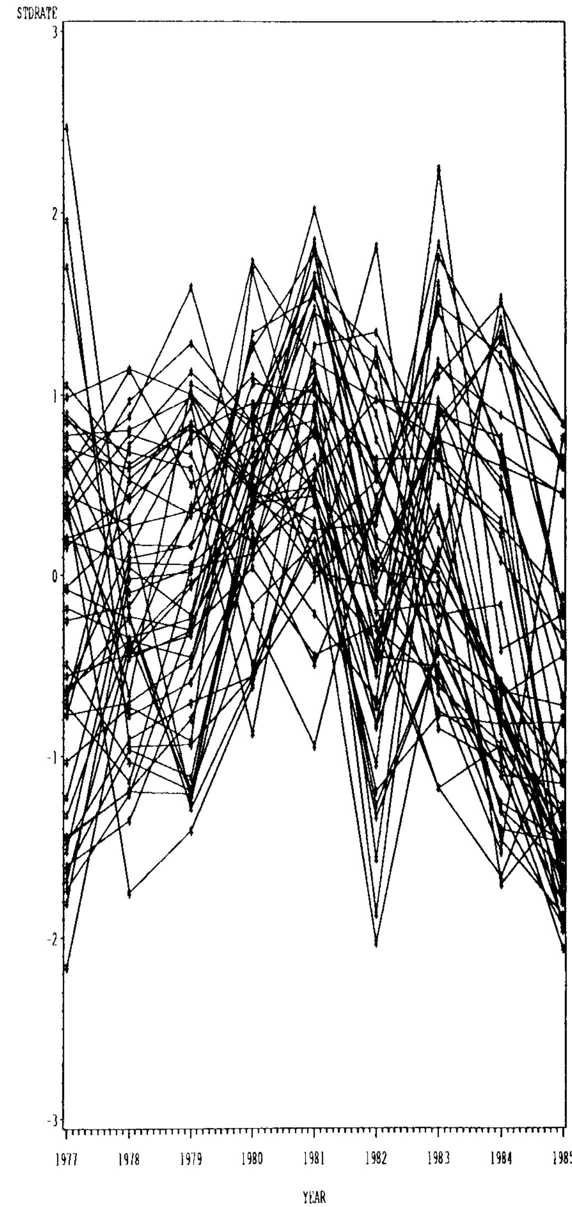
Standardization by Time Series Mean, SD Filter

Plot of Michigan Out-Migration Standardized Rates
Time is 1977 - 1985, inclusive

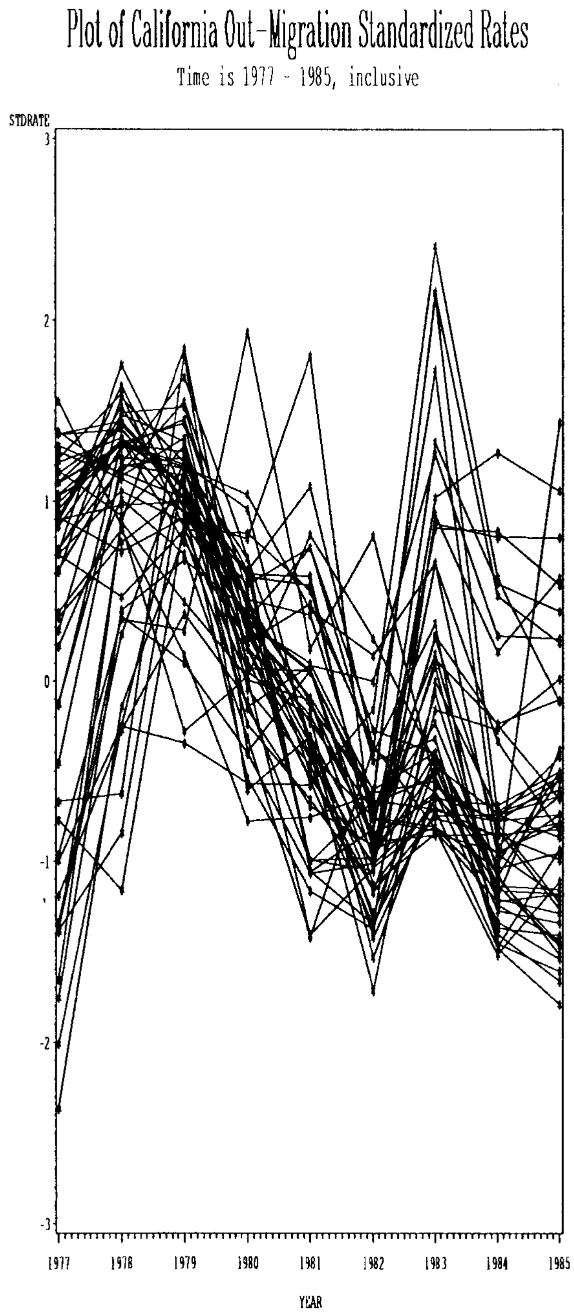



By State of Destination
Data is in Square Roots
Standardization by Time Series Mean, SD Filter

Plot of Michigan Out-Migration Standardized Rates
Time is 1977 - 1985, inclusive

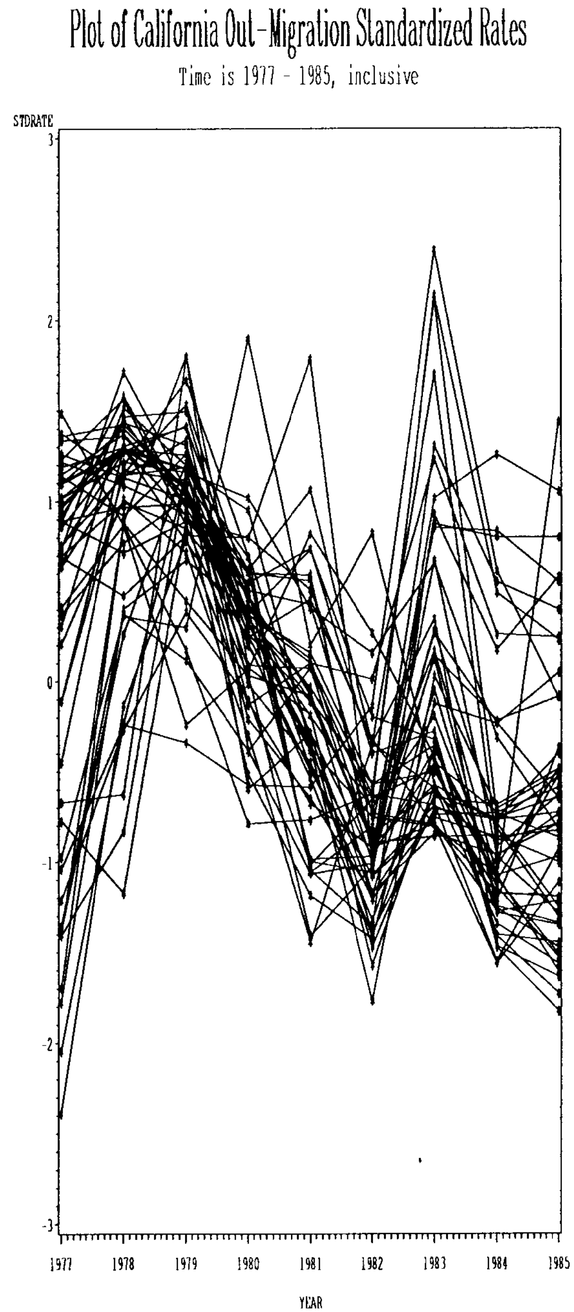



By State of Destination
Data is in Natural Logarithms
Standardization by Time Series Mean, SD Filter

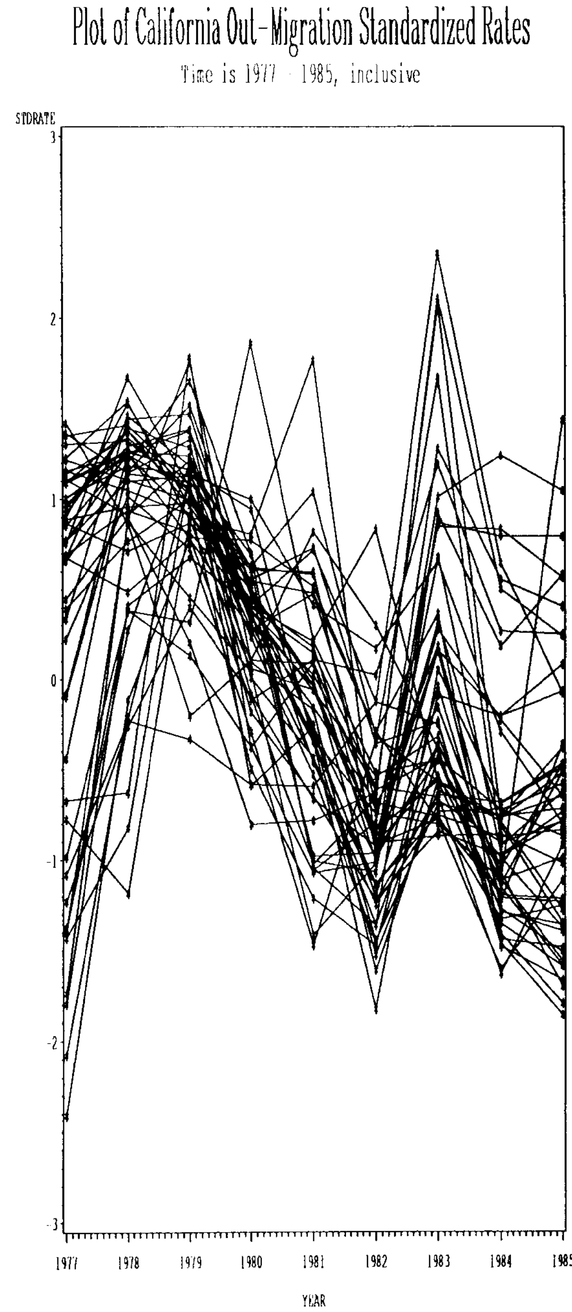



By State of Destination

Standardization by Time Series Mean, SD Filter



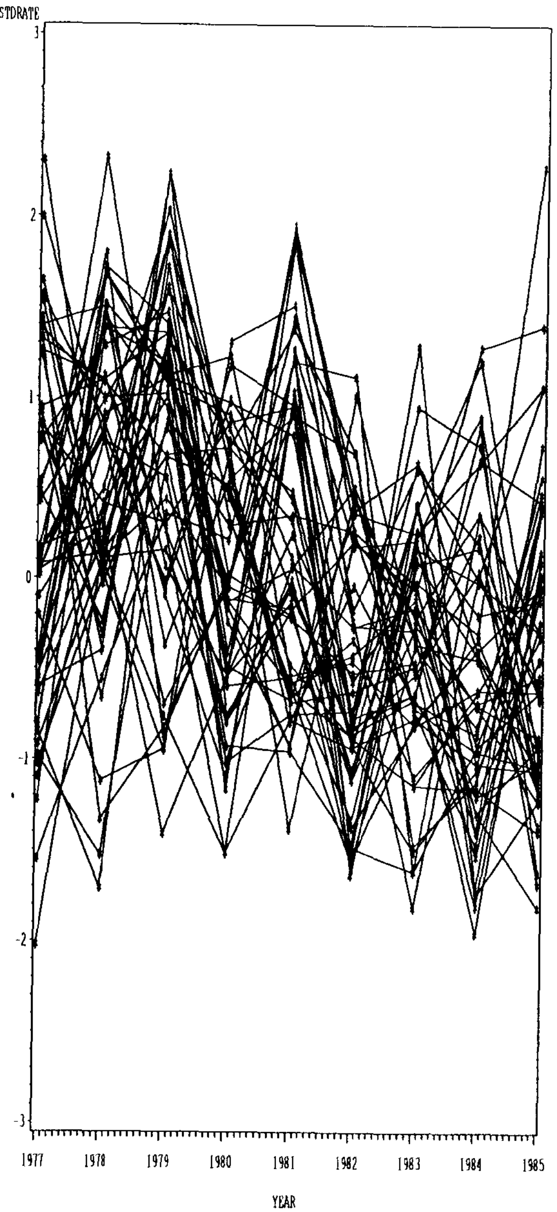

By State of Destination
Data is in Square Roots
Standardization by Time Series Mean, SD Filter




By State of Destination
Data is in Natural Logarithms
Standardization by Time Series Mean, SD Filter

## Plot of D. C. Out-Migration Standardized Rates

*Time is 1977 - 1985, inclusive*

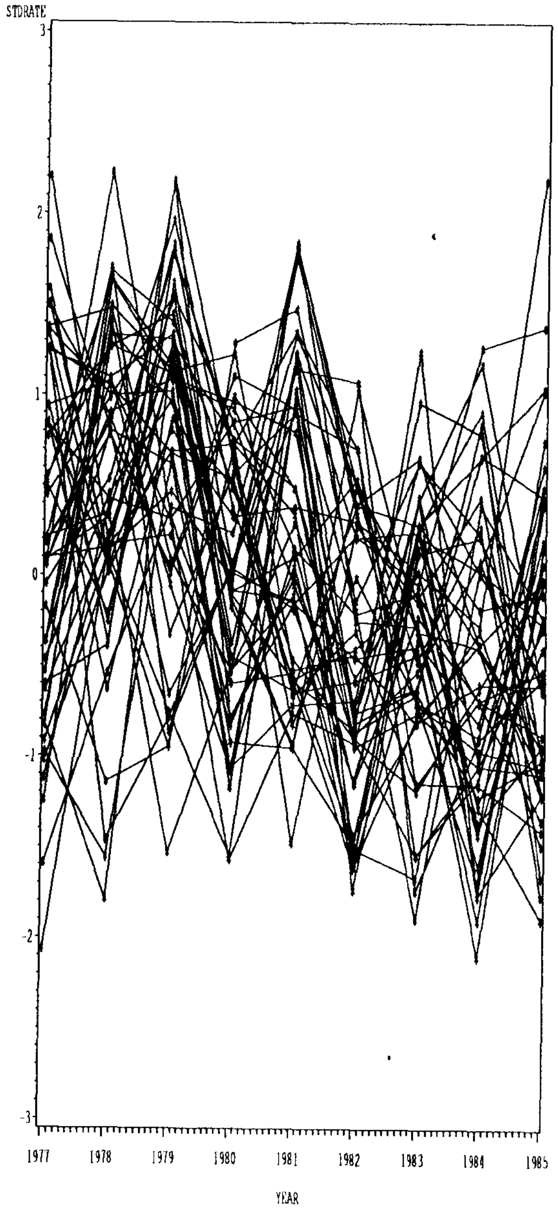



By State of Destination

Standardization by Time Series Mean, SD Filter

## Plot of D. C. Out-Migration Standardized Rates

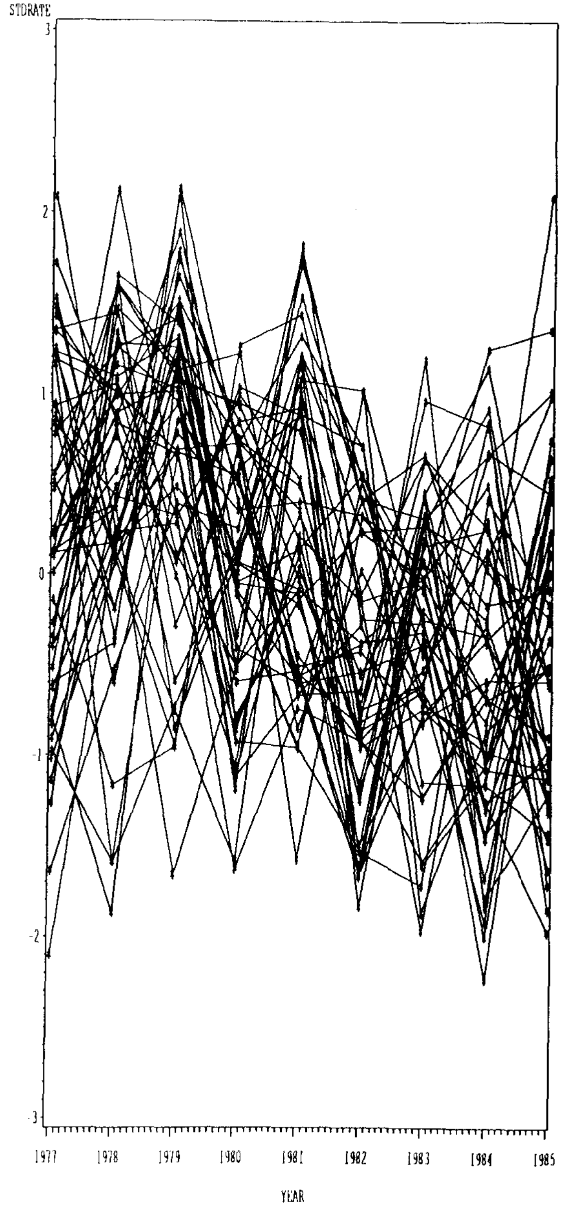
*Time is 1977 - 1985, inclusive*




By State of Destination
Data is in Square Roots
Standardization by Time Series Mean, SD Filter

## Plot of D. C. Out-Migration Standardized Rates

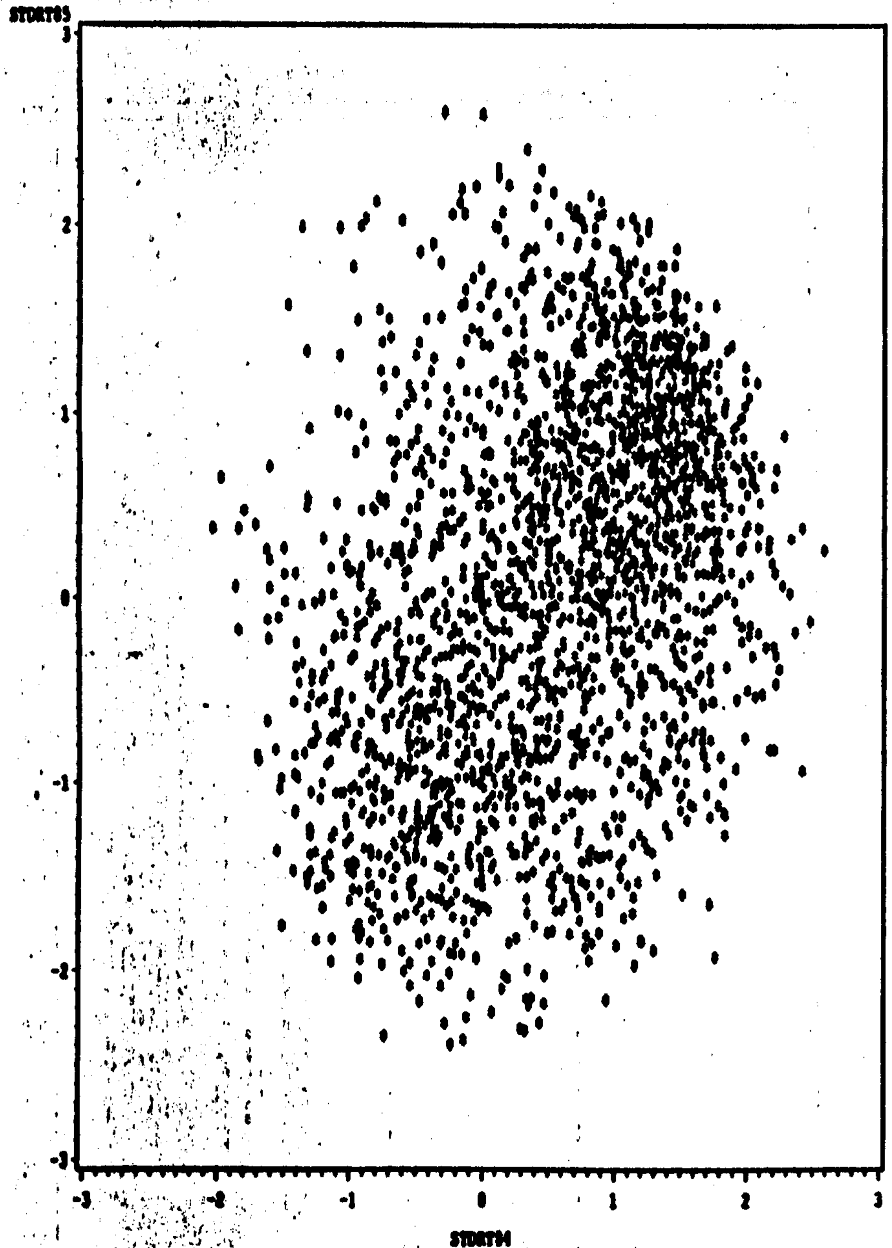
*Time is 1977 - 1985, inclusive*




By State of Destination
Data is in Natural Logarithms
Standardization by Time Series Mean, SD Filter

## 1985 vs 1984 Out-Migration Rates

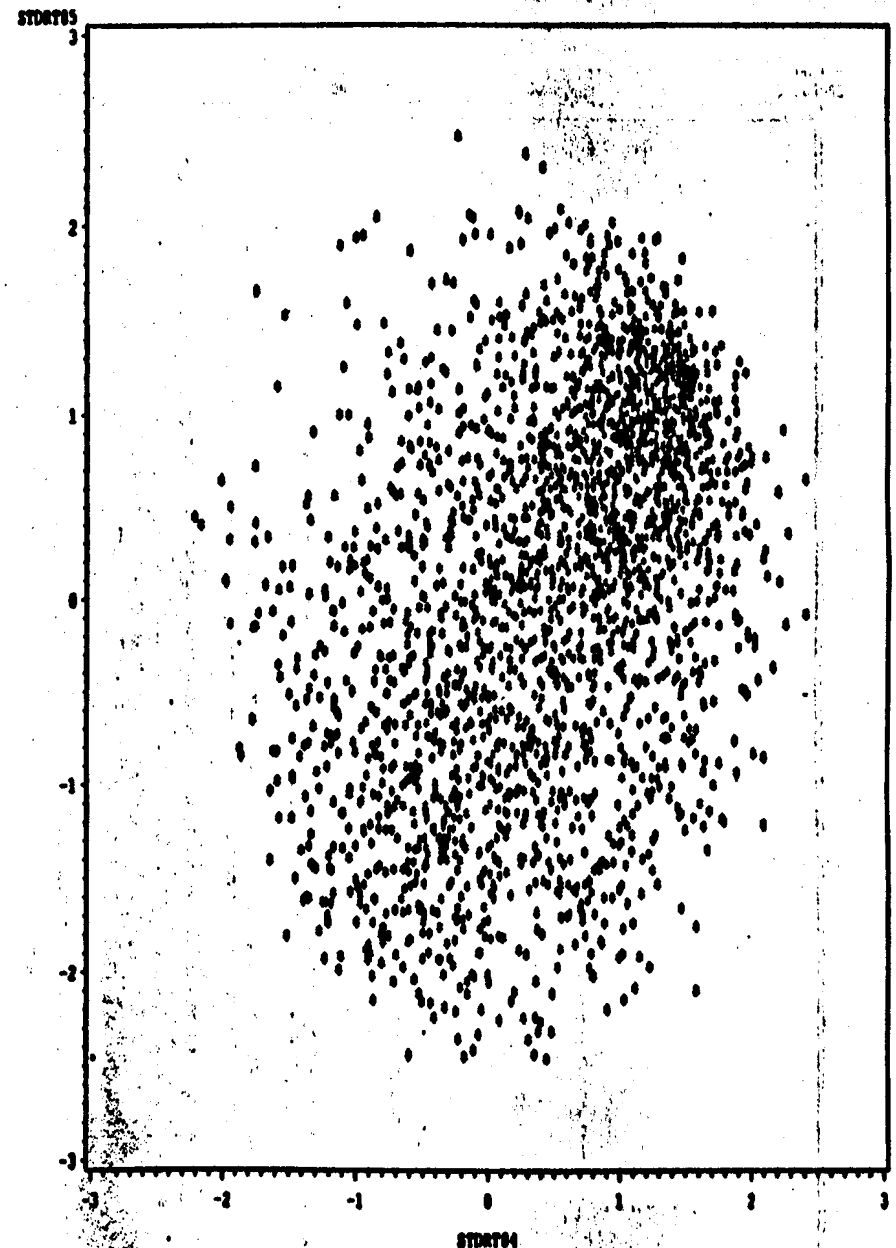
Rates are on a Standardized Basis




Stanardization by Time Series Mean, SD Filter

## 1985 vs 1984 Out-Migration Rates

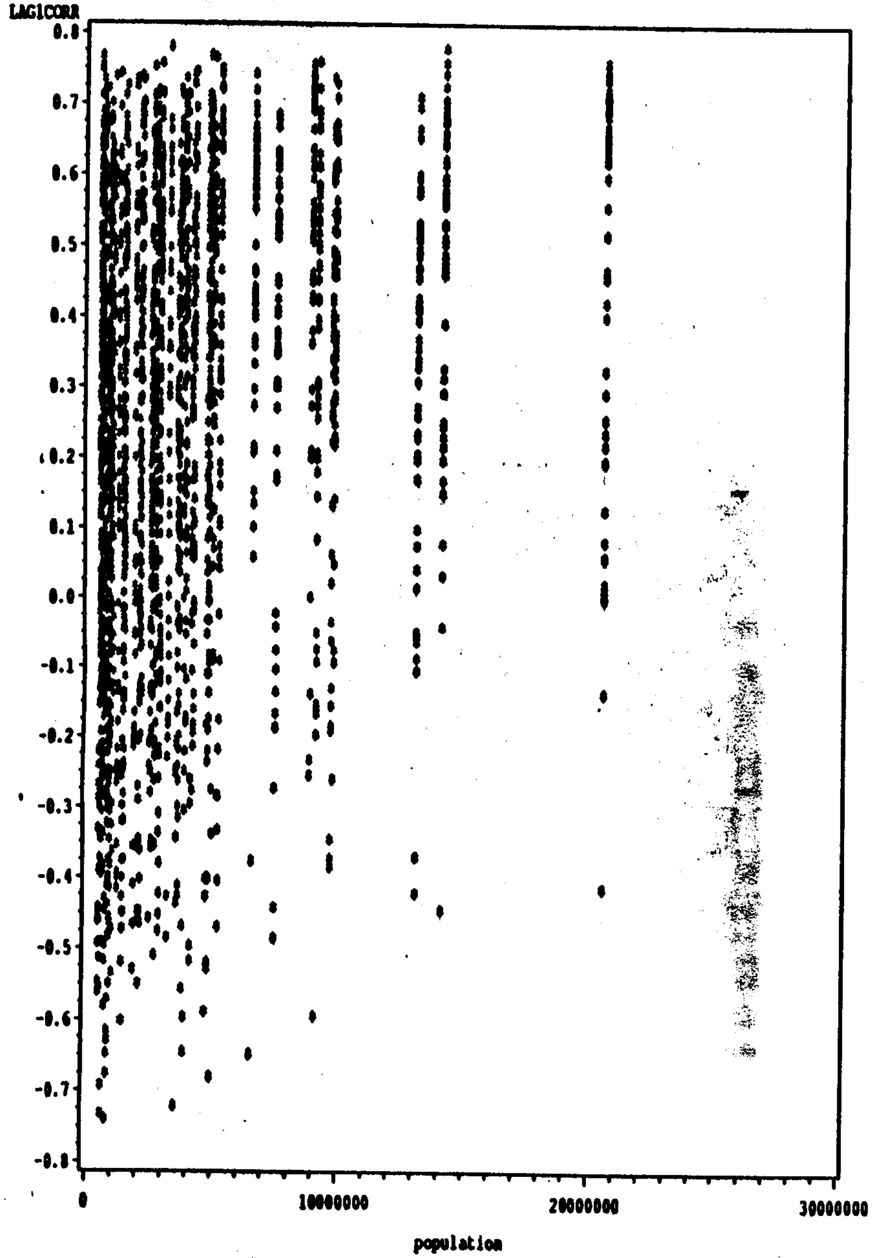
Rates are on a Standardized Basis




Stanardization by Time Series Mean, SD Filter
Data is in Natural Logarithms

## Plot of Lag 1 Autocorrelation vs Population

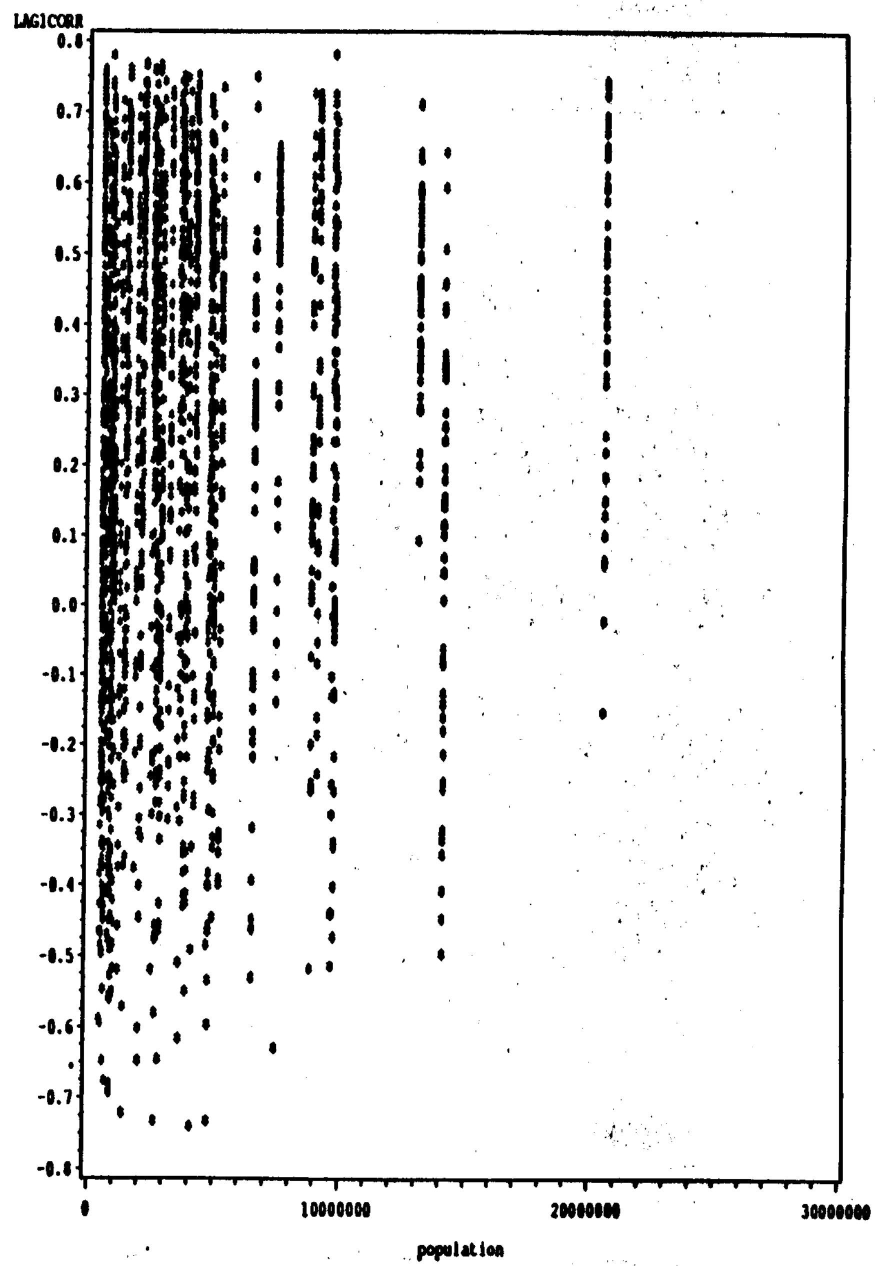
Rates are on a Standardized Basis




Standardization by Time Series Mean, SD Filter
Population is Origin-Specific

## Plot of Lag 1 Autocorrelation vs Population

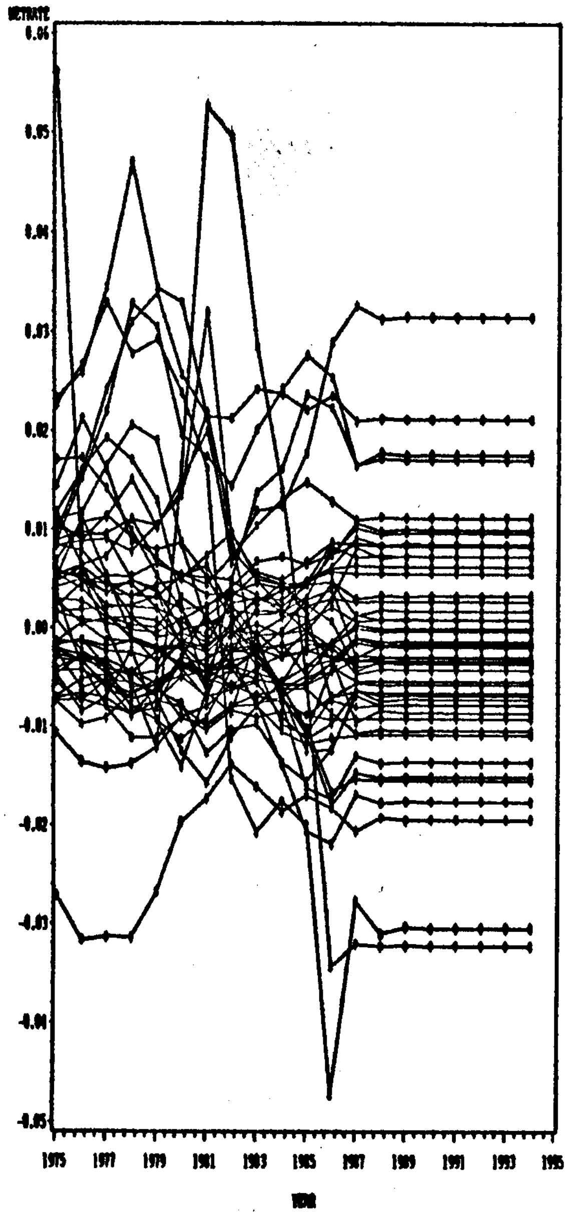
Rates are on a Standardized Basis




Standardization by Time Series Mean, SD Filter
Population is Destination-Specific

## Plot of Net Migration Rates by State

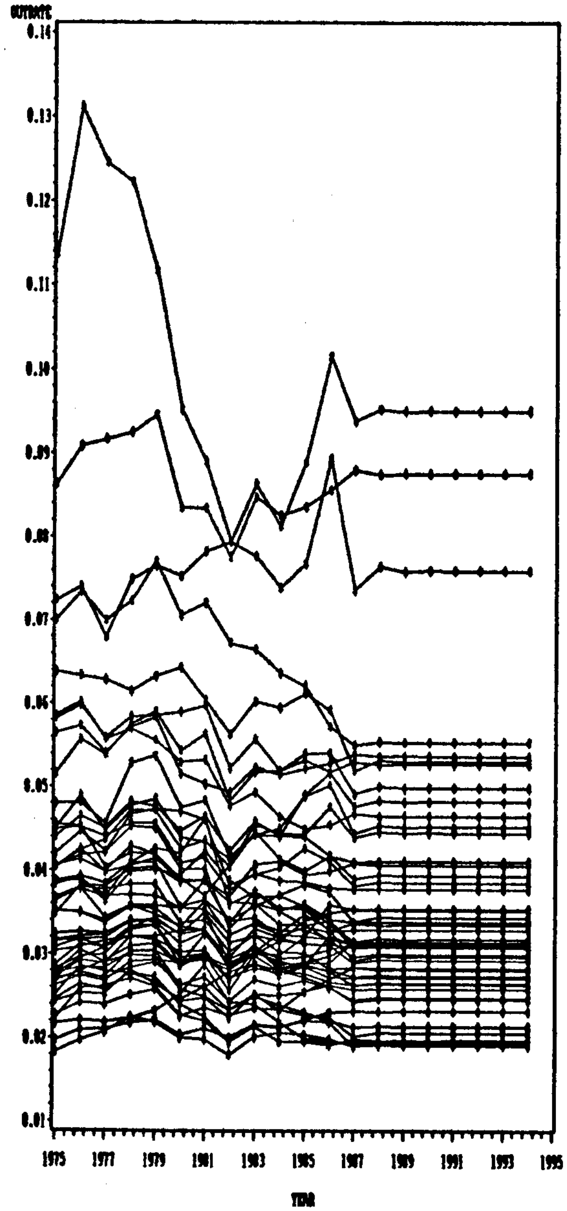
Year is 1975 - 1994, inclusive




Net-Rate is defined as Annual Population Migration Change
Divided by Initial State Population
Model 7b

## Plot of Out-Migration Rates by State

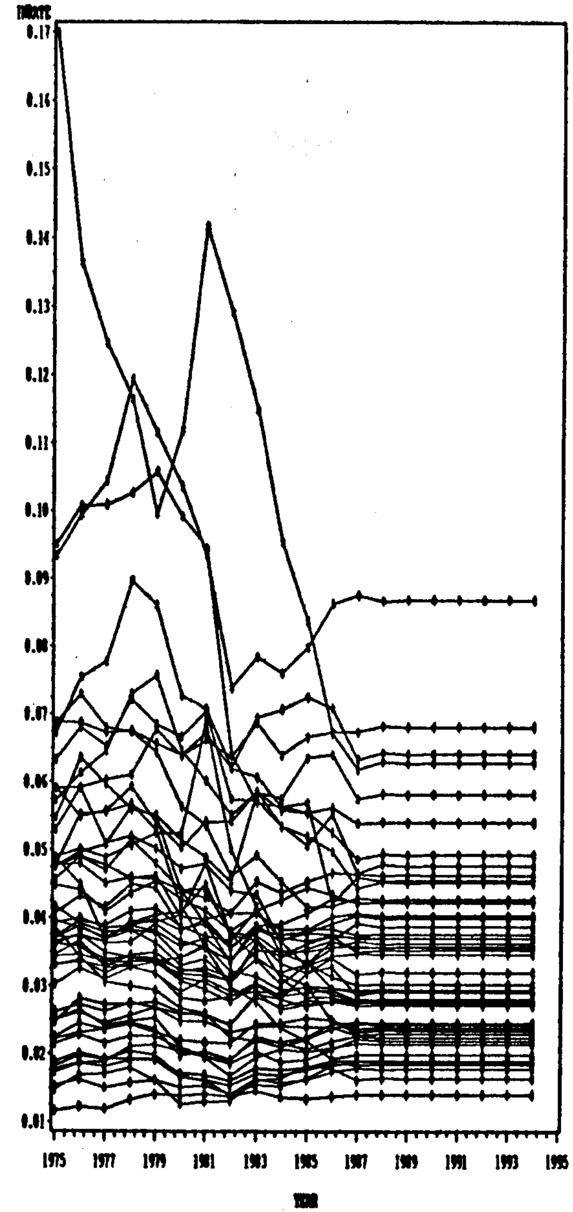
Year is 1975 - 1994, inclusive




Out-Rate is defined as Annual Population Outflow
Divided by Initial State Population
Model 7b

## Plot of In-Migration Rates by State

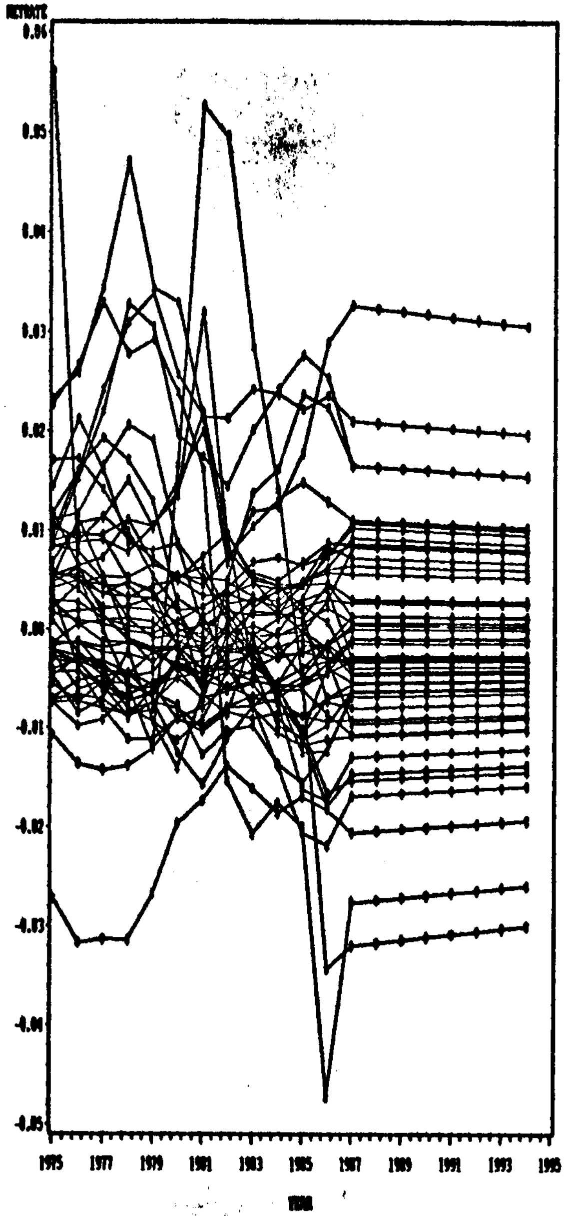
Year is 1975 - 1994, inclusive




In-Rate is defined as Annual Population Inflow
Divided by Initial State Population
Model 7b

## Plot of Net Migration Rates by State

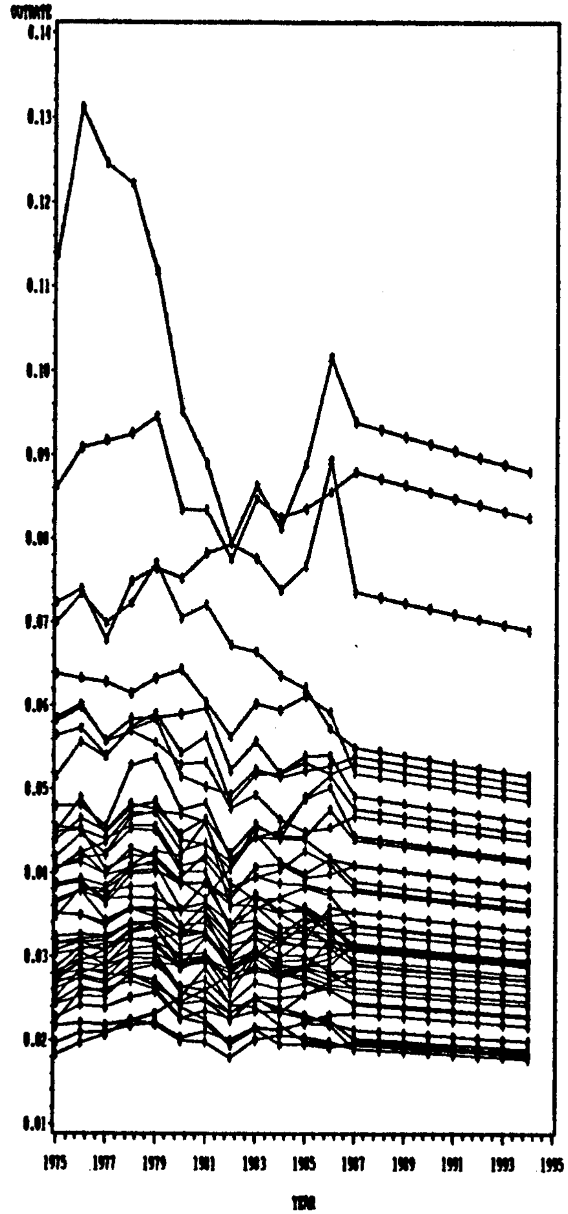
Year is 1975 - 1994, inclusive




Net-Rate is defined as Annual Population Migration Change
Divided by Initial State Population
Model 8b

## Plot of Out-Migration Rates by State

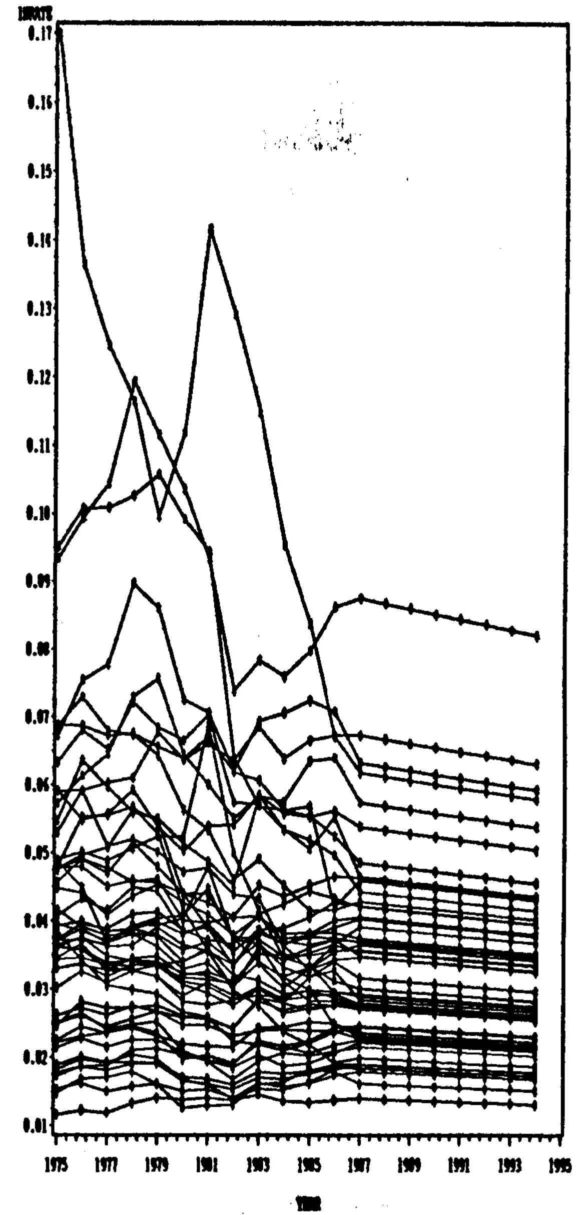
Year is 1975 - 1994, inclusive




Out-Rate is defined as Annual Population Outflow
Divided by Initial State Population
Model 8b

## Plot of In-Migration Rates by State

Year is 1975 - 1994, inclusive




In-Rate is defined as Annual Population Inflow
Divided by Initial State Population
Model 8b