Employment Projections Within A Shift-Share Framework

EDWARD M. MILLER*

There is a need for methods for forecasting employment (and hence population) which can be used with computer aid, to make projections for a large number of relatively small areas. The simplest models merely make extrapolations from the historical growth trends, but these cannot incorporate information about the changing nature of final demand or changes in the rate of technical progress in different industries. Regional input-output provides a very sophisticated method for handling the flow of materials between industries and regions but the lack of good empirically derived coefficients makes it very difficult to apply even for regions the size of states and renders it impractical when dealing with smaller areas.

A shift-share approach would seem to be a suitable intermediate level of sophistication. It has a number of the advantages of the input-output approach. The model is propelled by projections of national employment growth in a number of different sectors. This makes it possible to incorporate and experiment with differing compositions of final demand. Technical progress can be incorporated in making the national employment projections.

For the reader not familiar with the shift-share technique a brief description will be given in the next few paragraphs, which some readers may wish to skip. A detailed description will not be given since that has been done elsewhere.[1] After describing the technique and source of data, the paper will examine the forecasting accuracy of the methodology, and discuss the negative results obtained by another researcher.

The shift-share technique starts with a disaggregation of the economy into sectors. The growth in employment within each sector is then divided into three components: (1) The national growth component is the local growth that would have been experienced if local employment had grown at the same percentage rate as total national employment. (2) The industrial mix component is the local growth that would have been experienced if the local sector had experienced the same growth as the sector nationwide minus the national growth component. (3) The regional share component is the algebraic difference between the observed local growth in employment and the sum of the national growth components

and the industrial mix components. The separate components for each sector may be summed to give an overall national growth component, an industrial mix component, and a regional share component.

An example may be useful. Consider a county with an employment of 1000 in a particular sector. The national growth rate for all employment is 5%, and 10% in that sector. Within the county the sector's employment grows by 15% or 150 people. If the county's employment had grown at the national rate of 5% for all sectors it would have increased by 50. This would be the national growth component. If the county's employment in the sector had grown by the 10% that national employment in the sector did, it would have increased by 100. The difference between this and the national growth component of 50 is the industrial mix component for that sector (50). The part of the county's growth that remains to be explained (150-100 or +50) is the regional share component.

By making separate projections for each local sector, account is taken of the differences in industrial composition between areas. The method gives separate projections for each sector, a feature that is essential where growth in different sectors produces different effects. Examples would be the prediction of the production of pollutants or transportation demand.

The model suffers in incorporating factors other than industrial composition only through historical projections of the regional share component. However, this ability to incorporate historical trends (and indirectly the large number of variables that have influenced them) is an advantage over current input-output techniques which lack a historical series of coefficients from which to make projections. However, there is no reason why the effects of various variables (accessability, wage rates, climate etc.) on the regional share component could not be estimated using standard econometric techniques, and these econometric functions used to project the regional share component.

The industrial mix component provides a measure of how rapidly the industries in a community are growing nationwide. A positive industrial mix component indicates that the county's industrial structure includes more fast growing than slow growing industries. A negative coefficient indicates that slow growing industries (nationwide) predominate. The regional share component is the residual left after controlling for the county's industrial mix. An overall positive regional share component indicates that within most industries the county is increasing its share of national empolyment. If the county is specialized in declining industries (indicated by a negative industrial mix component), a positive regional share component could still be consistent with the county having a below average growth of employment. The regional share component is often considered a measure of the general competitiveness of an area.

The shift-share approach has an obvious utility for describing regional growth, and in particular for indicating whether a particular community's growth was due to being specialized in rapidly growing industries and the extent to which it was unusually competitive within the industries in which it was specialized.

Used as a device for making projections, a shift-share approach would seem to be most useful where it was desired to prepare projections for a large number of areas quickly, especially if projections for the different sectors within each industry were required. The advantage of shift-share over simple extrapolation is that information on the pattern of national demand in different sectors could be incorporated. Growth rates for each sector nationally would be predicted and used to calculate the industrial mix components for each local area. The regional share components would be projected from the historical regional share components. This paper will examine the accuracy of this approach and compare the accuracy of the projections with those obtained by a naive extrapolation of historical growth rates.

The calculation of the components of employment growth for all the counties in the United States could be a very burdensome chore. Fortunately, this has already been done by the Bureau of Economic Analysis (BEA), U.S. Department of Commerce (formerly Office of Business Economics), for 3102 counties or county equivalents utilizing employment data from the 1940, 1950, and 1960 censuses for 32 different sectors.[2] The military is included as one sector, as is most government administration. Construction, farming, mining, and forestry-fishing are each a sector. Manufacturing is divided into ten sectors, and transportation into three. Trade and services constitute the remaining sectors. This is the data used here. Since the population census was used to provide the raw data, the employment is attributed to the county where the worker lives rather than the county where the job is actually located. This is good if the goal is to predict population, but not so good if the purpose is to predict industrial location.

A Naive Model

To provide a comparison with the shift-share analysis, a naive model will be used. This predicts future employment growth by projecting forward from the growth observed in the previous decade. It is a simple approach (frequently used by planners of physical facilities) that has the virtue of not requiring multi-sectoral projections of the national economy and of not requiring extensive computations. To provide consistency with the approach used later for projecting the regional share component, the projection formula will be estimated econometrically rather than merely assuming a continuation of the historical rate of growth of employment. Indeed, one of the significant findings will be that merely projecting the historical rate forward is not the best way to utilize the historical data. What theoretical basis the method has rests on a belief that the factors that make for employment growth persist for long periods of time and, although all of these factors may not be known, the historical growth rate provides a surrogate for many of them.

Thus, the percentage change in total employment 1950-1960 (TOTALC56) was regressed on the percent change in total employment 1940-1950 (TOTALC45) to give: (The numbers under the coefficients are standard errors for those coefficients)

(Equation 1)

TOTALC56 = −7.33 + .6185 TOTALC45

$$(.403)\,(.0166) \qquad R^2 = 30.9\%$$

There is no doubt about the usefulness of the previous decade's employment growth in predicting future employment growth. The t statistic[3] for the employment growth during the forties was 37.25. This high level is a result of having 3102 observations. However, the previous decade's growth of employment is not a very good predictor since it is able to explain only 30.9% of the next decade's employment growth. Even this 30.9% may be an overstatement of the accuracy that can be obtained because the coefficients in the equation probably change over time. Thus, there is certainly a need to seek improved methods for forecasting employment.

One of the most significant findings is that the best predictor is not merely to assume that the past rate of employment growth will be observed in the future (which would give TOTALC56 = TOTALC45). The coefficient for the previous decade's employment growth of .6185 is 23 standard deviations away from unity. Simple extrapolation of employment growth rates will overestimate the growth of rapidly growing counties and underestimate it for declining counties.[4] These results show that a county that grew more rapidly in one decade than the national average will also grow more rapidly in the next decade, but that its rate of growth will move closer to the national average. Likewise, counties growing at a below average rate in one decade will grow at a below average rate in the next decade, but their rate of growth will more closely approach the national average.

The reason for this result is that employment for any one time (such as the day of census enumeration) is affected by many transient factors.[5] Among these are strikes, temporary lay-offs, late harvests, etc. If one of these lowered employment for one year the growth of employment for the decade before will be understated, and the growth of employment for the next decade overstated. When this happens a negative correlation between employment in two adjacent decades is produced. In addition there are isolated events which produce fairly long term changes in employment such as the opening or closing of a single plant or military base, but which may not necessarily repeat themselves in successive decades.

Thus when an unusually rapid growth in employment is observed for one decade, it is not known whether this reflects (1) conditions favorable for employment growth which are likely to persist into the next decade, or (2) whether there was merely an isolated event during the decade which lead to a one time increase in employment unlikely to be repeated (such as opening of a military base), or (3) whether employment was abnormally depressed at the beginning of the period (perhaps by a temporary layoff), or (4) the employment at the end of the decade was abnormally high for some reason (such as a large construction project). If it is impossible to tell which is the most likely cause of unusually rapid growth, the best prediction is that the county will continue to grow faster than the nation but that the difference in growth rates will be substantially less in the future (about two-fifths less).

It should be noted that the constant term in equation 1 is significantly (t = −18.18) negative. This indicates that the typical county had an employment growth during the fifties that was 7.33% percentage points less than in the forties. This is due to the decline in the national rate of growth of employment from 26.7% for the forties to 15.5% for the fifties.[6]

Equations in the form of equation #1 were run for six regions of the country.[7] Equations of the form discussed here were able to explain only 19% to 38% of the variance. In all cases the coefficient for historical employment growth was significantly less than unity. Having seen that a simple projection of employment growth is only moderately successful, let us see how much better a shift-share approach can do.

Use of Shift-Share Analysis for Forecasting

In using shift-share analysis for purposes of predicting employment it is necessary to have methods for predicting each of the components of employment change. The national growth and industrial mix components can be obtained by predicting national employment in each of the 32 national industries and then multiplying base year employment by these predicted growth rates. This mechanical procedure is especially suited for use when projections must be prepared quickly for a large number of geographical units. The regional share component is most easily predicted from the values it has historically taken.

The examination of the forecasting accuracy of the shift-share approach will start by regressing the industrial mix and regional share components for 1950-1960 on the same components for 1940-1950. The components have been converted to a percentage by dividing through by the average employment for the period (one half of the initial and final year employment figures). This provides compatability between counties of different sizes and makes the data more homoscedastic. It is of course realized that if both components are actually to be projected by extrapolation from historical data that there is little reason for separating the change in employment into these different components. Essentially the same projections could be obtained more directly simply by extrapolating the changes in total employment in the manner discussed earlier. Thus, the justification for examining the correlation between industrial mix components in different years must be the insight into the processes at work it gives us.

Results for the States

Before discussing the results for the 3,102 counties in the United States, results for the 48 contiguous states and the District of Columbia will be presented. This will be done both because of the intrinsic interest of these results, and because there are some complexities in interpreting the county results that are less important for larger units such as states.[8] The county results will be easier to understand once the state results have been presented.

The results of a regression of the industrial mix component for the fifties on the industrial mix component for the forties (both in percentage form)[9] was:

(INDM45 = Industrial Mix Component 1940-1950
INDM56 = Industrial Mix Component 1950-1960)

(Eq. 2)

INDM56 = −7.643 + .7714 INDM45
                    (.0402)
            Standard Error = 2.494% R^2 = 88.69%

Thus, reasonably good predictions of the industrial mix component can be obtained through simple historical projections (even using a decade that included a World War). As will be seen when the regional share component is discussed, the difficulty in making state employment projections is in predicting the regional share component, not the industrial mix component.

It should be noted that the coefficient for the historical industrial mix component is significantly (5.7 standard deviations) below unity. It appears that the differences between states in the extent to which they benefitted from favorable industrial mixes were less in the fifties than in the forties. The most likely reason is the industrialization of many of the traditionally rural states, which reduced the differences between states in their industrial compositions.

The Regional Share Component for States

The regional share component is the part of the state change in employment that is not due to the industrial mix component. The following equation was obtained when the percentage regional share component for the fifties was regressed on the percentage regional share component for the forties:

(REGS45 = Regional Share Component 1940-1950
REGS56 = Regional Share Component 1950-1960)

(Eq. 3)

REGS56 = −.88402 + 1.032 REGS45
                      (.1141)
            Standard error = 9.321% R^2 = 63.55%

The coefficient for the historical regional share component is highly significant (t = 9.05) and within one standard deviation of unity. Thus, state regional share components can be predicted by assuming that they will retain the same value they previously had. This procedure seems capable of explaining almost two thirds of the variance in regional share components.[10] Still, the standard error over one decade was 9.32 percentage points. Any forecasting error in the industrial mix and the national growth components must be added to this. Thus, errors in forecasting over one decade will be at least 10%.[11] This is enough so that investment planning should always consider the possibility that the forecasts of demand will be wrong, and should hence endeavor to retain flexibility.

Results for Counties

The method of analysis applied to the county data is that used above for the states. When the industrial mix component for 1950-1960 was regressed on the industrial mix component for 1940-1950, the result was:

(Eq. 4)

$$INDM56 = -3.620 - .7448 \ INDM45$$
$$(.175) \ (.00889)$$
$$\text{Standard Error} = 6.759\% \qquad R^2 = 69.39\%$$

When the regional share component for 1950-1960 was regressed on the regional share component for 1940-1950 the result was:

(Eq. 5)

$$REGS56 = 1.156 + .3151 \ REGS45$$
$$(.327) \ (.0169)$$
$$\text{Standard Error} = 18.08\% \qquad R^2 = 10.03\%$$

Perhaps the most striking characteristic of these equations is the contrast between the high explanatory power of the industrial mix equation ($R^2 = 69.39\%$) and the much lower explanatory power ($R^2 = 10.03\%$) of the regional share equation. This contrast indicates that the correlation between employment changes in successive decades shown earlier is primarily due to the high correlation between the industrial mix components in successive decades. In turn, this high correlation is due to the slowness of changes in the industrial composition of counties and the tendency for rapidly growing industries in one decade to be rapidly growing industries in the next decade. For instance, agriculture showed large declines in employment in both the forties and the fifties.

If the shift-share projection procedure outlined above is to be superior to a simple historical projection, the advantage must be in more accurate projections of the industrial mix component since both the shift-share method and simple historical projection of total employment project the regional share component forward. Unfortunately, the scope for obtaining significantly better projections of total employment through improving on simple historical projections of the industrial mix component is limited. The standard error of the above industrial mix equation is only 6.759% percentage points. Even if this source of error was completely eliminated there would still remain the 18.08% standard error in the regional share component. Thus the simple shift-share predictive model given above can (with perfect predictions of growth of national employment in 32 separate industries) achieve a standard error that is fractionally superior to the 18.81% standard error obtained earlier by historical projection of total employment; a much simpler procedure.[12]

One reason for the small improvement from knowing the industrial mix component is that there is an inverse correlation between the industrial mix component and the regional share component; in particular, there is an inverse correlation between changes in the two components of employment growth.[13] As a

result, errors in the industrial mix component are offset by errors in the regional share component and the corresponding effect on total employment is small.

The reader will notice that the coefficient for the industrial mix component in Equation 4 is slightly less than .75. This indicates a tendency for the differences in industrial mix components to converge. This tendency has been noted by other researchers.[14] It is apparently due to an increasing similarity in industrial structures for the counties of the United States. In particular, the industrialization of areas formerly dependent on the extractive industries is reducing the difference between these areas and the remainder of the United States. This reduced difference in industrial mix components is presumably one reason for the tendency for employment growth rates to converge.

The Regional Share Component

Although the percent of the variance in the regional share component explained by the previous decade's regional share component is low (10.03%), the correlation between the regional share components in the two decades shows a very high level of statistical significance (t = 18.59) due to the large number of observations. There can be no doubt of the existence of a positive correlation between the regional components for successive decades.

The coefficient for the historically observed regional share is only .3151, which is even further from unity than it is from zero. Part of the difference from unity may be due to a real tendency for the differences in the "competitiveness" of counties to diminish over time. However, the primary reason for the low value for the coefficient is the same one noted earlier in projecting total employment. A positive or negative regional share effect may be produced by many single events which affect employment in only one year such as temporary layoffs, late harvests, construction projects, or the opening or closing of a single plant. Many of these short term fluctuations in employment will be offset by employment changes in the opposite direction during the next decade, introducing a negative correlation between employment changes in successive decades. Since the regional share component is the residual left after subtraction of the national growth and industrial mix components, it includes the effects of all these short run disturbances. This interpretation is supported by the large coefficient of 1.0 found for state regional shares (Eq. 3) and about 1.4 for populous counties (Eq. 10). A state is a much larger geographic unit and much more averaging occurs within its borders.

The low value for the percentage equation can also be explained by the inclusion of a large random element in the regional share component with the result that the best fitting regression equation is a weighted average of a zero coefficient for the random element and some positive coefficient for the part that does reflect the "competitiveness" of a community. We can view the regional share components as being a function of the "competitiveness" of the county, plus an error term. The regional share component for the previous decade is composed of this "competitiveness" plus certain randomly distributed transient terms.

In essence there are measurement errors in the independent variable, a condition that is known to bias estimates of coefficients downwards.[15]

Comparison with Other Studies

The above conclusion that historical information on the regional share component can be used for predicting future regional share components appears to be in conflict with a study of shift-share analysis by H. James Brown.[16] He examined the accuracy of predictions of the regional share component (which he refers to as the "competitive" component) from the previous period's regional share component for 17 Standard Metropolitan Statistical Areas (SMSA's) using data from the Census of Manufactures for 1947, 1954, 1958, and 1963. His conclusions were: "The data gives no doubt that for predicting the percentage growth in manufacturing employment the simple extrapolation of every regional industry at the historical national industry growth rate gives better results than the best formulation of the S and S projection model" and "The evidence indicates that the competitive component may appropriately be treated as a random variable." The first statement is not in direct conflict with the conclusion here in that the standard error of the historical equation for predicting the regional share component is not appreciably less than the standard error of a similar equation for predicting total employment.

Still, the statistical significance of the coefficient (t = 18.59) obtained for the historical value of the regional share variable seems to go against the tenor of his conclusions. There are many differences in the two studies that could explain the differences in conclusions such as Brown's use of the Census of Manufacturing instead of the Census of Population, the measurement of employment on a place of work basis, the examination of SMSA's instead of counties, and the choice of years.

However, the primary difference between the studies is in the range of projection methods considered. The only utilization of the historical regional share component Brown examines is "that the competitive component in the current period equals that from the previous period." This is compared with a model that "projects each regional industry at that industry's national growth rate over the previous period (Ingrow) and a model that projects each regional industry at the national industry's projected growth (Super Ingrow)." The reader will recall that the model used in the current study takes form REGS = C + B × REGS (previous period) with the coefficients for C and B estimated empirically. In contrast to this fairly general model, Brown examined only a model where C was set at 0, and B was constrained to be either 0 or 1. Considering that the best value for B was found to be .3151 for counties,[17] it is not surprising that Brown's predictions made with B set equal to 0 proved superior to those made with B set equal to 1. However, this showing does not support the conclusion that historical values for the regional share component are useless in making forecasts. As shown above they are useful. It is very likely that if Brown had examined some procedure that sets B at a number between 0 and 1, this would have proved superior to the simple models he examined.

Robert Kirt, using the same BEA data used in this paper, but a different methodology, has shown that a shift-share approach gives accurate predictions of employment for regions in Iowa.[18] The reader should also note that the ability to predict the regional share component for large urban counties (Eq. 10) and states is much better than for all counties.

The Importance of County Population

Changing the weights given to counties of different population will only make a large difference if other characteristics are correlated with county size.

Thus the industrial mix and regional share components for the fifties were regressed on the logarithm of the 1950 employment (a measure of the size of the county) giving:

(Eq. 6)

INDM56 = 69.30% + .0628 LOGEMP
                     (.00149)
              Standard Error = 9.74% R^2 = 36.44%

and

(Eq. 7)

REGS = − 7.12110% + .00764 LOGEMP
                        (.00292)
              Standard Error = 19.04% R^2 = 0.22%

There is a strong contrast between the high percent of the variance in the industrial mix component (36.44%) explained by county population, and the very low percentage of the variance in the regional share component explained by population. It is clear that the explanation for the more rapid growth of the populous counties is to be found in a higher proportion of fast growing industries, rather than to gains within particular industries. In turn, the industrial mix advantage of the populous urban counties is due largely to having little employment in the extractive industries (agriculture, mining, and forestry and fishing). Thus, the very high t value of 42.114 for the logarithm of employment is not surprising.

Although the positive coefficient for the regional share component is significant at the 95% level, it is not significant at the 99% level (here t = 2.620). The explanation for even this level of significance for a factor that explains only 0.22% of the variance is that there are 3102 observations. For all practical purposes, employment within particular industries grew at the same rate in the small rural counties as in the more populous urban counties.

The showing of the small, rural counties is actually even better than indicated above. There was a strong positive correlation between the regional share and the industrial mix components during the fifties.[19] This apparently arose because the presence of growing export industries led to above average growth in the local service industries. Since the more populous counties tended to have the larger industrial mix components, they would also have been expected to have

large regional share components. The fact that they did not suggests that within any one industry there were certain diseconomies of agglomeration. This finding is contrary to the current "growth center" argument that a certain minimum size is needed if a community is to increase its share of national employment within particular industries (i.e. attract new industry rather than merely maintain its relative position).

Since there are a number of significant differences between large and small counties, changes in the weighting function to reflect the greater importance of certain counties might prove profitable. Thus an alternative set of regressions were run with the dependent and independent variables expressed as the numerical change in employment. With ordinary least squares estimation this emphasizes the large counties with their numerically large changes in employment. The changes in employment for the small rural counties are numerically small and hence the points representing these counties are near the origin. The changes in employment for the large urban counties will usually be numerically large and lie further from the origin. When a least squares regression line is fitted its slope will be little influenced by the points near the origin. Such a regression is thus an indicator of the behavior of the large urban counties.[20]

The Naive Model Revisted

When the numerical change in employment 1950-1960 is regressed on the numerical change in employment 1940 to 1950 for the 3102 counties in the United States, the following equation was obtained:

(Eq. 8)

$$TOTALC5060 = -491.6 + .8614 \; TOTC4050$$
$$(220.0) \quad (.0125)$$
$$R^2 = 60.56$$

The negative value for the constant (significant at the 95% level) is probably due to the deceleration in the rate of employment growth. In particular, there was only a 15.5% increase in national employment in the fifties, versus 26.7% during the forties.

It may be noticed that Equation 8 explains 60.56% of the variance while the equation using percentages was only half as successful ($R^2 = 30.91\%$). The primary reason for the low explanatory power of the percentage equations is that there are many events such as temporary layoffs or the opening or closing of new plants which raise or lower growth for a particular decade, but which are not indicative of persistant structural factors making for growth or decline. For the typical county, with only a small employment such one-time events account for a significant fraction of the variance in employment. For larger counties such one-time events average out and the observed changes in employment are more likely to be indicative of persistent underlying structural causes.

This is probably the primary explanation for the employment change coefficient in the previous decade being .8614 when the absolute values of employment

change are used, but only .6185 when percentage changes in employment are used. Since the absolute growth in employment during the fifties was 8,897,700, versus 12,099,100 during the forties, the coefficient for historical employment growth would be .735, if all counties grew at the same rate. The fact that the coefficient is greater than .735 indicates that the forces making for differences in the rates of growth in the large urban counties were at least as strong in the fifties as in the forties. This is in contrast to the earlier conclusion (when all counties were weighted equally) that differences in the rate of growth may be diminishing. This was explained by increased industrialization reducing the differences between the industrial mixes of the rural and urban areas. Such an explanation is not necessarily inconsistent with the observation that for the large urban counties there has been no convergence in the rate of growth.

The Industrial Mix Component in Numerical Form

When the industrial mix component for 1950-1960 was regressed on the industrial mix component for 1940-1950 the following equation was obtained:

(Eq. 9)

$$INDM5060 = .4302 + .8225 \ INDM4050$$
$$\qquad\qquad (47.3306) \quad (.0677)$$
$$\qquad\qquad \text{Standard Error} = 2636.1 \qquad\qquad R^2 = 82.63\%$$

Although the extremely impressive t statistic (121.5) overstates the significance of the industrial mix component for the previous decade, there can be little doubt about the high predictive power of the previous decade's industrial mix component. Since the constant term is essentially zero, the industrial mix components for the fifties average 82% of the industrial mix components for the forties.

The Regional Share Component (Numerical)

When the regional share component for 1950-1960 was regressed on the regional share component for 1940-1950 (both expressed in numerical form) the resulting equation was:

(Eq. 10)

$$REGS5060 = .03379 + 1.3889 \ REGS4050$$
$$\qquad\qquad (141.939) \qquad (.0130)$$
$$\qquad\qquad\quad \text{Standard Error} = 7905 \qquad\qquad R^2 = 78.57\%$$

While the impressive t statistic (106.6) overstates the significance of the regional share component from the previous decade, there is little doubt that the historical regional share component is useful in predicting future regional share components. The above equation differs significantly from the equation for regional shares expressed in percentage form which was:

(Eq. 11)

$$\text{REGS56} = 1.156\% + .31051 \text{ REGS45}$$
$$(.327\%)\ (.01695)$$
$$\text{Standard Error} = 18.077\% \qquad R^2 = 10.03\%$$

The most significant difference is in the overall explanatory power of the two equations. The equation for predicting the numerical change in employment explains 78.57% of the variance, while the equation for predicting the percent regional share component explains only 10.03% of the regional share component. Apparently, historical values for regional share components are much better guides to the future for the large, urban counties than for the small, rural counties.

Earlier, when discussing the percentage regional share equations it was hypothesized that their low explanatory power was due to the many random events (such as the opening or closing of single plants) that affect a decade's rate of growth but are unlikely to be repeated. For the large counties, most such single events cancel out and the regional share component reflects only factors that affect a number of establishments at one time. These are much more likely to be conditions (such as lack of space for expansion, high labor costs, poor transportation, bad government etc.) that persist from decade to decade. Thus, the regional share component is a much better indicator of the "competitiveness" of a large community than of a small community.

The other major difference between the two equations is in the magnitude of the coefficients for the regional share effects. These were .315 for the percentage equation and 1.389 for the numerical equation. While the errors in variables bias probably tends to lower the coefficient for the numerical equation, it has much less of an effect than for the percentage equation. However, the evidence of any downward bias does serve to make the high coefficient for the numerical equation more puzzling.

Part of the explanation for the high coefficient is the 26.7% growth in employment between 1940 and 1950. If the regional share effect remained a constant percentage of employment for all counties we would expect the coefficient to be 1.267. The fact that the coefficient is greater than even this indicates that not only does the regional share component measure the competitiveness of a county, but that whatever is measured by the regional share component had a stronger impact on employment in the fifties than in the forties.

One possible factor would be the trend toward the suburbs. The regional share effect in the forties appears to be often negative for counties containing central cities, but positive for the suburban counties. With the increased movement toward the suburbs during the fifties the relative decline in the central city counties has accelerated, while the growth in the suburban counties increased. Because the signs of the components remain unchanged while the magnitudes increase, the correlation between the decades is high and the coefficient exceeds unity.

Other Determinents of Regional Share Components

The primary purpose of this paper is to show how shift-share data might be used for projections, with the regional share component projected using only historical data. However, it would be desirable to use additional information to project the regional share component. In addition, shift-share data can aid in analyzing the causes of growth. The industrial mix component represents the direct influence of having industries that are fast or slow growing on the national scale. Thus, the aggregate regional share component is the part of total employment change that is not due to the presence of fast growing or slow growing industries and the part of employment change that may have been influenced by other factors. Since many variables of interest are likely to be correlated with the industrial mix, it is important to exclude the effect of industrial mix before attempting to estimate the effects of other factors.

One common methodology in the study of industrial location is to take a particular industry and try to determine why certain areas were able to increase their share of employment in that industry while others were not. Such studies always have the problem of being limited to one industry, which may not be typical, while the cost of conducting that many separate studies is likely to prove prohibitive. However, since the regional share component is merely the algebraic sum of the amounts by which employment growth has exceeded or fallen short of the national average for each industry, the regional share component can be used as the dependent variable when it is desired to examine the sum of the effects of a particular factor on a number of industries. The procedure is made simplier since there is a published set of regional share components covering two decades for all counties and states.[21] Aggregations of counties into larger units can be made simply by addition of components. To illustrate this procedure the influence of two variables on regional share components will be discussed.

To examine the hypothesis that low wages encourage employment growth the state regional share (in percentage form) for the fifties was regressed on family income in 1960, used as a surrogate for wage rates. Surprisingly, no significant correlation was found ($R^2 = 0.40\%$, t = .436805). The simplest explanation is that in traditional economic theory low wages have two effects. They encourage employers to expand and encourage workers to migrate out. Whether the net effect of low wages is an increase or decrease in employment depends on which effect dominates, and on whether employers or workers are more mobile. There is no prior reason for expecting one effect to dominate, although students of industrial location typically expect the first effect to dominate, while students of labor economics and migration typically assume dominance of the latter. Apparently the two effects are cancelling each other out.

Greater success was obtained with a climatic variable. When the state regional share component was regressed on the mean January temperature of the largest city in each state the following equation was obtained.

(Eq. 12)

REGS56 = −16.0351 + .005356 TEMP
                                (.001677)
          Standard Error = 13.99% $R^2 = 17.85\%$

There is a significant relationship between climate and employment growth after controlling for the role of industrial mix. The areas with warm winters show a higher increase in employment, although this explains less than a fifth of the variance between states.

Obviously, much more research is needed before we can say confidently what determines the regional share component.

Regional Equations

Similar regressions to those reported have been run for six regions of the country. In general, these equations are similar to those reported above. In particular, there is a highly significant correlation between the regional share components in successive decades.

TABLE 1

Regression of County Industrial Mix, 1950-1960 on County Industrial Mix 1940-1950
For the United States and for Selected Regions

	Numerical Form					Percentage Form				
	C	Coeff.	St. Error	T	R²	C	Coeff.	St. Error	T	R²
All	.430204	.822467	2636.10	121.484	82.63	−03.62044	.744755	−06.75944	83.8008	69.39
New England	−275.244	.961869	2570.08	24.5510	90.27	−01.27465	.810420	03.67055	16.3563	80.45
Mid East	185.642	.882854	4648.79	37.5334	88.89	−00.672348	.722339	04.43085	19.8227	69.07
Lakes	−145.586	.671968	3782.83	46.7456	83.43	−01.64932	.814601	04.1922	39.1656	77.95
Mfg. Belt						−01.38443	.808365	04.23961	50.4118	78.92
Plains	−17.2692	.926657	658.081	80.3407	90.29	−01.81336	.936881	05.22161	45.5392	74.92
South	55.4428	.821719	1302.97	51.6637	67.17	−05.30043	.644319	07.86244	43.3682	59.05

New England comprises: Connecticut, Massachusetts, Rhode Island, New Hampshire, Vermont, Maine
Mid East comprises: New York, New Jersey, Pennsylvania, Delaware, Maryland, District of Columbia
Lakes comprises: Ohio, Indiana, Illinois, Michigan, Wisconsin
Manufacturing Belt comprises: the Mid East, New England, the Great Lakes
Plains comprises: Minnesota, Iowa, Missouri, the Dakotas, Nebraska, Oklahoma, Kansas
South comprises: Virginia, the Carolinas, Georgia, Florida, West Virginia, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Texas

TABLE 2

Regressions of County Regional Share, 1950-1960 on County Regional Share 1940-1950
For the United States and for Selected Regions

	Numerical Form					Percentage Form				
	C	Coeff.	St. Error	T	R^2	C	Coeff.	St. Error	T	R^2
All	.0337857	1.38869	7905.38	106.596	78.57	−01.15583	.315067	18.0766	18.5936	10.03
New England	2398.85	1.01995	9131.86	11.7564	68.01	00.581247	.351885	10.0782	3.35654	14.77
Mid East	2439.47	1.40328	16,669.0	33.5809	86.50	−00.85990	.598190	15.2891	8.06796	27.00
Lakes	−533.535	1.39468	5408.57	76.4860	93.09	−02.42951	.545623	11.5961	9.63104	17.61
Mfg. Belt						−015.7177	.510225	12.6127	13.0998	20.17
Plains	−255.10	1.92822	2803.72	62.5392	84.93	−04.97844	.351369	14.6972	8.32806	9.08
South	−94.3919	1.07605	5121.52	30.3777	41.44	−00.289371	.275082	18.9585	11.0317	8.54

FOOTNOTES

[1]See Ashby, Lowell D., *Growth Patterns in Employment by County 1940-1950 and 1950-1960*, Office of Business Economics, Department of Commerce, 1965 and Ashby, Lowell D., "The Geographical Redistribution of Employment: An Examination of the Elements of Change," *Survey of Current Business*, Vol. 44, No. 10, Oct. 1964, 13-20.

[2]The raw data is published in Office of Business Economics, *Growth Patterns in Employment by County, 1940-1950 and 1950-1960*.

[3]The t statistic is the ratio of the standard error of a coefficient to the value of the coefficient. Under certain conditions, including homoscedasticity, the probability that the observed correlation was due solely to chance can be calculated from the t statistic. The critical values for the t statistic decrease with the numbers of observations, eventually converging on 1.96 for 95% probability that the coefficient is significantly different from zero, and 2.58 for 99% probability of being significantly different from zero. For large numbers of observations such as used in the regressions reported in this paper, the above values can be used to provide tests of significance for homoscedastic data. Unfortunately, as discussed in the paper, much of the data used in this paper does not meet the conditions required for the application of traditional tests of significance. To avoid implying more statistical significance than can be justified, the t statistics have been reported without converting them to a percentage probability.

[4]Assuming that some adjustment is made for expected changes in the national growth rates. (In effect making the constant term in the estimating equation zero.)

[5]In more technical terms there is a downward bias due to measurement error in the dependent variable.

[6]The typical county does not necessarily have a decline in growth rates that is equal to the weighted average for all counties.

[7]These results are included in a longer paper available from the author.

[8]These are as follows: (1) Less heteroscedasticity because the difference in absolute size of states between large and small states are less than the differences in county sizes; (2) Most states are a mixture of rural and urban areas while many counties are entirely rural or urban. Thus, it makes a large difference whether one works with the regional share and industrial components in a percentage or a numerical form.

[9]The figures were deflated by half of the sum of the employment at the beginning and end of the period.

[10]As will be seen in the discussion of county projections, this is not the best procedure when forecasting employment at the county level. The reason for the difference is that for small areas such as counties, the historical regional share component is less likely to reflect accurately the competitive position of the area.

[11]Errors in forecasting at the county level are likely to be about twice as large.

[12]It should be noted that to obtain this accuracy it would be necessary to know the coefficients for Equation #5. As these would not be known the actual forecasting accuracy would be worse.

[13]See "The Role of Factor Supplies in Regional Growth," *Review of Regional Studies*, Vol. III, No. 3 (Spring 1972-1973) pp. 61-78.

[14]See Ashby, Lowell D., "The Geographical Redistribution of Employment: An Examination of the Elements of Change," *Survey of Current Business*, Vol. 44, No. 10, October 1964, pp. 13-20.

[15]See Johnston, J., *Econometric Methods*, McGraw-Hill 1963, Chapter 6, "Errors in Variables," pp. 148-150.

[16]"Shift and Share Projections of Regional Economic Growth: An Empirical Test," *Journal of Regional Science*, Vol. 9, No. 1, April 1969, pp. 1-18.

[17]The value was 1.0318 for states (Eq. 3) and 1.389 for populous counties (Eq. 10).

[18]See "Growth Potential Identification and Public Investment Strategy" in *Regional Science Perspectives*, Vol. 1, 1971, p. 115.

[19]See "The Role of Factor Supplies in Regional Growth," *Review of Regional Studies*, Vol. 111, No. 3 (Spring 1972-1973), pp. 61-78.

[20]An alternative approach would have been to divide the sample into two groups, rural and urban, and examine each separately. Unfortunately, the author's period with the New England Regional Commission was too short to permit the conducting of this experiment.

[21]*Growth Patterns in Employment by County—1940-1950 and 1950-1960,"* Office of Business Economics, U.S. Department of Commerce.