ON THE ANALYSIS OF REGIONAL GROWTH PATTERNS*

Richard Emmerson, R. Ramanathan, and Wolfhard Ramm†

1. INTRODUCTION

A frequently used device in the analysis of regional growth patterns is shift and share analysis. Although the procedure is primarily descriptive, it has often been used to project regional growth rates,[1] and to identify key industries and regional effects which might explain growth differentials.[2] This type of analysis has a number of well-known shortcomings when used for purposes other than description. Two major shortcomings are: (1) The procedure is not based on an explicit theoretical economic model, and (2) the procedure is not explicitly econometric. In other words, it does not allow rigorous tests of hypotheses or permit construction of confidence intervals for projections A recent paper by James and Hughes [8] attempts to carry out such a test but, in our opinion, it does not go far enough. For instance, their approach cannot be used to test for interactions between regions and industries, nor can confidence intervals be obtained.

In this paper we suggest the use of well-known techniques of covariance analysis[3] as an alternative to shift and share analysis. We argue that the approach is vastly superior on statistical grounds. The theoretical rationale for the procedure used is also discussed. In Section 2 we develop a covariance model which can be used to analyze growth rates. Possible extensions are discussed in Section 3. The approach is then applied to employment data for a number of Standard Metropolitan Statistical Areas (SMSA's) in California in Section 4. Section 5 briefly discusses possible economic interpretations of the approach and conclusions.

2. COVARIANCE ANALYSIS OF RATES OF GROWTH

A powerful technique which permits a systematic decomposition of the growth rate into regional share, industry mix, etc., and which also enables us to formally test

---

* The authors are indebted to the U.S. Department of Housing and Urban Development through the National League of Cities for financial support for portions of this project which is part of a larger study of regional growth patterns and implications of alternative regional policies. Alan Friedman, Harold Nelson, Tom Shipley and Paul Sommers provided able assistance in the collection of data, computer programming, and preparation of tables.

† Assistant Professor, Associate Professor, and Assistant Professor, Department of Economics, University of California, San Diego, respectively.

[1] See, for example, Dunn [4] and National Planning Association [10]. A critique and statistical test of shift and share projections can be found in Brown [3] and James and Hughes [8].

[2] See Perloff, *et al.* [11], Ashby [1, 2], Fuchs [6] for examples of applications and Houston [7] for a critique.

[3] For an introduction to covariance analysis see Frank [5]. More advanced discussions are in Scheffé [12] and Malinvaud [9].

Date received: March, 1974; revised, July, 1974.

the statistical significance of the local and industry effects is covariance analysis. This is a very well-known technique which unfortunately has not been well exploited by investigators of regional growth patterns. One reason for the lack of widespread use of this tool is that it generally requires more data than shift-share analysis. But considering the benefits derived from this method, the extra costs of data gathering (when possible) would be well spent. The procedure we adopt here is essentially the estimation of exponential (or other) time trends. In this section we specify the details of this approach in the context of a specific regional employment model. Define

$$x_{ki}(t) = \log[X_{ki}(t)/X_{ki}(0)] \qquad t = 0, 1, \cdots, T \tag{1}$$

where $X_{ki}(t)$ is the employment at time t in the ith industry within the kth location; $x_{ki}(t)$ is thus the exponential rate of growth between periods 0 and t. A covariance model with complete interactions is set up below:

$$x_{ki}(t) = (\alpha + \beta_k + \gamma_i + \delta_{ki})t + \epsilon_{ki}(t) \tag{2}$$

for $k = 1, \cdots, K$ (SMSA's); $i = 1, \cdots, I$ (Industries); $t = 1, \cdots, T$ (Time) and $\sum_k \beta_k = 0$, $\sum_i \gamma_i = 0$, $\sum_k \delta_{ki} = 0$, for all i, and $\sum_i \delta_{ki} = 0$ for all k.

The overall rate of growth is α, β_k is the class (SMSA) effect, γ_i is the industry effect and δ_{ki} is the interaction between the kth SMSA and the ith industry. It is assumed that $\epsilon_{ki}(t)$ is a normally distributed random error with zero mean and constant variance. The interpretation of the parameters of this model can be facilitated by expressing them in terms of industry-SMSA growth rates. Let u_{ki} be the growth rate of industry i in SMSA k. Then[4]

$$\begin{aligned} \alpha &= u_{..} \\ \beta_k &= u_{k.} - u_{..} \\ \gamma_i &= u_{.i} - u_{..} \\ \delta_{ki} &= u_{ki} - u_{k.} - u_{.i} + u_{..} \end{aligned} \tag{3}$$

where $u_{..} = \sum_k \sum_i u_{ki}/KI$, $u_{k.} = \sum_i u_{ki}/I$, and $u_{.i} = \sum_k u_{ki}/K$.

Thus α is the overall rate of growth of employment, β_k is the difference between the growth rate of SMSA k and the overall (state) rate, γ_i is the difference between the growth rate of industry i and the overall rate, and δ_{ki} represents the difference between the growth rate of a particular industry within an SMSA and the growth rate of all industries in the state adjusted for industry and SMSA effects.

2.1 *Testing for the Significance of Industrial Mix and Regional Share*

In the context of the covariance model specified above, tests of the significance of regional effects and the existence of key industries are readily developed. Assuming that the data refer to K SMSA's within a state, to test whether industry i within SMSA k (industry ki) differs significantly from statewide trends, the

[4] It is easy to show that industry and SMSA growth rates can be expressed as simple averages of within industry-SMSA growth rates in the context of Model (2) by computing the averages indicated in Equations (7a)–(7d) below. It is shown in Section 3 that these relations need not hold in more general models.

hypothesis is

(4a) $$H_0: u_{ki} - u_{..} = 0$$

The hypothesis that industry ki differs significantly from statewide trends for industry i is

(4b) $$H_0: u_{ki} - u_{.i} = 0$$

The following hypothesis tests for the significance of the difference between industry ki and the growth trends of other industries within SMSA k:

(4c) $$H_0: u_{ki} - u_{k.} = 0$$

The interaction term δ_{ki} provides the hypothesis that industry ki differs significantly from the statewide trend after adjusting for the effects of industry i and SMSA k. This hypothesis is specified as

(4d) $$H_0: u_{ki} - u_{k.} - u_{.i} + u_{..} = 0$$

An industry which is very sensitive to local effects may, however, not appear so since its growth rate may equal the SMSA and industry-adjusted statewide growth rate by chance. In this case, the hypothesis in (4d) would not be rejected, indicating the lack of a local effect for that industry.[5] The following set of tests for industries across SMSA's are therefore of considerable interest:

(4e) $$H_0: u_{1i} = u_{2i} = \cdots = u_{Ki} = u_{.i}$$

or equivalently

$$H_0: \delta_{1i} = \delta_{2i} = \cdots = \delta_{Ki} = 0$$

Such a test would indicate whether industry i is subject to local effects anywhere in the state and hence possibly within the SMSA of interest.[6]

The signs and magnitudes of individual parameters should also be useful in the identification of key industries. In particular, ranking industries by the value of γ_i indicates which industries are growing fastest within the state, ranking by $\gamma_i + \delta_{ki}$ indicates the fastest growing industries within SMSA k, etc. The magnitude of the deviations in the hypotheses (4a) through (4d) will also be important, since small deviations may turn out to be very significant given a sufficiently large sample.

Other tests which are of interest are:

[5] This problem also arises in the interpretation of the results from a shift and share analysis.

[6] Tests for the equality of industry i growth rates within subsets of particular SMSA's (e.g., densely populated SMSA's, coastal SMSA's, etc.) can be readily computed in an analogous manner. Such a procedure would permit the investigator to implicitly control certain SMSA characteristics.

(4f) $$H_0\colon \beta_k = u_{k.} - u_{..} = 0$$

(4g) $$H_0\colon \gamma_i = u_{.i} - u_{..} = 0$$

(4h) $$H_0\colon u_{k1} = u_{k2} = \cdots = u_{kI} = u_{k.}$$

2.2 *Estimation of Parameters and Computation of Test Statistics*

One approach to the estimation of the parameters in Equation (2) is to use regression analysis with dummy variables. We define KI dummy variables as follows:

$$z_{ki} = \begin{cases} 1 & \text{if the observation is for industry } i \text{ in SMSA } k \\ 0 & \text{otherwise.} \end{cases}$$

Estimates of the desired parameters can then be computed from the following regression:

(5) $$x_{ki}(t) = \textstyle\sum_k \sum_i b_{ki}(z_{ki}t) + \epsilon_{ki}(t)$$

Thus, for example, in the case of two industries, two SMSA's and three time periods, the vector of dependent variables, x' and the corresponding matrix of independent variables, Z', would be:

$$x' = [x_{11}(1)\ x_{11}(2)\ x_{11}(3)\ x_{21}(1)\ x_{21}(2)\ x_{21}(3)\ x_{12}(1)\ x_{12}(2)\ x_{12}(3)\ x_{22}(1)\ x_{22}(2)\ x_{22}(3)]$$

$$Z' = \begin{bmatrix} 1 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 \end{bmatrix}$$

The cross product matrix of the independent variables thus has a particularly simple structure in this case:

(6) $$Z'Z = \sum_t^T t^2 \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} = \frac{T(T+1)\,(2T+1)}{6} \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

where the matrices on the right are $KI \times KI$ identity matrices. Least squares estimates of the coefficients of Equation (5) can therefore be readily computed without inverting the $Z'Z$ matrix. This means that special computer programs can readily be written which are less costly to run than standard regression programs and which can readily compute any number of coefficients. It is easy to show that the least squares estimators of the coefficients of the above model are:

(7a) $$b_{ki} = \textstyle\sum_t x_{ki}(t)z_{ki}(t) / \sum_t z_{ki}(t)^2 = \sum_t x_{ki}(t)t / \sum_t t^2 = \hat{u}_{ki}$$

(7b) $$b_{k.} = \textstyle\sum_i b_{ki}/I = \hat{u}_{k.}$$

(7c) $$b_{.i} = \textstyle\sum_k b_{ki}/K = \hat{u}_{.i}$$

(7d) $$b_{..} = \textstyle\sum_k \sum_i b_{ki}/KI = \hat{u}_{..}$$

where the hats denote estimated values of the growth rate for the designated classifications. Estimates of the parameters of Equation (2) can thus be computed from least squares estimates of the coefficients of Equation (5) using the definitions given in (3).

The hypotheses (4a) through (4d) can also be expressed as weighted sums of these coefficients. Thus, for example, (4c) becomes:

$$H_0\colon b_{ki} - (1/I) \textstyle\sum_i b_{ki} = 0 \tag{8}$$

A test of this hypothesis can be based on the t-statistic:

$$t = \frac{b_{ki} - (1/I) \sum_i b_{ki}}{[\mathrm{var}\,(b_{ki} - (1/I) \sum_i b_{ki})]^{1/2}} \tag{9}$$

The variance in the denominator of (9) can be computed from the covariance matrix of the coefficients of Equation (5). Making use of Equation (6), it is obvious that the covariance matrix is given by

$$s_e^2[Z'Z]^{-1} = (s_e^2/D)\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} \tag{10}$$

where s_e^2 is the residual variance of Equation (5) estimated as given by

$$s_e^2 = [1/(TKI - KI)][\textstyle\sum_k \sum_i \sum_t x_{ki}^2(t) - D \sum_k \sum_i b_{ki}^2] \tag{11}$$

and

$$D = T(T+1)(2T+1)/6 \tag{12}$$

The individual variances are:

$$\mathrm{var}\,(b_{ki}) = s_{ki}^2 = s_e^2/D \qquad \text{for all } k,\ i. \tag{13}$$

Therefore, since all the covariances are zero, the t-statistic for testing (4c) becomes:

$$t = (b_{ki} - b_{k.}) \Big/ \left[s_{ki}\left(1 - \frac{1}{I}\right)^{1/2} \right] \tag{14c}$$

with $TKI - KI$ degrees of freedom.

By proceeding similarly, it is readily shown that the t-statistics for the other tests are as below:

$$t = (b_{ki} - b_{..}) \Big/ \left[s_{ki}\left(1 - \frac{1}{KI}\right)^{1/2} \right] \tag{14a}$$

$$t = (b_{ki} - b_{.i}) \Big/ \left[s_{ki}\left(1 - \frac{1}{K}\right)^{1/2} \right] \tag{14b}$$

$$t = (b_{ki} - b_{k.} - b_{.i} + b_{..}) \Big/ \left[s_{ki}\left(1 - \frac{1}{I} - \frac{1}{K} + \frac{1}{KI}\right)^{1/2} \right] \tag{14d}$$

$$t = (b_{k.} - b_{..}) \Big/ \left[s_{ki}\left(\frac{1}{I} - \frac{1}{KI}\right)^{1/2} \right] \tag{14f}$$

$$t = (b_{.i} - b_{..}) \Big/ \left[s_{ki}\left(\frac{1}{K} - \frac{1}{KI}\right)^{1/2} \right] \tag{14g}$$

To test the joint hypotheses in (4e) of the equality of the growth rates across SMSA's, or equivalently, of the absence of interaction effects across SMSA's for the ith industry, we reestimate model (2) under the additional restrictions imposed by (4e). We thus estimate for each industry, $j = 1, \cdots, I$,

$$x_{ki}(t) = (\alpha + \beta_k + \gamma_i + \delta_{ki})t + \epsilon_{ki}(t) \tag{15}$$

under the restrictions $\sum_k \beta_k = 0$, $\sum_i \gamma_i = 0$, $\sum_k \delta_{ki} = 0$ for all i, $\sum_i \delta_{ki} = 0$ for all k, and $\delta_{1j} = \delta_{2j} = \delta_{3j} = \cdots = \delta_{Kj} = 0$, and set up an analysis of variance table for each industry, $j = 1, \cdots, I$, as follows:

Source of Variation	Sum of Squares	Degrees of Freedom	Sum of Squares	Degrees of Freedom
Regression	RS_2	KI	RS_{15j}	KI − K + 1
Error	ES_2	TKI − KI	ES_{15j}	TKI − (KI − K + 1)
Total	TS_2	TKI	TS_{15j}	TKI

The F-statistic to test $u_{1j} = u_{2j} = \cdots = u_{Kj} = 0$ is given by

$$F = \frac{(ES_{15j} - ES_2)/(K - 1)}{ES_2/(TKI - KI)}$$

or

$$F = D[\textstyle\sum_k b_{kj}^2 - Kb_{.j}^2]/(K - 1)s_e^2 \tag{16}$$

since it is easily verified that

$$ES_{15j} = ES_2 + D[\textstyle\sum_k b_{kj}^2 - Kb_{.j}^2]$$

Under the null hypothesis, (16) has the F-distribution with $(K - 1)$ and $(TKI - KI)$ degrees of freedom. A similar F-test can be constructed for (4h), giving one F-value for each SMSA.

Thus, the analysis of covariance provides us with a systematic decomposition as well as statistical tests of a variety of hypotheses regarding regional and industrial effects and interactions.

2.3 *Projections of Employment*

The growth rates estimated in the model may be used to make projections of employment in the kith SMSA-industry. The employment, $\hat{X}$, at time T_0 is given by

$$\hat{X}_{ki}(T_0) = X_{ki}(0)e^{b_{ki}T_0} \tag{17}$$

It is also possible to obtain a 95 percent confidence interval for $\hat{X}$. The large sample confidence interval for $x_{ki}(T_0)$ is given by

$$b_{ki}T_0 \pm 1.96\, s_{ki}T_0 \tag{18}$$

where s_{ki} is given by (13). The confidence interval for $X_{ki}(T_0)$ is obtained by substituting (18) into (17)

$$\hat{X}_{ki}(T_0) = X_{ki}(0) \exp\,[(b_{ki} \pm 1.96\, s_{ki})T_0] \tag{19}$$

2.4 *Changing the Base Period*

It will be interesting to examine how sensitive the results are to a different choice of the base period. Let period 1 be the new base period and

$$y_{ki}(t) = \log [X_{ki}(t)/X_{ki}(1)] \tag{20}$$

Then

$$x_{ki}(t) = y_{ki}(t) + x_{ki}(1) \tag{21}$$

Let

$$a_{ki} = \sum_t y_{ki}(t)t/\sum t^2 \tag{22}$$

We then have

$$a_{ki} = b_{ki} - [3/(2T + 1)]x_{ki}(1) \tag{23}$$

It is evident that the estimated growth rate does generally differ with the choice of the base period, but if T is relatively large, the difference is not likely to be much. However, the *deviations* of a_{ki} from $a_{k.}$ and $a_{.i}$ will have zero expectations under the null hypotheses (4c) and (4b) respectively. We have

$$a_{ki} - a_{k.} = (b_{ki} - b_{k.}) - [3/(2T + 1)][x_{ki}(1) - \bar{x}_{k.}(1)] \tag{24}$$

where $\bar{x}_{k.}(1)$ is the average over industries of $x_{ki}(1)$. Under the hypothesis (4c) that $u_{ki} = u_{k.}$, all industries in the kth SMSA grow at the same rate, and hence the expected value of (24) will be zero. Thus, a change in the base period has no effect on the expected value of differences in growth rates. The choice of a base period is therefore not a serious issue, since interest centers on the significance of differences in growth rates. This is in substantial contrast to shift-share analysis which in most applications is quite sensitive to choice of base years. In the case of a period-to-period model specified in the next section, (26), the question of such sensitivity does not arise.

3. ALTERNATIVE SPECIFICATIONS, EXTENSIONS AND LIMITATIONS OF THE APPROACH

(1) Although the covariance model used in this study is set up in terms of the exponential growth rates, in some situations it may be more appropriate to use other formulations. For instance, if a linear trend is more appropriate, we have

$$X_{ki}(t) = \alpha_0 + (\alpha + \beta_k + \gamma_i + \delta_{ki})t + \epsilon_{ki}(t) \tag{25}$$

In a situation in which a growth rate has initially fallen and later risen (or vice versa), the following quadratic formulation may be more appropriate:

$$x_{ki}(t) = (\alpha + \beta_k + \gamma_i + \delta_{ki})t + (\lambda + \mu_k + \theta_i + \pi_{ki})t^2 + \epsilon_{ki}(t)$$

Another alternative is to express the period-to-period change as the dependent variable with no time trend. This gives

$$X_{ki}(t)/X_{ki}(t-1) = \alpha + \beta_k + \gamma_i + \delta_{ki} + \epsilon_{ki}(t) \tag{26}$$

Note that this becomes a pure analysis of variance model with interactions.

(2) It should be emphasized that a part of the maintained statistical hypothesis for the tests discussed in Section 2.1 is $E[\epsilon_{ki}^2(t)] = \sigma^2$, a constant, for all k, i, t. If it is believed that this specification does not hold, a generalized least squares approach should be considered.[7]

(3) It should be noted that class and subclass coefficients are not in general given as simple arithmetic means of regression coefficients for individual cells (i.e., b_{ki}). Suppressing subclass effects for the moment, the sums of squares of individual variables about the grand mean can be decomposed as follows:

$$\sum_k \sum_t (z_{ki} - z_{..})^2 = \sum_k \sum_t (z_{kt} - z_{k.})^2 + T \sum_k (z_{k.} - z_{..})^2$$

Similar decompositions follow for x_{ki} and the cross products of x_{kt} and z_{kt}. The overall regression coefficient for the sample would thus be given by[8]

$$\frac{\sum_k \sum_t (x_{kt} - x_{k.})(z_{kt} - z_{k.}) + T \sum_k (x_{k.} - x_{..})(z_{k.} - z_{..})}{\sum_k \sum_t (z_{ki} - z_{k.})^2 + T \sum_k (z_{k.} - z_{..})^2}$$

This expression reduces to the type of expressions given in (6) only because $z_{k.} = z_{..}$ for the independent variables used in this study (i.e., between class and subclass, variances and covariances are all zero).

(4) The major limitation of the approach adopted in this study is the lack of a behavioral model of employment determination. The specification is simply in the form of a time trend and does not incorporate specific economic variables which determine the levels of employment in various industries. Our purpose here is limited to demonstrating the advantages of covariance analysis relative to shift-share analysis as a technique to identify significant differences in patterns of regional economic growth.

4. AN APPLICATION TO INDUSTRIAL EMPLOYMENT IN CALIFORNIA

The model developed in Section 2 has been used to help identify industry and SMSA effects in the growth of total employment in California. Data on total employment by industry and SMSA were obtained for the period 1963–1971 from the U.S. Department of Labor [13]. The year 1963 was chosen as the initial period since the data were incomplete for a number of SMSA's in earlier years.

The covariance model was estimated using data normalized by base period employment, i.e., Equation (5), and using the previous period's employment for normalization. The fit of the base period model was judged to be superior on pragmatic grounds, since the base period model produced more significant coefficients and much higher R^2 (for the period 1963–1971, the R^2 was .925 for the base period model

[7] A general estimation and hypothesis testing approach which could be applied in this case is Zellner's [14] seemingly unrelated regressions model. It should be noted that there would be no gain in efficiency in this case since the observations for the independent variable are identical in each equation. The approach would, however, produce asymptotically unbiased tests and substantially greater computational costs.

[8] This estimator can also be readily modified to allow for different numbers of observations in each industry-SMSA cell.

TABLE 1: Selected Results for the San Diego SMSA, 1963–1971

Industry	(1) b_{ki}	(2) $t(b_{ki} - b_{..})$	(3) $t(b_{ki} - b_{.i})$	(4) $t(b_{ki} - b_{k.})$	(5) $t(b_{ki} - b_{k.} - b_{.i} + b_{..})$	(6) $F(b_{1i} = b_{2i} = \cdots = b_{ki} = b_{.i})$	(7) $b_{.i}$	(8) $t(b_{.i} - b_{..})$
Mining	.028	−1.49	4.03	−4.25	1.48	93.59	.007	−22.71
Construction	.037	.05	8.34	−2.63	6.02	31.29	−.008	−33.82
Manufacturing	.019	−3.27	−1.65	−6.12	−4.50	43.30	.027	−7.04
Transportation and Utilities	.054	3.16	3.08	.64	.48	13.43	.038	.68
Wholesale Trade	.056	3.55	1.82	1.05	−.85	30.04	.047	7.48
Retail Trade	.062	4.51	2.49	2.07	−.15	10.75	.048	8.83
Finance	.055	3.29	1.32	.78	−1.37	8.71	.048	8.44
Insurance and Real Estate	.060	4.27	4.69	1.81	2.18	44.95	.035	−1.21
Service	.073	6.63	1.25	4.29	−1.46	7.83	.067	22.80
Government	.063	4.80	.88	2.36	−.184	8.94	.058	16.57
All Industry	.051						.037	
Standard Deviation of Estimate	.0055	.0055	.0054	.0052	.0051			.0013

Residual Variance = .0062601, Individual Variance = .000030687, Individual Standard Deviation (S_{ki}) = .0055396, $K = 16, I = 10, T = 8, b_{k.} = .051$ (for k = San Diego), $b_{..} = .037$, $b_{k.} - b_{..} = .014$, $t(b_{k.} - b_{..}) = 8.301$. Critical $t^*_{.01} = 2.576$, critical $t^*_{.05} = 1.960$, critical $t^*_{.01} = 1.645$, $F^*_{.01}(9, \infty) = 2.41$, $F^*_{.05}(9, \infty) = 1.88$.

and .260 for the period-to-period model).[9] We therefore only report further results for the base period model.

4.1 *Application to San Diego*

In the interest of conserving space, the tables of individual coefficients and test statistics for all SMSA's are not presented but are available from the authors. Results are presented for only one SMSA, San Diego, with some mention of other SMSA's in the next section. All the results which relate to San Diego are summarized in Table 1. Column (1) gives the growth rates of industries within San Diego. Column (2) indicates that every one of these rates, except mining and manufacturing, was greater than the overall growth rate of employment in California. Column (3) shows that when compared to industry growth rates, only manufacturing grew more slowly in San Diego than the state. This is consistent with the fact that San Diego is growing significantly faster than California overall ($b_{k.} - b_{..} = .014$ i.e., employment in San Diego is growing 1.4 percent faster per year).

A much better indication of the strengths and weaknesses in the San Diego economy is given by the significance of the interaction effects, Column (5). Both the insurance-real estate and construction industries grew significantly faster than industry and SMSA growth rates would lead one to expect, while the growth rates

[9] The models are, of course, not directly comparable since the dependent variable, whose variation we are attempting to explain, is defined differently in the two models.

for manufacturing and government were significantly lower.[10] Reinspection of Columns (3) and (4) yields some insight into the source of these interactions. The positive interaction for construction and real estate-insurance can be explained by their ability to grow faster than one would have expected on the basis of industry growth rates ($b_{ki} - b_{.i}$ highly significant). Manufacturing's negative interaction can be attributed to its inability to grow as fast as total employment growth in San Diego would lead one to expect.

Columns (6), (7), and (8) summarize general industry characteristics for all of California. The F-statistics in Column (6) can be interpreted as an index of the extent to which the growth of employment in an industry is affected by conditions specific to particular regions. The identification of such conditions or factors, if they exist, is of considerable interest since they may be subject to local policy manipulation. A high F does not, however, necessarily imply that the industry can be readily influenced by local policies. For example, the extremely high F for mining is probably due to the heterogeneous distribution of natural resources, rather than differences in local institutions or policies. The high F's of manufacturing, construction, real estate-insurance and wholesale trade may, however, indicate that the growth of employment in these industries is influenced by local conditions, and these conditions may be subject to influence by local policy makers.

It is interesting to note that the service, and service-related (finance, retail trade, government) industries have the lowest F-statistics. This is as one would expect since the demand for these industries probably depends primarily on local population and income growth, which vary less across SMSA's and are less likely to be locally influenced than many other conditions. This suggests, for example, that efforts to encourage or discourage the growth of these industries may not be highly successful.

Column (7) simply gives the overall growth rates for industries within the state, and Column (8) shows which industries in California are growing significantly slower or faster than the state total.

Similar tables can be constructed for the other SMSA's. Sufficient information to carry out such constructions for California SMSA's may be obtained from the authors.

4.2 *Coastal Versus Interior SMSA's*

It is possible to extend the covariance analysis to take account of additional control variables by stratifying the SMSA's or industries into groups. As an illustration, we arranged the SMSA's into two groups, coastal and interior, and carried out a covariance analysis of each group. Table 2 lists, for each industry, the areas which had a significant (at a 5 percent level) positive or negative interaction. This has been done for coastal and noncoastal metropolitan areas.

It is interesting to note that Los Angeles does not appear anywhere in the table. This indicates that if we adjust the growth rate of Los Angeles for SMSA, coastal and industry effects, no significant interaction is left. In other words, among coastal SMSA's, Los Angeles does not have a unique industry. As one would expect, Sacra-

[10] The interaction effects are significant at a confidence level of over 90 percent for government, 95 percent for the real estate and insurance industry and 99 percent for the construction and manufacturing industries.

TABLE 2: Coastal Vs. Interior SMSA Comparison

Industry	Coastal SMSA's		Interior SMSA's	
	Positive	Negative	Positive	Negative
Mining	Santa Rosa, Salinas	San Diego, Oxnard	Bakersfield, Riverside	Fresno, Sacramento, Vallejo
Construction	San Diego, San Francisco	Santa Barbara, Salinas	Bakersfield, Fresno, Modesto, Vallejo	Anaheim, Riverside, San Jose
Manufacturing	Santa Rosa, Salinas	San Diego, Santa Barbara	Modesto, Riverside, Vallejo	Anaheim, Sacramento
Transportation Utilities	San Francisco	Oxnard	San Jose, Vallejo	Bakersfield, Fresno, Riverside
Wholesale Trade	Santa Barbara, Salinas	San Francisco, Oxnard	Anaheim, San Jose	Modesto, Riverside, Stockton
Retail Trade	Oxnard	None	Sacramento, Vallejo	Modesto, Stockton
Finance	San Francisco, Oxnard	San Diego, Santa Rosa, Salinas	Sacramento	Modesto, Riverside
Insurance and Real Estate	Oxnard	Santa Barbara, San Francisco, Santa Rosa, Salinas	Anaheim, Sacramento	Modesto, Vallejo
Service	Oxnard	Salinas	Sacramento	Bakersfield, San Jose
Government	Santa Barbara	San Diego, Salinas	Fresno, Sacramento	Bakersfield, Riverside

mento has its key industries concentrated in government, service, finance, insurance and retail trade. Manufacturing and mining show a significant negative interaction for that SMSA. In the case of Modesto, the opposite result holds. The export base industry, such as manufacturing, shows a significant positive interaction indicating that in Modesto manufacturing employment is growing faster, even after adjustment for growth in all interior SMSA's and all industries as a whole. The softer industries like finance, retail trade, insurance and real estate show a negative interaction indicating a decline relative to industry and SMSA effects.

Among coastal metropolitan areas, Oxnard shows a pattern similar to Sacramento. Retail trade, finance, insurance, real estate and service industries grow faster relative to SMSA and industry growth rates, whereas wholesale trade, mining, transportation and utilities showed negative interactions. The results for other metropolitan areas may be obtained from Table 2.

5. CONCLUSIONS

In this paper we have pointed out some deficiencies in using shift and share analysis to examine growth patterns and identify key industries. We have proposed covariance analysis as an alternative tool not only because it provides a systematic procedure for decomposing observed growth rates into regional and industry effects and interactions, but also because it enables us to test for the statistical significance of a given component. Furthermore, it is possible to make projections and also obtain their confidence intervals. As mentioned earlier, the procedure still has the weakness that there is no underlying economic model which explains the differences between the growth rates of regions or industries. The analysis, however, can be readily extended to include explanatory variables other than a time trend or add other controls, such as coastal and interior SMSA's. Of particular interest to regional decision makers would be the inclusion of explanatory policy-related variables. The extended model would thus have several explanatory variables (perhaps several endogenous variables as well), the coefficients of which can be decomposed into region or industry effects and interactions.

REFERENCES

[1] Ashby, L. D. "The Geographical Redistribution of Employment: An Examination of Elements of Change," *Survey of Current Business*, 44 (1964), 13–20.

[2] ———. "Regional Change in a National Setting," Staff Working Paper in Economics and Statistics, No. 7, U.S. Dept. of Commerce, April 1964.

[3] Brown, H. J. "Shift and Share Projections of Regional Economic Growth: An Empirical Test," *Journal of Regional Science*, 9 (1969), 177–188.

[4] Dunn, E. S., Jr. "A Statistical and Analytical Technique for Regional Analysis," *Papers, Regional Science Association*, 6 (1960), 97–112.

[5] Frank, C. R., Jr. *Statistics and Econometrics*. New York: Holt, Rinehart and Winston, 1971.

[6] Fuchs, V. R. *Changes in the Location of Manufacturing in the United States Since 1929*. New Haven: Yale University Press, 1962.

[7] Houston, D. "Shift and Share Analysis: A Critique," *Southern Economic Journal*, 32 (1967), 577–581.

[8] James, F. and J. Hughes. "A Test of Shift and Share Analysis as a Predictive Device," *Journal of Regional Science*, 13 (1973), 223–231.

[9] Malinvaud, E. *Statistical Methods of Econometrics*. Chicago: Rand-McNally, 1966.

[10] National Planning Association. "Projection Procedure for Susquehanna Area Economic Base Study," Washington, D.C., 1964, mimeographed.

[11] Perloff, H. S. *et al. Regions, Resources and Economic Growth*. Baltimore: Johns Hopkins Press, 1960.

[12] Scheffé, H. *The Analysis of Variance*. New York: Wiley, 1959.

[13] U.S. Bureau of Labor Statistics. *Employment and Earnings Statistics for States and Areas*. Washington, D.C.: U.S. Government Printing Office, 1972.

[14] Zellner, A. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57 (1962), 348–368.