A DATA-POOLING APPROACH TO ESTIMATE EMPLOYMENT MULTIPLIERS FOR SMALL REGIONAL ECONOMIES

Jeff V. Conopask

U.S. Department of Agriculture
Economics, Statistics, and Cooperatives Service
In Cooperation with the
U.S. Environmental Protection Agency

Technical Bulletin No. 1583
ABSTRACT

The level of residentiary employment in areas such as counties is an important component of rural development, environmental, and energy impact analysis. Volatile employment changes lead to migration, population, and wage changes. This study demonstrates a procedure for estimating disaggregated and lagged economic base multipliers for short to intermediate term forecasting models. Using an application of pooled cross-section/time-series data, a covariance model and an error components model are demonstrated for Northern Great Plains coal development. The only sector identified as having a consistent time lag was mining, and this lag was estimated to be 1 year.

Keywords: Economic base multiplier, Pooled cross-section time-series, Covariance, Error components.

ACKNOWLEDGEMENT

This study extends earlier work initiated by Lloyd D. Bender. Lynda Lynn and Hymen Weingarten provided computer analysis assistance. Other research team members are Fred K. Hines, Andrea Lubov, George Temple, Paul R. Myers, Stanley W. Voelker, and Thomas F. Stinson. Typing was accomplished by Lucia V. Wesley.
## CONTENTS

Summary........................................................................................................ iv

Introduction...................................................................................................... 1

Economic Base Analysis................................................................................ 2

Methodology.................................................................................................... 4

Econometric Considerations in Using Pooled Data................................. 7
  Correcting for Cross-Section Influences................................................... 7
  Correcting for Autocorrelation in the Time Series...................................... 8

Empirical Results............................................................................................ 9
  OLS, Covariance, or Error Components..................................................... 9
  Basic Sectors as Sources of Residency Activity......................................... 9
  Differential Residency Effects of Economic Base Sectors....................... 11
  Lagged Residency Employment Adjustment............................................ 11
  Economic Base Model in Economic Space............................................... 14
  Other Differences in Local Residency Demand......................................... 16
  Evaluation of the Pooling Procedure.......................................................... 16

Conclusions..................................................................................................... 18

Appendix A - Economic Base and Related Literature of Interest............. 19

Appendix B - Additional Econometric and Statistical Considerations.... 20
  Multicollinearity............................................................................................ 20
  Additional Cross-Section, Time-Series Considerations............................ 20
  Predictive Power Measures....................................................................... 22
  The Covariance Model............................................................................... 23
  Additional Error Components Notes........................................................ 24
  Measuring Autocorrelation in a Pooled-Data Set...................................... 26

Footnotes.......................................................................................................... 27
This study demonstrates a data-pooled econometric methodology for estimating economic base multipliers in small regional economies and for testing hypotheses concerning these multipliers. An economic base multiplier indicates how many residentiary industry employees are demanded per each additional basic industry employee. A residentiary industry produces goods and services principally for the local market; a basic industry produces goods and services locally, but sells elsewhere principally.

The economic base model to be estimated has two principal features. First, the model is disaggregated which postulates that different export sectors demonstrate different linkages to residentiary sectors. Second, the time-series component of the data pool allows lags in residentiary employment adjustment to be investigated.

The model is estimated by either of two econometric techniques using a cross section (15 counties) of time series (5 years). The covariance approach and an error components approach are applied to a purposive sample of 15 Northern Great Plains coal development counties for the years 1970-74. Except for agriculture, manufacturing, and mining, the data are all location quotient transformed. The error components approach is shown to provide a slightly better forecasting capability than the covariance approach.

The results demonstrate the efficacy of the pooling approach, with the multipliers from the significant coefficients ranging from -1.64 for wholesale and retail trade to 5.15 for services. The negative wholesale and retail trade multiplier indicates a dampening of the export portion on the residentiary portion of that sector. Only the mining sector has a lagged residentiary employment effect. Manufacturing and construction evidence little, if any, lag in the model with the greatest forecasting power.

The estimation of economic base multipliers through pooling cross-section and time-series data offers three advantages over traditional methods. First, pooling allows a substantial increase in degrees of freedom over a short time frame so that disaggregated models may be estimated. Second, using a limited time series (with pooling) minimizes structural change problems inherent in a long time series (without pooling), which tends to bias multiplier estimates. Finally, by expanding the data set to include time series, it is possible to estimate the effect of time lags upon multipliers.

The pooling approach allows detailed analysis of economic base theory applications through time. Used with appropriate statistical tools, the pooling approach can provide disaggregated and lagged multiplier estimates for planners in energy impacted regions. Levels of prediction accuracy can be evaluated for each region, and the specifications of the estimating equations can be adapted to regional conditions.
A DATA-POOLING APPROACH TO ESTIMATE EMPLOYMENT MULTIPLIERS FOR SMALL REGIONAL ECONOMIES

Jeff V. Conopask

INTRODUCTION

Emphasis on rural development and current interest in the impacts of environmental regulations and energy development (e.g. coal development in the Northern Great Plains) requires continued improvements in economic measurement. Through the use of econometric techniques, principally regression analysis, more accurate multipliers can be estimated. This is a considerable improvement over the earlier multiplier construction method of simply taking the ratio of nonexport to export employment for a given region/year.

These econometric investigations of small regional economic relationships often utilize cross-section data, although time series is used when available. However, intertemporal forecasting bias inherent in cross-section data and the inability to test for lagged economic effects has increasingly turned the attention of researchers toward less accessible, small regional time-series data. In turn, degrees of freedom problems often result when differential multipliers with lagged effects are estimated with a limited time-series. This report presents an alternative econometric method to alleviate these measurement problems for small economies.

As a first priority, this study demonstrates an econometric methodology for estimating small regional economic base multipliers in order to forecast nonexport or residentiary employment levels. A residentiary industry produces goods and services principally for the local market; a basic industry produces goods and services locally, but sells principally beyond the borders of the region in question. A corollary objective is to test hypotheses concerning multiplier disaggregation and lagged effects, location in economic space, and varying levels of local aggregate demand for residentiary goods and services in coal development counties of the Northern Great Plains (NGP).

A multiple regression model, which uses pooled time-series and cross-section data, is constructed to estimate economic base multipliers. Two estimating techniques are used: the covariance approach and the more recently developed error components approach.

1 Refer to page 27 for this and other footnotes.
The following hypotheses are tested.

(1) Basic sectors such as mining are important sources of change in a nonmetropolitan region, spurring change in residiary sectors such as services, banking, and finance; that is, sectors that rely on export of their products are the major source of change.

(2) Each base sector has a different affect on residiary sectors; that is, each industry is expected to have different demands for factor inputs, only part of which can be fulfilled locally.

(3) Time lags are a part of the multiplier process. The timing of residiary sector employment expansion varies according to basic sector; that is, a single time period (i.e., annual) adjustment does not reflect the actual expansion of residiary employment, which tends to lag behind the initial expansion of basic industry employment.

(4) The position of the economic unit in space will affect the relationship between change in basic employment and change in residiary employment (the multiplier); that is, not all communities can be expected to provide the same mix of services. Since each employment center (county) can be expected to provide different services according to its size and distance from a major trade center, the multiplier associated with given centers varies accordingly.

(5) Differences in demand for residiary services vary according to differences in local per capita income; that is, different per capita income levels will reflect preferences for different levels of necessity and luxury goods purchased.

ECONOMIC BASE ANALYSIS

Economic base theory postulates that economic activity in a region expands (contracts) primarily because of export activity to other regions. The determinants of supply (e.g., labor availability, production costs, and technology) are not considered. The simple theory assumes exports are independently and autonomously determined. As export demand expands, residiary sectors of the region are stimulated to generate additional economic activity. The employment multiplier representing this effect is most commonly expressed as the ratio of total or residiary employment to basic employment. For this study, the multiplier is represented as the change in residiary sector employment associated with a unit change in basic employment.

The economic base theory is used widely because it is inexpensive and easy to implement at any level of regional aggregation. Both shortrun and longrun changes were assumed to be explained by the theory in earlier years. Currently, however, it is generally held that multipliers have predictive value only as long as the industrial structure of the local economy remains unchanged; that it is valid only in the short run. A structural change is defined as a change in the types of industries existing in a region. For example, if
a mining equipment manufacturing firm located in a region where none existed before, the change in industrial composition could change the ratio of resi-
dentiary to basic employment and change the multiplier accordingly.

Criticisms of the theory center around its failure to consider specific community industrial structure, and its failure to consider the effect of community access to other trading centers. First, the classifications of basic and residentiary sectors may change with the size of communities. An industry might be considered basic in a small community, and residentiary oriented in a large one. An example would be a coal mine which provides fuel for a local industry versus a coal mine which exports its product. If the model cannot account for unique adjustment associated with differences in community industrial structure, then the multiplier represents an average for all basic activities and all regions under examination. There is no reason to assume a priori that all industrial sectors have the same linkages with the residentiary sectors. Differences in factor input requirements by capital intensive and labor intensive basic sectors, as well as the consumption tastes of employees of differing basic sectors, will manifest themselves in different economic base multipliers.

The failure of economic base theory to account for community access to service/trade centers was anticipated by both Alexander and Isard. More reliable estimates of base multipliers could be obtained by incorporating size of community and accessibility to economic activity outside the community into the multiplier. The conceptual basis for this measure involves an expansion of economic base theory to accommodate central place theory with the objective of allowing for differences in the multiplier associated with differences in community industrial structure, size, and access to markets other than that served in the immediate region.

Other writers have realized the importance of improving economic base multiplier estimates. In particular, timelags between basic industry expansion and residentiary expansion have been recognized as important in order to obtain more reliable estimates. This relationship is often computed without incorporating a timelag adjustment between the expansion in basic industries and the resulting expansion of residentiary industries. Therefore, the overall relationship may be distorted in the measurement period, because the adjustment of residentiary industries to changes in basic industries has not yet taken place. If this supposition is correct, this implies that there is a timelag between the expanded activity in the basic sector and the resulting activity in the secondary sectors. The inconclusive results of some empirical base studies may be partially attributed to unspecified time delays in nonexport sector adjustment patterns.

Unfortunately, there have been very few cases where base models have been implemented using small region accounts constructed in the same way for the same region at two or more continuous points in time. This situation has led to the use of cross-section studies to estimate regional multipliers. The potential exists for serious prediction errors, according to one author, when cross-section models are used for time projection.
One suggestion for improving multiplier estimates involves pooling cross-section and time-series data. By pooling data, it would be possible "to make the most efficient use of data across individuals to estimate that part of the behavioral relationship containing variables that differ substantially from one individual to another, in order that the lesser amount of information over time can be used to the best advantage in the estimation of the dynamic part of the relationship studied." Thus, a limited time series for multiplier estimation can be enhanced by pooling with cross-section (e.g. county) data. Objections relating to the absence of specified lagged relationships and the estimation of the multiplier at a point in time can be mollified by this approach. The increased degrees of freedom allow not only disaggregated economic base sectors to be estimated, but also the extension of the model to reflect economic space and income differences in local sector demand. Thus far, to this author's knowledge, there has been no attempt to estimate base multipliers in this manner.

METHODOLOGY

A modified economic base model using a time series of cross sections of employment data from the U.S. Department of Commerce's Bureau of Economic Analysis (BEA) is estimated in this study. The export base is disaggregated into various basic sectors in order to reflect differences in industrial structure. Whereas all employment in the mining, agriculture, and manufacturing sectors is assumed to be basic, location quotients are used to estimate the basic (export) portion of other industries. The core of the model is the economic base hypothesis. Two theoretical specifications extend the economic base core of the model to reflect the effect of economic space and differences in local aggregate demand.

Annual data for the years 1970-74 on employment and personal income for 15 energy impacted NGP counties comprise the data sample. Table 1 lists average annual sector employment for the 15-county region. The sample is limited to counties actually experiencing mining or thermoelectric generator construction and operation by 1974 to emphasize current energy impact activity.

The residential employment estimation model is specified as follows:

\[ R_{it} = f(AG_{it-r}, MAN_{it-r}, MIN_{it-r}, CON_{it-r}, CPU_{it-r}, WRT_{it-r}, SERV_{it-r}, GOV_{it-r}, D_{it}, I_{it}) \]

where \( R \) = sum of residential sectors employment in nonexport portion of construction, transportation, communication, public utilities, wholesale and retail trade, finance, insurance real estate, services, nonfarm proprietors, and Federal, State, and local government

\( AG \) = all wage and salary workers in agriculture, forestry, and fisheries, plus self-employed farm proprietors
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture, forestry, and fisheries</td>
<td>870</td>
<td>868</td>
<td>849</td>
<td>839</td>
<td>846</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>251</td>
<td>241</td>
<td>228</td>
<td>235</td>
<td>247</td>
</tr>
<tr>
<td>Mining</td>
<td>317</td>
<td>324</td>
<td>368</td>
<td>428</td>
<td>543</td>
</tr>
<tr>
<td>Construction</td>
<td>255</td>
<td>281</td>
<td>366</td>
<td>483</td>
<td>661</td>
</tr>
<tr>
<td>Transportation, communication, and public utilities</td>
<td>384</td>
<td>379</td>
<td>377</td>
<td>399</td>
<td>416</td>
</tr>
<tr>
<td>Wholesale and retail trade</td>
<td>972</td>
<td>995</td>
<td>1,014</td>
<td>1,085</td>
<td>1,181</td>
</tr>
<tr>
<td>Finance, insurance, and real estate</td>
<td>116</td>
<td>121</td>
<td>127</td>
<td>135</td>
<td>147</td>
</tr>
<tr>
<td>Services (including nonfarm proprietors)</td>
<td>1,320</td>
<td>1,357</td>
<td>1,401</td>
<td>1,452</td>
<td>1,518</td>
</tr>
<tr>
<td>Government</td>
<td>1,393</td>
<td>1,436</td>
<td>1,456</td>
<td>1,563</td>
<td>1,579</td>
</tr>
</tbody>
</table>

MAN = all manufacturing employment
MIN = all mining employment
CON = construction employment (export portion as calculated by location quotient procedure as in footnote 10)
TCPU = transportation, communication, and public utility employment (export portion as calculated by location quotient procedure as in footnote 10)
WRT = wholesale and retail trade employment (export portion as calculated by location quotient procedure as in footnote 10)
SERV = services employment (export portion as calculated by location quotient procedure as in footnote 10)
GOV = Federal Government employment (export portion as calculated by location quotient procedure as in footnote 10)
D = spatial variable reflecting adjacent county influence—employment in county \( i \) in year \( t \) as a percent of total employment in year \( t \) in all counties adjacent to county \( i \)
I = per capita income
\( i = 1, \ldots, 15 \) counties
\( t = 1, \ldots, 5 \) years
\( r = 0, \ldots, 1 \) year

Two principal assumptions underlie the model. These assumptions are (1) that the export sectors are independent of one another, and (2) that workers and firms within the same economic base sector each have similar consumption patterns (although consumption may be different across export sectors).

Agriculture, manufacturing, and mining are assumed to be the major basic sectors of the regions. Other sectors such as construction and transportation, communications, and public utilities are expected to demonstrate energy impacts from export effects of electric generating plant construction and operation. The latter sector should also reflect the use of rail transportation of coal. The remaining sectors, except for the Federal Government, are expected to demonstrate the influence of regional trade centers. Not all counties can be expected to have wholesale and retail trade plus service sectors which have export activity. Those counties that do export this activity can be expected to be influenced by coal development. The location quotient technique, in addition to helping identify the base sectors, identifies regional trade center base sectors that normally would serve only a local area.

The location of a county in economic space may affect the local employment multipliers stemming from export base activity. Workers in the basic sector may commute to the larger county for residiary services. A possible example
of this is a coal-mining county which has limited population centers (where some residiency services would be provided) but is adjacent to a county with larger population centers (larger variety of residiency services and stronger capability to react to regional growth).

The economic base multiplier may be affected in another way. Differences in income will result in different demands for the quantity of residiency services. Certainly as the per capita income of a county increases, the quality and mix of residiency services will change positively. Per capita income is used to test the hypothesis that increasing per capita income will increase the demand for residiency sector employment.

ECONOMETRIC CONSIDERATIONS IN USING POOLED DATA

Estimating economic base multipliers using pooled data presents problems in estimating efficient estimators. While ordinary least squares techniques yield linear and unbiased estimators, the structure of the disturbance term when using pooled data may be more complex than one resulting from time-series or cross-section data alone. However, proper specification of the disturbance term may yield additional information of a qualitative nature which can improve the efficiency of the estimators.

In designing more efficient estimators from pooled data, two characteristics of the pooled disturbance term are of immediate concern: disturbances associated with each cross section and those associated through time. Disturbances associated with cross-section data (interpreted as resulting from county differences in factors such as social capital, institutions, and degree of urbanization) distort the estimation of residiency employment generation across all counties. The second disturbance characteristic, correlation among the residuals of each succeeding time period (autocorrelation), violates an assumption of the classic regression model: that the errors are independently distributed. Correlation in the error term may result from misspecification of the model; that is, error term correlation resulting from cumulative effects of the omission of variables or errors in measurement.

Correcting for Cross-Section Influences

Two basic approaches to data-pooling estimation are used: the covariance approach, and the more recently developed error components technique. A covariance model is an extension of ordinary least squares by the introduction of N-1 dummy variables to allow the intercept term to vary over cross-section units (counties). In other words, annual observations for each county are grouped chronologically according to dummy variable to test whether the standard error of the estimate can be reduced from the level obtained when estimating with ordinary least squares without dummy variables.

The covariance approach has limitations. Degrees of freedom are lost by inclusion of the dummy variables. Furthermore, a large percentage of the error variation may be explained without identifying the exact source of the variation. Finally, the approach assumes that the error variance is constant.
across all regions, an assumption that may be untenable. These limitations led to the development of the error components approach.

The error components refinement of Balestra and Nerlove uses a generalized least squares procedure to incorporate the additional information on regional specific effects through the disturbance term of the equation. By dividing the error term into regional specific effects and other random effects, a more efficient estimating procedure is permitted. This procedure eliminates the loss of degrees of freedom associated with the covariance model. This methodology is represented symbolically as:

\[ R_{it} = a + bX_{ikt-r} + e_{it} \]

where \( R \) = residentiary employment
\( X \) = exogenous variables
\( i = 1,...,15 \) counties
\( t = 1,...,5 \) years
\( k = 1,...,K \) variables
\( r = 0,...,1 \) year
\( b \) = parameter
\( a \) = constant term
and \( e_{it} \) = error term

and \( e_{it} = u_i + v_{it} \)
\( u_i \) = regional effect but time invariant
\( v_{it} \) = remainder varying over regions and time

\[ E(u_i) = 0 \]
\[ E(v_{it}) = 0 \]
\[ E(u_i v_{it}) = 0 \]
\[ E(u_i u_j) = \sigma_u^2, \ i=j \]
\[ \quad \qquad \quad 0, \ otherwise \]
\[ E(v_{it} v_{js}) = \sigma_v^2, \ i=j \ and \ t=s \]
\[ \quad \qquad \quad 0, \ otherwise. \]

Specific information on the generalized least squares procedure used to calculate the error components estimates is in appendix B.

**Correcting for Autocorrelation in the Time Series**

Corrections must be made by using a procedure such as the Cochrane-Orcutt method if autocorrelation is present in the data. However, a modified Durbin-Watson statistic produced as a part of the covariance model results indicates that autocorrelation is not a problem when the regional effects are accounted for by dummy variables in this analysis.
EMPIRICAL RESULTS

This section first presents a comparison of the results using the ordinary least squares, covariance, and error components techniques. Then, test results of the five hypotheses are presented and discussed. Finally, an evaluation of the methodology is presented.

OLS, Covariance, or Error Components

The covariance and error components results of the core economic base model demonstrate the advantage of using surrogates to account for the regional effects (table 2). These regional effects are accounted for by allowing the intercept term to vary (as in the covariance model) or through the disturbance term (as in the error components model). A simple F test comparing the sums of squared error of the ordinary least squares (OLS) approach with the covariance approach rejects the null hypothesis that the equal intercepts are appropriate.

The percent standard error is 16.2 and 4.3, respectively, for the OLS and covariance models. The error components approach is a slightly better predictor than the covariance technique based on a comparison of their respective standard errors of the estimate. The relative standard error is 3.9 percent when the error component is used.

Autocorrelation appears to be a problem when the OLS technique is used. The autocorrelation reflected in the OLS Durbin-Watson can be due to relationships among regions or to the time-series effect. Severe autocorrelation is not indicated by the Durbin-Watson d produced as a part of the covariance technique. Apparently, the autocorrelation is imbedded in the regional cross-section effects of the data and not in the time series. One can expect autocorrelation if a longer time series is used. The sum of squared errors of the covariance and error components models are similar; hence, it is assumed that little autocorrelation remains in the error components results.

Basic Sectors as Sources of Residency Activity

The empirical results of the models specified in table 2 demonstrate that export base industries are associated with local industry sectors in each county. The coefficient of determination for the error components shows that 76 percent of the variation in residency employment is associated with the eight export sectors. The associated F test is significant at the F,01 level. Therefore, the hypothesis that higher levels of export sector employment are associated with higher levels of residency sector employment is accepted. Five out of six statistically significant coefficients had the expected positive sign.

The coefficients on agriculture and the combined transportation, communication, and public utilities were not statistically significant even at the t,10 level. The insignificant coefficients of the agriculture sector could reflect data limitations. Year-to-year changes in agricultural employment in
Table 2—Comparison of regression results using BEA employment data from 15 NGP
major coal producing counties for 1970-74

<table>
<thead>
<tr>
<th>Variable 1/</th>
<th>Statistical model</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Covariance</td>
<td>Error components</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1,250.100</td>
<td>3,461.770</td>
<td>(2.493)</td>
<td>(2.094)</td>
</tr>
<tr>
<td>Agriculture, forestry, and fisheries t (AG )</td>
<td>-.543</td>
<td>-4.519</td>
<td>(1.949)</td>
<td>(1.010)</td>
</tr>
<tr>
<td>Manufacturing t (MAN t )</td>
<td>6.531</td>
<td>1.996</td>
<td>(7.858)</td>
<td>(3.034)</td>
</tr>
<tr>
<td>Mining t-1 (MIN t-1 )</td>
<td>2.463</td>
<td>1.012</td>
<td>(6.847)</td>
<td>(2.451)</td>
</tr>
<tr>
<td>Construction t (CON t )</td>
<td>.256</td>
<td>.471</td>
<td>(.604)</td>
<td>(2.756)</td>
</tr>
<tr>
<td>Transportation, communication, and public utilities t (TCPU t )</td>
<td>-2.356</td>
<td>-2.60</td>
<td>(1.908)</td>
<td>(.216)</td>
</tr>
<tr>
<td>Wholesale and retail trade t (WRT t )</td>
<td>-1.959</td>
<td>-1.454</td>
<td>(1.050)</td>
<td>(1.80)</td>
</tr>
<tr>
<td>Services t (SERV t )</td>
<td>7.625</td>
<td>4.519</td>
<td>(2.653)</td>
<td>(1.736)</td>
</tr>
<tr>
<td>Government t (GOV t )</td>
<td>.108</td>
<td>1.417</td>
<td>(1.636)</td>
<td>(3.508)</td>
</tr>
<tr>
<td>Sum of squared errors</td>
<td>23,481.600</td>
<td>1,210.190</td>
<td>1/ 1,360.300</td>
<td></td>
</tr>
<tr>
<td>Standard error of estimate</td>
<td>678.500</td>
<td>180.850</td>
<td>163.300</td>
<td></td>
</tr>
<tr>
<td>Percent standard error</td>
<td>16.200</td>
<td>4.300</td>
<td>1/ 3.900</td>
<td></td>
</tr>
<tr>
<td>R^2 2/</td>
<td>.980</td>
<td>.990</td>
<td>1/ .760</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>314.400</td>
<td>1640.400</td>
<td>20.600</td>
<td>(8, 51)</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>.515</td>
<td>2.161</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the t .05 level = 1.677.
**Significant at the t .01 level = 2.403.
***Significant at the t .005 level = 2.682.
****Significant at the F .01 level = 2.90.
NA = Not available.
1/ All employment is assumed to be basic for agriculture, manufacturing, and mining. Basic employment for all other sectors was calculated using location quotients as in text footnote 10.
2/ With the error components approach, the dependent variable is transformed and the available computer software measures only intraclass variation as the total sums of squares; hence, the error component R^2 is not comparable with the covariance result. Attention should be addressed to the other statistics derived from the sum of squared errors of both approaches.
the BEA data reflect only percentage changes at the State level. Annual changes in agricultural employment in all counties within a State reflect the State's rate of change. The nonsignificance of a sector coefficient also may reflect nonlocal spending patterns, not only of the industry but of a majority of its workers. This could be true for transportation such as employment in intercontinental railroading and longhaul trucking which dominate this sector in the NGP.

The negative sign of the coefficient of wholesale and retail trade is not totally implausible. The data were transformed by location quotients and constrained to be $\geq 0$. Counties which serve as regional trade centers likely have positive export values for the wholesale and retail trade centers, while all others have a calculated zero export base in that sector. There are several reasons to speculate that regional trade centers with export bases in wholesale and retail trade may exhibit lower residentiary employment than other counties. First, economies of scale in the sector could mean the services of the sector are more labor efficient in regional trade centers. Expansion of the export portion of that sector could mean that local residents merely absorb the excess capacities of the sector firms. Second, there can be substitution of services in the wholesale and retail sector for services of the other residentiary sectors. This would imply a degree of specialization in towns and a hierarchy of centers as implied by central place theory. Third, there may be spillover effects to adjacent counties which implies that service markets spread out geographically around a regional trade center. Fourth, the export portion of wholesale and retail trade may compete for labor with other basic sectors. As growth occurs, capital may be substituted for smaller labor intensive activities and firms. Finally, there could be rather complex and perhaps irregular adjustment patterns in the wholesale and retail trade sector which cannot be described in specifications of econometric models. This implies that basic activities in the sector may have expanded, but that the residentiary sectors have not adjusted or that the adjustment path is uncertain.

Differential Residentiary Effects of Economic Base Sectors

Data in table 3 demonstrate that 22 of 28 paired coefficients are statistically different from each other. Four involved coefficients with insignificant t statistics for the null hypothesis $B_j = 0$, hence, are probably unreliable. The two remaining insignificant relationships both involved the coefficients on government employment paired once with services and once with manufacturing. The t statistic for the coefficients of the government and services sector could have resulted from military activity (Minot Air Force Base) or Bureau of Indian Affairs activity. No explanation is offered for the manufacturing/government result.

Lagged Residentiary Employment Adjustment

The lagged adjustment of residentiary employment was tested for selected economic base sectors. The results of a 1-year lag test for manufacturing, mining, and construction are shown in table 4. The covariance model was used with both the t test and standard error of the estimate as the criteria
<table>
<thead>
<tr>
<th>Basic sectors</th>
<th>Agriculture, forestry, and manufacturing&lt;sub&gt;f&lt;/sub&gt;</th>
<th>Mining&lt;sub&gt;t-1&lt;/sub&gt;</th>
<th>Construction&lt;sub&gt;t&lt;/sub&gt;</th>
<th>Transportation, communication, and public utilities&lt;sub&gt;t&lt;/sub&gt;</th>
<th>Wholesale and retail trade&lt;sub&gt;t&lt;/sub&gt;</th>
<th>Services&lt;sub&gt;t&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing&lt;sub&gt;t&lt;/sub&gt;</td>
<td>4.00**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mining&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>2.18*</td>
<td>6.05**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction&lt;sub&gt;t&lt;/sub&gt;</td>
<td>1.49</td>
<td>4.05**</td>
<td>2.14*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transportation, communication, and public utilities&lt;sub&gt;t&lt;/sub&gt;</td>
<td>4.00**</td>
<td>4.00**</td>
<td>.90</td>
<td>.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wholesale and retail trade&lt;sub&gt;t&lt;/sub&gt;</td>
<td>.12</td>
<td>27.9*</td>
<td>7.51**</td>
<td>4.15**</td>
<td>6.12**</td>
<td></td>
</tr>
<tr>
<td>Services&lt;sub&gt;t&lt;/sub&gt;</td>
<td>8.47**</td>
<td>1.74*</td>
<td>2.13*</td>
<td>2.13*</td>
<td>3.16**</td>
<td>4.24**</td>
</tr>
<tr>
<td>Government&lt;sub&gt;t&lt;/sub&gt;</td>
<td>2.75**</td>
<td>1.06</td>
<td>9.40**</td>
<td>9.50**</td>
<td>1.99*</td>
<td>7.93**</td>
</tr>
</tbody>
</table>

*Significant at \( t_{0.05} \) level.
**Significant at \( t_{0.01} \) level.

1/All employment is assumed to be basic for agriculture, manufacturing, and mining. Basic employment for all other sectors is calculated using location quotients as in text footnote 10.
Table 4—Covariance results of timelag employment multiplier estimates for selected sectors for 15 NGP major coal producing counties 1970-74 1/

<table>
<thead>
<tr>
<th>Variable 2/</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unlagged</td>
</tr>
<tr>
<td>Agriculture, forestry, and fisheries</td>
<td>-3.56</td>
</tr>
<tr>
<td></td>
<td>(-1.69)*</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>2.15</td>
</tr>
<tr>
<td></td>
<td>(3.65)**</td>
</tr>
<tr>
<td>Mining</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>(3.95)**</td>
</tr>
<tr>
<td>Construction</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>(2.19)*</td>
</tr>
<tr>
<td>Transportation, communications, and public utilities</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>(.61)</td>
</tr>
<tr>
<td>Wholesale and retail trade</td>
<td>-1.90</td>
</tr>
<tr>
<td></td>
<td>(2.19)*</td>
</tr>
<tr>
<td>Services</td>
<td>5.13</td>
</tr>
<tr>
<td></td>
<td>(2.18)*</td>
</tr>
<tr>
<td>Government</td>
<td>1.57</td>
</tr>
<tr>
<td></td>
<td>(4.28)**</td>
</tr>
<tr>
<td>Standard error of the estimate</td>
<td>163.57</td>
</tr>
</tbody>
</table>

*significant at the t .01 level.
**significant at the t .05 level.
1/ Data from U.S. Department of Commerce, Bureau of Economic Analysis.
2/ All employment is assumed to be basic for agriculture, manufacturing, and mining. Basic employment is calculated by using location quotient as in text footnote 10 for all other sectors.
3/ Manufacturing, mining, and construction lagged.
4/ Mining and construction lagged.
for testing lag significance. In the result in which all three sectors are lagged, the manufacturing coefficient is insignificant, while coefficients on mining and construction are significant. The second lagged regression, however, demonstrates that if manufacturing is unlagged, then only the mining coefficient maintains t significance and the construction coefficient is not significant. The nature of the linkage between lagged manufacturing and lagged construction is not evident even from the correlation matrix; hence, the evidence suggests that unlagged versions are appropriate. Neither regression has a better standard error of the estimate than the regression without lags.

Since the mining sector exhibits stable (little magnitude fluctuation, but continuous statistical significance) behavior during experiments with lagged and unlagged manufacturing and construction sectors, it was decided to fit a final model with a lagged mining sector. The evidence (t statistic) suggests a lag exists for mining, although there is a slight loss of predictive power with that specification. This is a statistically defensible position since a loss of efficiency can be tolerated if bias is reduced.

The residentiary sector expansion resulting from mining activity increases can be expected to include a broad range of local activities if mining is of a permanent nature. Mining activity will stimulate such sectors as local government; health; personal, legal, and financial services; amusements; and local housing construction. On the other hand, more transitory employment may stimulate a narrow array of activities such as in mobile home sales and the food and lodging sectors.

All other sector coefficients are unlagged. The assumed adjustment is that residentiary sector employment increases occur in the same annual period as the respective export sector demand increases. The fact that manufacturing, services, and government have the largest values of the six significant coefficients indicates strong linkages within the local economy.

The largest multiplier effect is in the services' sector. The magnitude of this multiplier could reflect the labor intensiveness of that sector, but is also related to the use of location quotients in calculating the independent variable. However, few other disaggregated base studies have demonstrated such a magnitude.

The magnitude of the manufacturing multiplier is within an anticipated range based on previous studies. The government sector coefficient also appears to be in an acceptable range.

Economic Base Model in Economic Space

The economic base model was extended to include economic effects from adjacent counties under the hypothesis that the position of the economic unit in space affects the relationship between economic base employment and residentiary employment. The effects of adjacent county employment on residentiary employment generation in county i are statistically significant (equation II, in table 5). However, the inclusion of this variable in the economic base model results in an interaction with the construction sector variable. This
Table 5—Comparison of other error components estimates of residentiary employment multipliers for 15 NGP major coal producing counties 1970-74

<table>
<thead>
<tr>
<th>Variable</th>
<th>Equation</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td>3,461.77</td>
<td>2,401.15</td>
<td>1,782.65</td>
<td>3,796.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.09)</td>
<td>(2.60)</td>
<td>(1.00)</td>
<td>(1.99)</td>
</tr>
<tr>
<td>Agriculture, forestry, and fisheries t</td>
<td></td>
<td>-1.54</td>
<td>-2.25</td>
<td>.31</td>
<td>-3.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.01)</td>
<td>(1.98)*</td>
<td>(.18)</td>
<td>(1.79)*</td>
</tr>
<tr>
<td>Manufacturing t</td>
<td></td>
<td>2.26</td>
<td>1.97</td>
<td>2.10</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.89)**</td>
<td>(3.63)**</td>
<td>(3.55)**</td>
<td>(2.13)*</td>
</tr>
<tr>
<td>Mining t-1</td>
<td></td>
<td>.99</td>
<td>1.00</td>
<td>.93</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.71)**</td>
<td>(3.11)**</td>
<td>(2.53)**</td>
<td>(.34)</td>
</tr>
<tr>
<td>Construction t</td>
<td></td>
<td>.52</td>
<td>.18</td>
<td>.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.41)**</td>
<td>(1.10)</td>
<td>(3.00)**</td>
<td></td>
</tr>
<tr>
<td>Transportation, communication, and public utilities t</td>
<td></td>
<td>.38</td>
<td>.08</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.37)</td>
<td>(.09)</td>
<td>(.41)</td>
<td></td>
</tr>
<tr>
<td>Wholesale and retail trade t</td>
<td></td>
<td>-1.64</td>
<td>-1.59</td>
<td>-1.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.28)*</td>
<td>(2.46)**</td>
<td>(2.14)*</td>
<td></td>
</tr>
<tr>
<td>Services t</td>
<td></td>
<td>5.15</td>
<td>5.42</td>
<td>5.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.22)*</td>
<td>(2.57)**</td>
<td>(2.20)*</td>
<td></td>
</tr>
<tr>
<td>Government t</td>
<td></td>
<td>1.93</td>
<td>1.24</td>
<td>1.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.03)**</td>
<td>(4.07)**</td>
<td>(7.20)**</td>
<td></td>
</tr>
<tr>
<td>(County employment ÷ employment in adjacent counties) X 100 percent</td>
<td></td>
<td></td>
<td></td>
<td>67.73</td>
<td>89.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.14)**</td>
<td>(6.16)**</td>
</tr>
<tr>
<td>Per capita income</td>
<td></td>
<td></td>
<td></td>
<td>.06</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.48)</td>
<td>(.79)</td>
</tr>
<tr>
<td>R²</td>
<td></td>
<td>.76</td>
<td>.86</td>
<td>.78</td>
<td>.70</td>
</tr>
<tr>
<td>Standard error of the estimate</td>
<td></td>
<td>163.30</td>
<td>151.74</td>
<td>162.68</td>
<td>180.53</td>
</tr>
<tr>
<td>Percent standard error</td>
<td></td>
<td>3.90</td>
<td>3.60</td>
<td>3.90</td>
<td>4.30</td>
</tr>
</tbody>
</table>

*significant at the t .05 level.
**significant at the t .01 level.
1/Data from U.S. Department of Commerce, Bureau of Economic Analysis.
2/All employment is assumed to be basic.
3/Basic employment is calculated by using location quotient as in text footnote 10.
interaction is not fully understood in economic terms, but is represented by
the loss of the significance of the construction sector’s coefficient. Since application of the estimating procedure to construction sector impacts of energy projects is a higher order goal of this research than spatial specification, the latter effort’s statistical complications prompted its deletion from equation I.

The adjacent county employment variable may measure some of the same effects as the following location quotient calculated sectors: construction; transportation, communication, and public utilities; wholesale and retail trade; services; and government. Consequently, an alternative equation was also specified and tested. This alternative equation (IV) in table 5 has as its independent variables agriculture, forestry, and fisheries; manufacturing; mining lagged 1 year; adjacent county employment effects; and per capita income. This specification considers three of the four major economic base sectors (construction is deleted because of the interaction with the space variable cited above). The mining and per capita income coefficients are insignificant and the \( R^2 \) lower. It is interesting to note that the agriculture, forestry, and fisheries coefficient is significant here as it is in equation II in table 5, but still has an a priori unexpected sign. The economic space coefficient has a larger value and a smaller standard error as compared to the similar coefficient in equation II in table 5.

Other Differences in Local Residuary Demand

The extension of the economic base model to account for income induced differences in local demand for residuary goods and services is not successful. Using per capita income as the variable, the results of equation III in table 5 demonstrate a very slight improvement in the standard error of the estimate and an insignificant t statistic. Consequently, these results are considered to add little, if anything, to the economic base model, and the variable is deleted. This approach, reflecting subregional differences, may mask demand differences due to income.

Evaluation of the Pooling Procedure

The evidence presented in these results indicates that a pooling methodology improves the efficiency of the estimators and alters the values of the coefficients and their statistical significance. The regional effects in the error components model accounts for 94.2 percent of the residual variation remaining in the ordinary least square results. The percentage of standard error drops from 16.2 in the ordinary least squares regression to 4.3 in the covariance and to 3.9 in the error components results. Of those coefficients which are significant, four are lower in value in the error components than in the ordinary least squares approach. The value of the manufacturing and services export sector coefficients drops dramatically to more acceptable levels based on a priori information.

The construction and the combined wholesale and retail trade sectors have significant coefficients which increase in magnitude when the regional effects are introduced. On the other hand, when the regional effects are accounted for,
transportation, communication, and public utilities become insignificant determinants.

The exact nature of regional effects is difficult to establish by changes in statistical significance. Outside of data anomalies, it is possible that a specific resideniary sector mix that responds to activity changes in certain sectors is sensitive to the regional effects. One possible surrogate variable to account for urbanizing influences (greater service sector mix) in these low density population regions is the size of towns. Further research requires more disaggregated data; e.g., by jurisdiction.

Additional perspective on the sensitivity of the error components estimated employment multipliers can be gained from table 6. Ranges of resideniary employment generation for the statistically significant coefficients are displayed using 95-percent confidence limits. \(^{31}\) Given a 100-unit increase in mining employment in year 1, the most likely resideniary sector expansion will be the addition of 99 employees by the end of year 2, a 1-year lag. The confidence limits show that the initial increase will result in a range from 72 to 126 additional employees within 0.95 probability limits.

Similarly, expansion in resideniary employment due to 100 construction workers will have a mean value of 52 additional workers if equilibrium level is achieved, but will range from 47 to 57 workers.

Table 6—Statistically significant multiplier coefficients and 95-percent confidence intervals \(^{1}\)

<table>
<thead>
<tr>
<th>Economic sector</th>
<th>Mean multiplier</th>
<th>Confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Manufacturing(_t)</td>
<td>2.26</td>
<td>1.58</td>
</tr>
<tr>
<td>Mining(_{t-1})</td>
<td>0.99</td>
<td>0.72</td>
</tr>
<tr>
<td>Construction(_t)</td>
<td>0.52</td>
<td>0.47</td>
</tr>
<tr>
<td>Wholesale and retail trade(_t)</td>
<td>-1.64</td>
<td>-2.67</td>
</tr>
<tr>
<td>Services(_t)</td>
<td>5.15</td>
<td>-5.68</td>
</tr>
<tr>
<td>Government(_t)</td>
<td>1.93</td>
<td>1.78</td>
</tr>
</tbody>
</table>

\(^{1}\) Based on error components regression results from table 2.
Only the export services sector exhibits a wide range of multiplier values for its 95-percent confidence limits. This may be due to difficulties in applying the location quotient technique to this sector—more so than the other disaggregated employment sectors. The remaining sectors' multiplier values vary from 8 to 30 percent of the mean multiplier, except for wholesale and retail trade which varies 63 percent.

CONCLUSIONS

This study demonstrates the application of data pooling econometric methodology for estimating economic base multipliers for small regions. An attempt also is made to extend economic base theory to account for central place tendencies and the influence of local income differentials on the generation of residentiary employment.

The results indicate that data pooling is a viable and practical means to improve multiplier estimation. Data pooling allows short- to intermediate-run forecasting by increasing the degrees of freedom without extending the time period beyond the short- to intermediate-run assumption of constant background conditions in economic and community structure. Furthermore, the results show that for nonmetro counties (1) export sectors generate residentiary employment, (2) a timelag for residentiary sector employment adjustment exists definitely in the mining sector, (3) different export base sectors have different effects on residentiary job creation, (4) economic spatial relationships require better model specification, and (5) differential income demand for residentiary services is insignificant in the NOP data.

Three avenues of future research are suggested from the results: (1) development of measures of economic space, (2) exploration of time adjustment paths, and (3) experimentation into the nature of the regional effects.

The first avenue of research relates a point in space to other economies in addition to considering the aggregated export to the rest of the world. The location quotient approach to identify economic base sectors implicitly contains a spatial component in that typically residentiary sectors that export to adjacent counties are identified in the NOP case. However, integration of a spatial effect into a data-pooled forecasting model will require further work.

Additional work on the second proposed avenue of research is required since other sectors showed evidence of timelags, which could not be incorporated into the data-pooled forecasting model without problems. In fact, the mining sector also proves to be significant in the (same year) adjustment. Use of another data base may improve the model's sensitivity. The time adjustment demonstrated is a simple specification. More complex specifications, such as a geometrically declining lag or a stock-adjustment approach, should be investigated.

The third avenue of future research should use the error components to isolate regional variables in order to replace dummy variables, which are difficult to interpret. Thus, identification of qualitative variables
(unless they are regionally different) would allow planners to use the simpler OLS technique.

The results suggest the potential of developing an employment prediction model that accounts for time invariant regional effects. The use of data pooling offers a way to maximize the information return on a limited time-series for a small regional economy by allowing the efficient estimation of a short- to intermediate-term prediction model. Data pooling econometric methods, when used to estimate base multipliers, have merit as practical tools for planners of energy impacted communities.

APPENDIX A

Economic Base and Related Literature of Interest

A number of other studies are worth mentioning in order to gain additional perspective on the empirical difficulties that have been experienced in applying economic base theory to estimate regional multipliers. Also, three studies (one extending economic base theory with central place hypotheses, and two relating residentiary sector employment with per capita income) will be cited.

In one of the earliest attempts to test the export base theory in a temporal design, Sasaki concluded that the effects of 1-, 2-, or 3-year lags were random. In that study, an aggregated base employment multiplier was investigated. The exceptionally high coefficient of determination ($R^2$) obtained was partially tautological since total employment was regressed on total employment lagged 1, 2, and 3 annual periods and on export employment. Prior to his work, the residentiary employment changes had been widely assumed to occur in a single time period (within 1 quarter or 1 year, depending on the study). That assumption was justified by assumed ease of entry into most residentiary activities. However, in Sasaki's work, the insignificant t statistics on the lagged endogenous variable lent little credence to the time adjustment concept. No mention was made of multicollinearity, although specification error could have contributed to the low t statistics. The poor statistical results can be at least partially explained by degrees of freedom limitations (9 annual observations; 2 exogenous variables; 1-, 2-, and 3-year lags).

In research published in 1970, Weiss and Gooding experimented with a precalculated half-year lag using annual data. They regressed residentiary employment on a partially disaggregated base of the Portsmouth, N.H., economy. While that lag scheme worked (others were tried and proved less satisfactory statistically), Weiss and Gooding concluded that their "a priori expectations (disaggregation) were only partially confirmed by significant statistical results". Again, the lack of adequate degrees of freedom (11 years, 3 exogenous variables, artificial half-year lag) provided empirical problems.

In another application of differential employment multipliers, but using a distributed lag structure, Luttrell and Grey expanded on Czamanski's similar, but earlier, efforts. The latter conducted research on Baltimore's economy.
in a simple simultaneous equation system of regional growth. The 1-year annual lag in Czamanski's partially disaggregated model was significant.

Luttrell and Grey estimated individual employment growth equations for nine midwestern cities using quarterly change data for 9 years in an Almon lag model. In encountering statistical problems (inconclusive results) with two out of four postulated employment growth relationships, they concluded that emphasis on the export base view (demand side) is inadequate. In their view, excluding other factors such as quantity and quality of labor and management skills leaves much of the growth process unexplained. Further, their results suggested that increases in employment in some sectors resulted in less employment in others, the latter being priced out of a highly competitive labor market. However, the use of quarterly change data postulates a very quick equilibrating response from the residiary sectors which imposes a large sensitivity burden in the employment data.

The only work to be cited with regard to economic space will be Bender's recent contribution. In a cross-section study of nonmetropolitan counties in the West, Bender found a statistically significant relationship between residiary employment and distance to major trade centers. His approach worked best in the Plains States where application of central place theory is least impeded by geographic barriers.

Fuchs has demonstrated that residiary sector employment increases as a percentage of total employment as per capita income increase. Much of the work examining the residiary sector expansion phenomenon has concentrated on per capita income. This hypothesized relationship has been successfully tested using both cross-section and time-series data for the United States according to Kuznets.

APPENDIX B

Additional Econometric and Statistical Considerations

Additional econometric and statistical topics pertinent to this research use of pooled data for multiple regression analysis are included in this appendix.

Multicollinearity

As the correlation matrix in appendix table 1 demonstrates, multicollinearity generally does not appear to be a problem in using regression techniques with this data set. Only in the case of the GOV and D variables is there a suspect correlation (0.9405). Investigation of the variance-covariance matrix of the error components transformed data demonstrates large covariances between D and the other variables.

Additional Cross-Section, Time-Series Considerations

In addition to the regional effects and autocorrelation effects discussed in the text, the following econometric estimating problems may also exist in
Appendix Table 1—Correlation coefficients for analysis of pooled regression results, NGP major coal county employment estimation, 1970-74

<table>
<thead>
<tr>
<th>Variable 1/</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000</td>
<td>0.704</td>
<td>0.889</td>
<td>0.145</td>
<td>0.132</td>
<td>0.598</td>
<td>0.437</td>
<td>0.045</td>
<td>0.908</td>
<td>-0.942</td>
<td>-0.465</td>
<td>0.125</td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
<td>0.745</td>
<td>-0.446</td>
<td>-0.324</td>
<td>-0.412</td>
<td>0.932</td>
<td>0.151</td>
<td>0.799</td>
<td>0.748</td>
<td>-0.291</td>
<td>-0.202</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.000</td>
<td>0.002</td>
<td>0.017</td>
<td>0.055</td>
<td>0.356</td>
<td>0.257</td>
<td>0.758</td>
<td>0.819</td>
<td>-0.307</td>
<td>-0.080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.000</td>
<td>0.752</td>
<td>0.668</td>
<td>0.517</td>
<td>-0.153</td>
<td>-0.152</td>
<td>-0.066</td>
<td>-0.101</td>
<td>-0.356</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.000</td>
<td>0.348</td>
<td>0.803</td>
<td>-0.114</td>
<td>-0.096</td>
<td>0.009</td>
<td>0.260</td>
<td>0.475</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.000</td>
<td>0.193</td>
<td>-0.168</td>
<td>-0.143</td>
<td>-0.180</td>
<td>0.146</td>
<td>0.276</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.000</td>
<td>0.148</td>
<td>0.252</td>
<td>0.368</td>
<td>-0.413</td>
<td>0.353</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.000</td>
<td>-0.112</td>
<td>0.048</td>
<td>-0.289</td>
<td>0.224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.000</td>
<td>0.940</td>
<td>-0.370</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.000</td>
<td>-0.391</td>
<td>0.059</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1.000</td>
<td>0.109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1/ Variables
1. Residential employment.
2. Agriculture, forestry, and fisheries employment.
4. Mining employment lagged 1 year.
5. Construction employment.
6. Transportation, communication, and public utilities employment.
7. Wholesale and retail trade employment.
8. Service employment.
10. Adjacent county employment as a percentage of county i employment.
11. Distance to trade center.
12. Per capita income.
pooled data: (1) time-series influences, (2) combined cross-section and
time-series influences, and (3) heteroscedasticity.

The first two facets refer to those arguments in the regression func-
tional form which have not been explicitly included as independent variables;
hence, they are aggregated in a random disturbance term. The pooling of time-
series and cross-section data creates a complex disturbance term in an empiri-
cal model, thereby adding a new dimension of difficulty in specification. As
Kuh and Meyer have pointed out, cross-section and time-series estimates do not
necessarily measure the same things. 39/

Early covariance computer runs with T-1 time dummy variables indicated
there were no time-series effects (i.e., business cycle shifts in employment
generation behavior) measured in this data set. Since there were no time-
series effects, there was no requirement to pursue the combined cross-section
and time-series effects.

The third facet refers to a violation of an assumption, homoscedasticity,
of the general linear model. This assumption expects independently distributed
errors from a normal population with an expected value of zero and a constant
variance. Problems with nonconstant variance of the error term are referred to
as heteroscedasticity. Johnston states that the constant variance assumption
is unrealistic when dealing with cross-section data. 40/ Cross-section data
that has large differences in the size of individual observations may exhibit
heteroscedasticity. In the empirical problem here, large employment-level
counties may have secondary employment levels which may be more volatile than
smaller counties. However, there is little empirical evidence on the type of
heteroscedasticity to be found in economic relationships.

For regression work, while the individual parameter estimates will be
unbiased, the estimated variances of those parameter estimates will be biased;
hence, the hypothesis testing on the parameters will be incorrect. A GLS pro-
cedure using weighted least squares can be used to attempt a correction.
First, however, the heteroscedasticity must be detected.

There are a number of tests for the detection of nonconstant error
variance. The one applied to the data in this study originated with Glejser
and consists of regressing the absolute value of the residuals on a simple
nonlinear function of each individual independent variable. 41/ The t statistic
is then used to test the homoscedasticity assumption.

The application of this procedure to the BEA data rejected the null
hypothesis of constant variance in the error term, thus strengthening the
suspicion of heteroscedasticity. However, the uncertainty in identifying and
specifying the exact nature of the heteroscedasticity, plus software difficul-
ties in handling a combined regional effect, autocorrelation, and heterosce-
dasticity econometric problem, precluded further work at this time.

Predictive Power Measures

Single equation regression models usually use the $R^2$ statistic as an
indicator of predictive power. Since it is not always a reliable indicator,
especially when a transformed dependent variable is involved (as in the error components model), other measures of predictive power are required. The method used here is to divide the standard error of the estimate by the mean of the dependent variable. This calculation, called the percentage standard error, is expected to be in a range of 10 to 15 percent for good prediction.

The Covariance Model

One of the earliest econometric methods used to estimate relationships from pooled data was the covariance approach. Since the approach is nothing more than ordinary least squares with N-1 dummy variables for the regional effects case, the model can easily be implemented. The approach allows for a varying intercept term to adjust for missing information. Symbolically, for the time invariant region varying export base model:

\[ R_{it} = a + bX_{itk} + c_{Wit} + \ldots + c_{N Nt} + e_{it} \]

where \( R_{it} \) = residiendiary employment

\( X_{itk} \) = exogenous variables such as basic employment, distance measures, or per capita income.

\( i = 1, \ldots N \) counties
\( t = 1, \ldots T \) time periods
\( k = 1, \ldots K \) sectors

\( W_{it} = 1 \) for the ith county
\( 0 \) otherwise \( i = 2, \ldots N \)
\( t = 1, \ldots T \)

\( a \) = constant
\( b, c \) = parameter estimates
\( e_{it} \) = residual term.

In order to ensure that an increase in efficiency occurs with the use of the covariance approach versus ordinary least squares, an F test is conducted on the model specification used in the text. Using the statistic:

\[ F = \frac{(RSS_1 - RSS_2)/(N+T-1)}{(RSS_2)/(NT-N-T)} \]

where \( RSS_1 \) = residual sum of squares from ordinary least squares
\( RSS_2 \) = residual sum of squares from covariance approach,

a comparison can be made to test the null hypothesis that the equal regional intercept restriction on the OLS model is correct. Using a 15-county, 5-year sample, the resulting F test as calculated is 38.44, highly significant at
the F.01 level. This refutes the null hypothesis and demonstrates the advantage of the covariance model by increasing the efficiency of the estimators.

Additional Error Components Notes

The error component theoretic design evolved out of a model developed earlier by statisticians. Developed from analysis of variance work, the random effects model was applied to an econometric use. The resulting estimator was found to have the properties of least bias, least mean square error, and the greatest robustness against specification error of the methods which were compared.  

The random effects model essentially divides the error term into two or more components, thus identifying hidden effects which had been heretofore locked in the residual. Applied to the estimation problem in question here, the regional effects in the disturbance term would be a random disturbance varying about the equation’s intercept, while the covariance approach would treat that component of the error term as a constant incorporated within a separate intercept representing each region.

Since error component variances are generally not known, a two-step estimation process is necessary. First, a covariance regression is run on the pooled data. The residuals of this dummy variable model are then used to calculate sample estimates of the variance components. Generalized least squares are then run on data transformed by the estimated variances to obtain the final parameter estimates.

Remembering that the error term’s composition is the following:

\[ e_{it} = u_i + v_{it} \]

for the region varying time invariant model, then a proportion of the error variance accounted for by the regional component can be calculated as follows:

\[ \rho = \frac{u_i^2}{e_{it}^2} \]

In the computer program which calculated the error components estimates, a proxy \( \hat{\rho} \) is estimated. This calculation of \( \hat{\rho} \) proceeds by taking the variance of the cross-section specific effects from the covariance regression as a proportion of the total variance. The latter is, of course, the sum of the variance of the cross-section effects, and the sum of squared residuals from the covariance regression. Symbolically:

\[ \hat{\rho} = \frac{\sum u^2}{\sum u^2 + S^2/NT} \]
where \[ \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} \left( \bar{S}_{i} - \bar{S} \right)^2 \]

where \( N \) = number of regions (counties)

\( T \) = number of years

\( \bar{S}_{i} \) = mean for the \( i \)th region of \( S_{it} \)

\( \bar{S} \) = grand mean over all regions of \( S_{it} \)

\( \bar{X}_{ki} \) = mean for region \( i \) of \( X_{kit} \)

\( \bar{X} \) = grand mean of \( X_{kit} \)

The data is then transformed:

\[ S_{it}^* = \frac{S_{it} - \bar{S}_i}{\sqrt{\eta}} + \frac{\bar{S}_i}{\sqrt{\xi}} \]

\[ X_{kit}^* = \frac{X_{kit} - \bar{X}_{ki}}{\sqrt{\eta}} + \frac{\bar{X}_{ki}}{\sqrt{\xi}} \]

where \( \xi = (1-\rho) \) and \( \eta = (1-\rho) \)

The data are then used in an OLS regression to obtain the generalized least squares results. These results can, in general, be found to provide better estimators to the covariance approach unless \( \rho \) approaches 1. This condition, as shown by the Monte Carlo studies from Maddala, among others, results in the error components method degenerating into the covariance model. Nonetheless, since a priori knowledge of \( \rho \) is unavailable and we are interested in improving our estimates and prediction capability, use of the error components procedure is warranted. Then modeling decisions can be made as to best statistical fit, best economic interpretation, best prediction, etc.

The \( \hat{\rho} \) estimate for the main result here is .9977, which obviously is approaching 1. While the error components approach is chosen as the best result mainly on the strength of a slightly better percentage standard error, the covariance approaches results are very similar in terms of coefficient value and statistical significance. For this particular model specification, the error components approach offers only a slight improvement over the covariance approach.
Measuring Autocorrelation in a Pooled-Data Set

Since the pooled-data set is not composed of a continuous time-series but a set of discrete, finite time series, a modified Durbin-Watson d statistic is utilized to measure any first order autoregressive scheme present. Symbolically, the d statistic is:

\[
d = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} (e_{it} - e_{it-1})^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} e_{it}^2}
\]

The statistic investigates autocorrelation only within a cross-section unit's time series and averages over all cross-section units. The test statistic is a proxy since it does have large sample properties, yet only 75 observations were used in the final run.

2/ This simple explanation of regional economic expansion should not be confused with the similar but more complex input-output expansion effects. For a good discussion of backward and forward linkages with intermediate goods and services production, see Eliahu Romanoff, "The Economic Base Model: A Very Special Case of Input-Output Analysis," The Journal of Regional Science, 14, 1:121-130, Apr. 1974. Similarly, this simple theoretical expansion should not be confused with one detailing many specific regions (e.g., an inter-regional multiplier).


4/ John W. Alexander, "The Basic Nonbasic Concept of Urban Economic Functions," Economic Geography, v. 30, July 1954, asserts that the division of a region's economic activities into export and local categories illustrates a space relationship. Walter Isard, Methods of Regional Analysis: An Introduction to Regional Science, Cambridge: M.I.T. Press, 1971, seventh printing, pp. 199-205, surmised the importance of an extension, but failed to develop the concept beyond a general theory. Appendix A briefly discusses a more recent study which attempted this economic base theory extension.


7/ Edwin Kuh, "The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Applications," Econometrics, 27, 2:197-214, Mar. 1959. Kuh states that "...observations on cross-section bias lend support to the rather obvious but pertinent proposition that the biases from excluded variables can be strikingly different in time series and cross sections. Therefore, the propriety for prediction purposes of applying in one context behavior relations estimated in another context is highly questionable. Observations on individuals at only one or a few points of time will often be structurally incomplete because the observation in a given cross-section is likely to be affected by prior observations." A Chow test on early OLS runs of a 21-NGP county sample of untransformed data suggested a structural change occurs late in the time series used here.


10/ The location quotient technique assumes for this study that if a given county is highly specialized relative to the Nation in production of a particular commodity/service, the product is presumed to be an export item. That is:

\[ e_{ikt} = E_{ikt} - \left( \frac{U_{ikt}}{U_t} \right) E_{it} \]

where
- \( e_{ikt} \) = estimated county i basic employment in industry k in year t
- \( E_{ikt} \) = total county i employment in industry k in year t
- \( U_{ikt} \) = U.S. employment in industry k in year t
- \( U_t \) = U.S. total employment in year t
- \( E_{it} \) = total county i employment in year t

and \( \left( \frac{U_{ikt}}{U_t} \right) E_{it} \) < \( E_{ikt} \) for \( e_{ikt} \) to be basic

if \( \left( \frac{U_{ikt}}{U_t} \right) E_{it} \) > \( E_{ikt} \) then all employment in industry k in county i is local or residiendiary oriented.

11/ U.S. Department of Commerce, Bureau of Economic Analysis; *Annual County Employment*, 1970-74. Employment data are fraught with their own pitfalls (such as changing labor/output ratios), but are commonly used due to their availability. Note that the 5-year period of the model should constrain this problem. In Tiebout's excellent discussion of appropriate data, the point is made that, "Employment...as a unit of measurement is not too sensitive to shortrun changes." See Charles M. Tiebout, *The Community Economic Base Study*, Supplementary Paper Number 16, N.Y.: The Committee for Economic Development, 1962, p. 67.

12/ Major coal-mining activity is defined here to mean mines producing 50,000 tons per year or more according to U.S. Department of Interior, Bureau of Mines, L. Westerstrom, *Coal-Bituminous and Lignite*, reprint from 1974 Minerals Yearbook, U.S. Government Printing Office, 1975. Data from the following counties were included: Big Horn, Richland, and Rosebud, Mont.; Bowman, Burke, Mercer, Oliver, Stark, and Ward, N. Dak.; Campbell, Carbon, Converse, Lincoln, Sheridan, and Sweetwater, Wyo.


14/ This is not to say that multicollinearity or other model misspecification will be ignored. See appendix B for further discussion of these econometric considerations.
15/ T. Paul Schultz, "An Econometric Model of Family Planning and Fertility," *Journal of Political Economy*, 77, p. 165, 1969, suggests another interpretation of regional effects as errors of measurement that differ from region to region. These errors might stem from administrative and institutional factors such as arbitrary political boundaries for a regional economy.


17/ See appendix B for additional detail on the covariance approach.


20/ As the empirical results will demonstrate, the attempt to expand the core economic base model was not successful.

21/ See appendix B for additional detail.

22/ See appendix B for an explanation of the prediction error measure.

23/ The Durbin-Watson statistic upper limit for K=8 variables and n=60 observations is d ≈ 1.80. Hence, using the formula 2 < 2.16 < (4-1.80), the null hypothesis of zero autocorrelation for the covariance model is accepted. See appendix B for details on the modified Durbin-Watson d statistic used here.

24/ The t statistic used was:

\[ t \approx \frac{B_j - B_k}{S_{B_j} - S_{B_k}^{1/2}} \]


25/ Agriculture was not tested for lagged effects due to the quality of the BEA data constructed by allocating State totals to the county level.

26/ Compare the unlagged regression in table 4 with mining sector lagged covariance result in table 2.

27/ See Pindyck and Rubenfeld, *op. cit.*, pp. 22-23, for a discussion of minimizing mean square error. Unfortunately, the computer software used did not produce a measure of bias.
While the simple correlation matrix indicates very low correlation of construction and the adjacent county employment variables, the Farrar-Glauber test for multicollinearity (comparing F tests of each independent variable regressed on the K-1 remaining independent variables) demonstrates that both government and the adjacent county variable are the largest potential sources of multicollinearity. Inspection of the variance-covariance matrix of the error components results shows a strong, absolute covariance between the adjacent county employment variable and construction. See J. Johnston, op. cit., pp. 159-64 for additional discussion.

Another version of economic space using distance interaction terms was tested early in this research. Distance from the center of a county to a Rand McNally trading center was multiplied times the level of employment in each sector. The covariance results were unimpressive; hence, the approach was not pursued. Generally speaking, the pooling approaches taken here do not appear to be the best statistical methodology for incorporating economic space into economic base theory.

A comparison of simple ordinary least squares (no dummy variables) with and without per capita income for this data set reveals that the result without the income variable has a slightly lower standard error of the estimate, with all other coefficients virtually the same and all t statistics the same. The demand differential hypothesis is not reflected here.

The usual 100(1-.05) percent confidence interval was used here:

$$B_{i1} \pm t_{.05/2} \sqrt{\frac{\sum e_{it}^2}{nt-k}} \sqrt{(aii)}$$

See J. Johnston, op. cit., p. 135.


42/ Note that time effects can also be estimated by T-1 dummy variables.


45/ Henry Huang of the University of Minnesota made the error components program available to the author.


© U. S. GOVERNMENT PRINTING OFFICE : 1978 261-496/77