

NBER WORKING PAPER SERIES

NON-COINTEGRATION AND ECONOMETRIC EVALUATION OF
MODELS OF REGIONAL SHIFT AND SHARE

Scott J. Brown

N. Edward Coulson

Robert F. Engle

Working Paper No. 3291

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1990

This paper is preliminary and should not be quoted without the permission of the authors. It was written while the second author was a visitor at UC San Diego and he thanks that institution for its hospitality. The third author thanks NSF SES-8705884 for support. This paper is part of NBER's research program in Economic Fluctuations. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3291
March 1990

NON-COINTEGRATION AND ECONOMETRIC EVALUATION OF
MODELS OF REGIONAL SHIFT AND SHARE

ABSTRACT

This paper tests for cointegration between regional output of an industry and national output of the same industry. An equilibrium economic theory is presented to argue for the plausibility of cointegration, however, regional economic forecasting using the shift and share framework often acts as if cointegration does not exist. Data analysis on broad industrial sectors for 20 states finds very little evidence for cointegration. Forecasting models with and without imposing cointegration are then constructed and used to forecast out of sample. The simplest, non-cointegrating models are the best.

Scott J. Brown
San Diego Gas & Electric
San Diego, CA 92903

N. Edward Coulson
Pennsylvania State Univ.
University Park, PA 16802

Robert F. Engle
Univ. of California
La Jolla, CA 92093

Introduction

One of the abiding concerns of regional economics is the temporal behavior of aggregate regional economies. An important dimension of this behavior is the comparison of the region to the economic behavior of the nation in which it is located. Both as they follow and as they deviate from macroeconomic trends and cycles, the movements of regional employment, earnings and other aggregates are very often modeled as functions of national data. This concern extends to the behavior of particular industrial sectors as well, since regional economic policy is often based on an analysis of the performance of industries which are important to the region, in both actual and potential senses.

Emphasis on this relationship between national and regional economies is not misplaced given the experience of the past decade or so, since during that span the response of regions in the United States, and more particularly of certain sectors on those regions, to various movements in macroeconomic cycles has varied tremendously. Moreover, the shocks which have apparently induced such responses appear to have had persistent regional effects; of special note is the effect of supply shocks on manufacturing sectors in the upper midwest region, which may never regain the employment peaks of the 1970s, and the recent slump in the oil-based states of the southwest.

The regional persistence of shocks is not what might be expected; standard analysis presumes that factor price adjustment and factor migration to equalize those prices across regions would force an interregional equilibrium wherein all regions would grow at the same national rate in the long run. A key by-product of such a model's steady state is that the share of national output, employment and earnings generated in the region will be constant over time.

On the other hand, practical regional forecasters, often using the "shift and share" paradigm, frequently forecast the share of output, employment or earnings in

the region to remain at its current value. Such a model not only economizes on data requirements, but also is efficient if the share is a random walk. In this case, the share will not have a long run equilibrium.

This paper applies modern time series methods to the regional forecasting problem. Using the concept of cointegration to test for a long run equilibrium relationship between regions, rather startling results are obtained. It is found that industrial output in individual states is not cointegrated with national output in the same sector. The finding is consistent with the "shift share" model and has rather important economic implications.

In Section 2 of this paper we present a model of the relationship between regional economies; the model, while extraordinarily simple, is sufficient to examine the concepts, of shift and share, productivity shocks and the possible causes of non-cointegration or the persistence of regional shocks. In Section 3 we present the basic shift share analysis and in 4 we formulate the statistical model for regional earnings. Section 5 tests for the existence of an equilibrium relationship between regional and national earnings directly by testing for the cointegration of these two variables using the Engle and Granger(1987) test procedure. The striking result is that cointegration is not found, which casts severe doubt on equilibrium based models. Additionally, the non-cointegration results provide a basis for empirical model construction, suggesting that unit root models be given strong consideration. In Section 6, we propose nine models of regional shift, many of which have been previously analyzed. Some are non-cointegrating, disequilibrium based, and others are cointegrating and equilibrium based. These are examined for "out of sample" forecasting performance. The conclusion is that the models of disequilibrium are superior, especially as measured by their forecasting ability. Section 7 provides some summary remarks.

2. Regional Economic Models

We develop a very simple model for the output and payroll of region i and industry j . We suppose all regions have access to the same technology but that there are some fixed factors which are specific to a region. These could be pure location or weather effects or merely inherited and immobile physical capital. We consider labor as the variable factor of production and model region i , industry j at time t but suppress the j subscript throughout. Letting q_{it} be the log of output, ℓ_{it} the log of employment, a_i the log of the fixed factor, b_t a productivity shift which is independent of region, and α , the output elasticity of labor in this sector, the production function can be written for industry j as

$$(1) \quad q_{it} = a_i + b_t + \alpha \ell_{it}.$$

Defining w_{it} and p_{it} to be the log wage and log price of output in region i , the first order conditions for profit maximization with respect to the variable factor, labor, are given by

$$(2) \quad w_{it} - p_{it} = \log \alpha + q_{it} - \ell_{it}$$

The quantity most easily observed in regional data is payroll. Denoting its log as

$$(3) \quad \begin{aligned} y_{it} &= w_{it} + \ell_{it} \\ &= \frac{1}{1-\alpha} (p_{it}) - \frac{\alpha}{1-\alpha} w_{it} + \frac{1}{1-\alpha} \log \alpha + \frac{1}{1-\alpha} (a_i + b_t) \end{aligned}$$

which depends on the fixed factor, productivity, wages and prices.

The difference between log earnings for sector j in two regions i and k which each satisfy (1) and (2), is then

$$(4) \quad y_{it} - y_{kt} = \frac{1}{1-\alpha}(p_{it} - p_{kt}) + \frac{\alpha}{1-\alpha}(w_{kt} - w_{it}) + \frac{1}{1-\alpha}(a_i - a_k).$$

While it is possible for the prices of a good to differ from one region to another in the short run, they should be equal in the long run. Similarly through labor migration or the factor price equalization theorem, wage rates should be equalized in the long run.

Economic arguments thus suggest that in the long run, the ratio of earnings in two regions should be a constant depending upon $a_i - a_k$. It is therefore reasonable to model the relative wages and relative prices between two regions as stationary stochastic processes. Since the sum of such processes will itself be a stationary process, equation (4) implies that the log ratio should itself be a stationary process. Thus, shocks to wages, prices, employment or national productivity do not affect the long run ratio.

If productivity is integrated of order 1 but relative wage and prices are $I(0)$, then this is a simple example of cointegration as discussed by Engle and Granger (1987). Log earnings in each sector is integrated of order 1 from equation (3), but the log ratio is stationary.

The finding of cointegration is robust to several variations. If there is more than one variable factor and the output elasticities are constant, then the relative price of the new factor across regions will also appear in (4). For example, if capital is considered to be a variable factor of production with a price which is equalized across regions in the long run (a plausible assumption because of the highly developed capital market), then payroll will still be cointegrated.. Similarly, if labor is disaggregated into several factors of production, each with a constant output elasticity, then the argument above applies for each of the types of labor as well as for the sum of the payrolls. Finally, if there are differing costs of living in different regions and labor migration is assumed to equalize real wages across

regions, then the stochastic process of $w_{kt} - w_{it}$ should still be stationary but with a non-zero mean. Consequently, the long run equilibrium share will depend upon the relative cost of living as well as on the specific factors $a_k - a_i$.

Under what conditions will cointegration fail in the model? If the a_i change over time either through differential state productivities or through endogenous permanent shocks to a_i , then cointegration will not be found. Alternatively, if wages and/or prices have no long run tendency to equalize across regions, then there will be no cointegration. If the output elasticity is not a constant, then cointegration will fail if the elasticity moves in a non-stationary fashion.

In the statistical analysis we test for cointegration between an industry earnings in one state and the corresponding industry earnings in the nation as a whole. Denote Y_{it} as earnings in the US outside region i for the same industry j . Because each state is a tiny fraction of the US, $\log(Y_{it} + Y_{it}^*) \approx \log Y_{it}^*$ so that cointegration between one state and the balance of the U.S. is equivalent to cointegration between that state and the nation. Conveniently, tests for cointegration do not rest on any exogeneity assumption. The use of the nation as the second region more closely accords with traditional regional economic considerations which we discuss below. Furthermore, it means that all tests for one industry are calculated relative to the same measure.

3. The Shift and Share Model

If region 2 is considered to be the US output of the sector under study rather than this total minus the state, then there are regional economic counterparts to most of these models.

Using the terms of Stevens and Moore (1980) which are derived from Perloff (1969) we define regional growth in earnings in this sector as $\Delta y_{1t} = RG_t$ while national industry growth is $\Delta y_{2t} = IG_t$. The log share in this region is s_t and the regional shift is $RS_t = \Delta s_t$. We have the identity

$$\Delta y_{1t} = \Delta y_{2t} + \Delta s_t$$

or

$$RG = IG + RS$$

Thus the forecasting problem becomes one of forecasting national industry growth and then forecasting the regional shift. In many cases IG is taken to be strongly exogenous and is forecast from some other source so only RS remains.

A very popular assumption favored by Brown(1969), and Hellman(1976) is that the best forecast of the regional shift is zero so that the share is a martingale. This is sometimes called rather misleading the constant share model, but we will call it the martingale share model. Although the forecast value of the share is that it will be the same in the future as it is today, the share is not at all constant. Since the share is therefore an integrated process, it is clear that regional and national output are not cointegrated in this case. A slight variation on this model would allow the share to be a random walk with drift so that a region will consistently grow faster or slower than the nation. Yet another variation is to suppose that the shift follows a simple time series process or perhaps even is itself a random walk.

An alternative simple assumption is that the share will approach its long term historical value in the forecast period. This is an example of an assumption implying cointegration with a unit elasticity. If the long run elasticity is different from one, then the region will be forecast to grow with this constant elasticity relation to the national economy. Any of these models can be given deterministic trends as well. A wide range of more complex models will be estimated and tested, however they all share these long run forecast properties.

4. The Statistical Model

We suppose that the earnings in regions 1 and 2 can be considered to be a

bivariate vector autoregression

$$(5) \quad A^*(B) \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}, \quad E \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \begin{bmatrix} \epsilon_{1t} & \epsilon_{2t} \end{bmatrix} = \Omega^*$$

where B is the backshift operator, $A^*(B) = I - A^*B$ and the ϵ_{it} are white noise processes. This first order case is the simplest to examine and conveniently appears to be the appropriate model for this data set. The process can be stationary, can have one unit root or can have two unit roots. These will be determined by the rank $I - A^*$. Statistically the case of one unit root corresponds to cointegration while two unit roots implies non-cointegration. The model can be parameterized into the difference $y_1 - y_2$ and y_2 . Letting $s_t = y_{1t} - y_{2t}$ which is interpreted as the log share earnings in region 1 as a fraction of region 2, the model becomes

$$(6) \quad \begin{bmatrix} s_t \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} s_{t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

where $A = PA^*P^{-1}$, $u = P\epsilon$

$$Euu' = \Omega = P \Omega^* P', \quad P = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

The rank of $I - A = \text{rank } I - A^*$. If this rank is zero, then both s and y_2 are random walks with correlated innovations; there is consequently no cointegration.

$$(7) \quad \begin{aligned} \Delta s_t &= u_{1t} \\ \Delta y_{2t} &= u_{2t} \end{aligned}$$

If u_1 and u_2 are correlated, a triangular representation may be appropriate for forecasting y_1 conditional on y_2 . This representation may be sensible if y_2 is taken

as strongly exogenous in the sense of Engle, Hendry and Richard (1983) in which case multi-step forecasts of y_2 can be made without forecasting s .

$$(7') \quad \begin{aligned} \Delta s_t &= \beta \Delta y_{2t} + v_t \\ \Delta y_{2t} &= u_{2t} \end{aligned}$$

If $I-A$ has rank 2, then both s and y_2 are stationary.

If the rank is one, there is cointegration, and in the case suggested by the theoretical model, s is the stationary linear combination. Then

$$A = \begin{bmatrix} a_{11} & 0 \\ a_{21} & 1 \end{bmatrix} \text{ with } a_{11} < 1.$$

The model becomes

$$(8) \quad \begin{aligned} s_t &= a_{11} s_{t-1} + u_{1t} \\ \Delta y_{2t} &= a_{21} s_{t-1} + u_{2t} \end{aligned}$$

where the second equation has the familiar error correction form. For conditional forecasting an equivalent representation again introduces Δy_2 into equation (8).

The assumption that y_2 is strongly exogenous would now imply that $a_{21}=0$.

$$(8') \quad s_t = \beta \Delta y_{2t} + a_{11} s_{t-1} + v_t$$

If the cointegrating vector is different from $s = (y_1 - y_2)$, then $I-A$ can be written as $\gamma\alpha'$ or

$$A = I - \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} (1, \alpha)$$

so that the model becomes

$$(9) \quad \begin{aligned} \Delta s_t &= -\gamma_1(s_{t-1} + \alpha y_{2t-1}) + u_{1t} \\ \Delta y_{2t} &= -\gamma_2(s_{t-1} + \alpha y_{2t-1}) + u_{2t} \end{aligned}$$

where the long run equilibrium relation is that $s_t + \alpha y_{2t}$ is stationary or $y_1 + (\alpha-1)y_2$ is stationary. Thus the long run elasticity between earnings in region 1 and region 2 is not 1. Just as before, this system can be triangularized by introducing Δy_2 into the first equations.

$$(9') \quad \Delta s_t = -\gamma_1(s_{t-1} + \alpha y_{2t-1}) + \beta \Delta y_{2t} + u_{1t}$$

Engle and Granger (1987) survey the testing for and estimation of cointegrating systems. The testing is of particular interest because it is not obvious that national and regional earnings are cointegrated. In the language of Engle and Granger a regression of y_{1t} on y_{2t} is called a cointegrating regression because if a long run equilibrium exists such a regression will give a consistent estimate of it. In fact Stock (1986) shows that convergence to the true parameters is extremely fast. The intercept term gives an estimate of log share and the slope coefficient ought to be close to unity if equilibrium shares exist. Granger and Newbold (1974) refer to this regression as "spurious" since inference is so hazardous but Engle and Granger (1987) note that the residuals from the cointegrating regression are convenient for testing for cointegration. If we cannot reject the null hypothesis that the residuals form a random walk sequence, then we infer that there is no linear combination which defines an equilibrium for these variables.

Several statistics are available for testing that there are no unit roots in the cointegrating residuals. For the first-order case, the most convenient is a test that the Durbin-Watson statistic is zero. This was analyzed by Bhargava and Sargan (1983) for fixed regressors and revised for the cointegration case in Engle and Yoo (1986) whose critical value calculation we use here. We use this simpler method because of the large number of cases we examine and because the data support first order models, however in other contexts it is possibly better to use the Engle-Granger(1987) tests based on the augmented Dickey and Fuller (1981) regression.

The possibility of non-cointegration has several important implications for our data:

- (a) There is no long-run equilibrium share; or put another way, there is no equilibrium relationship between national and regional earnings.
- (b) The stochastic process s_t is non-stationary and has infinite variance.
- (c) Shocks to the share s_t are permanent, so that tax or subsidy policies designed to impact particular industries, or local shocks such as environmental regulations or infrastructure investments will have persistent effects.
- (d) The optimal k -step-ahead forecast of s_{t+k} is in fact s_t

The last of these is probably the only one of the four which was specifically desired by those who formulated the shift-share framework.

The idea that key macroeconomic variables are well described by random walks (with the attendant implications) has recently gained currency among macroeconometricians (Nelson and Plosser (1982), Campbell and Mankiw (1986)). The implications are strong and much attention has been devoted to the appropriate testing procedure. The implications for regional forecasting and regional policy making seem especially important.

In the next section we present tests of cointegration for the models.

Following that we estimate specific forecasting models and compute the out-of-sample forecasts for a wide range of states and industries.

5. Results from Cointegration Tests

Quarterly data on US and state earnings in the nine broad industry categories, plus total earnings are examined from 1969:3 to 1981:4. Twenty states are included in the sample. For each state-industry combination two cointegrating regressions were performed, the first unconstrained, the second constraining the slope coefficient at unity. 400 cointegrating regressions were thus performed. With 50 observations the 5% critical value given in Engle and Yoo (1986) is 0.78 for the case where the coefficient is estimated unconstrained. At a 10% size the critical value falls to .69. For the constrained case, the critical values are simply obtained from the critical values of τ_{μ} as reported by Dickey and Fuller and quoted by Engle and Yoo. These are related to the DW critical values by the approximation:

$$DW = \frac{4 t^2 / T}{1 + t^2 / T}$$

This gives critical values for the constrained model of .59 for the 5% and .48 for the 10% test.

The results are presented in Tables 1 and 2.. For each pair of state and sector, two statistics are given. The first is the constrained case, the second is unconstrained. As can be seen, there is little or no evidence of cointegration.

In Table 1, results are presented for non-durable and durable manufacturing. These sectors are typically modelled by regional economists as basic or export oriented sectors and trade theorists would say they produce tradeables. Therefore

these sectors should satisfy the assumption that the relative price of such a good from one region relative to another should have a long run equilibrium value of one by long run purchasing power parity. Often the manufacturing sector is modeled as a linear function of national output in manufacturing. See for example Glickman(1971). Consequently, these sectors should be most likely to cointegrate.

From the table however, it is clear that only a single rejection of the non-cointegration hypothesis is obtained at the 5% level out of 40 constrained and 40 unconstrained tests. This surprising result is borne out in the other sectors as well. Only in Transportation, Communication, and Public Utilities, Services and Government are there any cases of rejections at all. Altogether however, there are only 9 rejections at the 5% level out of 200 unrestricted tests which is probably best interpreted simply as evidence of Type I error. As some of these sectors are local serving and produce non-tradables, perhaps the finding of non-cointegration is less surprising since local demand shocks could destroy cointegration. This is corroborated by the results on total state earnings which again find no evidence of cointegration. Even more interesting is the observation that there are only 5 rejections in the unit elasticity case. This is the case which gives a constant share in the long run. There seems to be no evidence to support such a model.

There are several interpretations of this striking finding. First, it is possible that cointegration within a single homogeneous sector would disappear when aggregates such as durables were examined together. The regional economist is however offered little choice on the ingredients for his accounts and forecasting is typically done from such aggregates. Second, it is possible that the test has little power, partly because there are only 50 observations. Cointegration could be true, yet the tests have so little power that it cannot be detected. This is examined in the next section where forecasting models are constructed with and without imposing cointegration. As is shown there, the cointegrated models are inferior in

out of sample forecasts. Also, the statistical assumptions underlying the test could be incorrect. In particular, the national series could have a more complex process than that embodied in (5) such as having deterministic trends. This would alter the size of the test. Again, however, the forecasting models should then reverse this conclusion.

The overwhelming conclusion is that cointegration does not exist, with the attendant implication that the share is appropriately modeled as a nonstationary process, rather than one converging to some dynamic equilibrium. Put another way, whatever causes shocks, the effects of the shocks seem to be permanent. This conclusion will be of interest to regional policymakers since it dramatically affects the cost benefit analysis of local industrial policy. A program to attract new industry will yield jobs and income over the long run if it yields them in the short run. That is, any increase in the share of output in a sector can be viewed as permanent so that the long run benefits may well outweigh short run costs. By the same analysis, actions in competitor states may permanently reduce industrial shares so that activist policies may be critical. This is in accordance with the game theoretic approach to local policy analysis which is commonly discussed and opposed to the equilibrium assumptions that any interference in the regional marketplace will incur a local welfare loss. In terms of the model of section 2, it appears that the a 's do change over time and the local policy maker should examine the costs of increasing them.

A further implication is that the traditional non-parametric models of martingale share and constant shift are more appropriate models of regional aggregates than might have been expected. We turn now to the comparison of these and other models..

6. Models of Regional Growth

In this section we propose, estimate and compare several models of regional growth in terms of out-of-sample forecasting as well as within-sample fit. Some of these models assume non-cointegration, while others impose cointegration. We consider the cointegrated models as well in order to corroborate our finding that the cointegration model is not a very good model for forecasting. We first comment briefly on each of nine models and then compare their theoretical properties, and how they are nested. Then estimation is discussed, followed by model comparison. For ease of exposition each model is formulated in terms of the regional shift term, $RS_t \equiv \Delta s_t$, however each can easily be recast in terms of $RG_t \equiv \Delta y_{it}$, by adding $IG_t \equiv \Delta y_{US,t}$ to the right hand side. The models are roughly listed by increasing generality. We denote $\epsilon_t = s_t - E_{t-1}s_t$, and the information set is assumed to include national variables dated at t .

Model 1: Martingale Share $E_{t-1}[RS_t] = 0$ or $\Delta s_t = \epsilon_t$ as in (7).

As stated above, this implies that the share is a random walk.

Model 2: Martingale Shift $E_{t-1}[RS_t] = RS_{t-1}$, or $\Delta^2 s_t = \epsilon_t$

Brown (1969) and others examine this model. This shift is a random walk; so that the share itself must be differenced twice to give an unforecastable series. That is, s_t is $I(2)$ which is also a unit root process, so that in this case as well, the non-cointegration of y_{it} and $y_{us,t}$ is implied.

Model 3: Constant Shift: $E_{t-1}[RS_t] = \alpha_0$, or $\Delta s_t = \alpha_0 + \epsilon_t$ as in (7) with an intercept.

The share is a random walk with drift.

Model 4: Autoregressive Shift $E_{t-1}[RS_t] = \beta_0 + \beta_1 RS_{t-1}$, or

$$\Delta s_t = \beta_0 + \beta_1 \Delta s_{t-1} + \epsilon_t$$

This is a generalization of Models 1, 2 and 3. It allows the shift component to be an autoregressive process which converges to a rate of change of $\beta_0/(1-\beta_1)$ (if $|\beta_1| < 1$). It also implies a unit-root model in two lags similar to Model 3 (unless $\beta_1 = 0$)

Model 5: Autoregressive earnings $E_{t-1}[RS_t] = \gamma_0 + \gamma_1 y_{it-1} - IG_t$, or

$$y_{it} = \gamma_0 + (\gamma_1 + 1) y_{it-1} + \epsilon_t$$

This is the only one of our models specifically designed to let the regional economy be independent of the national economy which is a possibility given the earlier results. Because of this, no assumption on cointegration is made. Earnings themselves may be a random walk ($\gamma_1 = 0$) or may converge to $-\gamma_0/\gamma_1$ (if $\gamma_1 < 0$).

Model 6: Constant short-run elasticity $E_{t-1}[RS_t] = \theta_0 + \theta_1 IG_t$, or

$$\Delta s_t = \theta_0 + \theta_1 \Delta y_{US,t} + \epsilon_t$$

This is the first model to allow national earnings to have a non-unit impact elasticity on regional earnings, (since $RG_t = \theta_0 + (\theta_1 + 1) IG_t$) and is an example of equation (7') with an intercept. Since the error-correction term s_{t-1} is missing, the long-run elasticity is undefined, and so this model is in the non-cointegrated unit-root class of models 1 through 4. It may be important to allow non-unit elasticity in the short run (as do 8 and 9 as well as this model) and in the long run (in 9) since there is a small amount of evidence in the cointegration results to suggest that the non-unitary response is preferred.

Model 7: Autoregressive share $E_{t-1}[RS_t] = \delta_0 + \delta_1 s_{t-1}$, or

$$\Delta s_t = \delta_0 + \delta_1 s_{t-1} + \epsilon_t$$

The model is (8) and can be given as $s_t = \delta_0 + (\delta_1 + 1) s_{t-1}$ so that if $\delta_1 < 0$, the equation forces the long run equilibrium share to be $-\delta_0/\delta_1$. This model therefore implies cointegration between regional and US earnings with a unit long run elasticity (unless $\delta_1 = 0$, in which case we have model 2).

Model 8: Unit Elastic Error Correction Model

$$E_{t-1}[RS_t] = \pi_0 + \pi_1 IG_t + \pi_2 s_{t-1}$$

$$\Delta s_t = \pi_0 + \pi_1 \Delta y_{US,t} + \pi_2 s_{t-1} + \epsilon_t$$

The unit elasticity error correction model given in (8') nests most of the models above, since it implies cointegration with a unit long run elasticity (if $\pi_2 < 0$) and allows the short-run elasticity to be non-unitary like model 7 (if $\pi_1 \neq 0$).

Model 9: Error Correction Model (9')

$$E_{t-1}[RS_t] = \psi_0 + \psi_1 y_{it-1} + \psi_2 IG_t + \psi_3 s_{t-1}, \text{ or}$$

$$\Delta s_t = \psi_0 + (\psi_1 + \psi_3) \left(y_{it-1} - \frac{\psi_3}{\psi_1 + \psi_3} y_{US,t-1} \right) + \psi_2 \Delta y_{US,t} + \epsilon_t$$

This can be rewritten as the first equation of a bivariate single lag vector autoregression of regional on national earnings; it generalizes the unit elasticity error correction model in 8 to allow non-unit long run elasticity which is now given by $\psi_3/(\psi_1 + \psi_3)$. That is, some kind of increasing or decreasing comparative advantage for this region is allowed. But unlike the disequilibrium models 1 and 2, the change in the long run is systematic and predictable.

The discussion above makes clear the way the models are nested. Note further that Model 9 nests all of the models except 2 and 4 since these two use a second lag of earnings. Also note that the "equilibrium" models 7, 8 and 9, nest unit root models under certain restrictions, hence it is appropriate to include them in model comparison even though the results of Section 5 indicate that they might

be inappropriate. In any case, since comparison of forecast performance is a major tool used below, it seemed sensible to include equilibrium models in the comparison, in case the data generation process radically changed in the post-sample forecast period.

Each of the models is estimated for each of the state-industry combinations. The data set is the same as that used in Section 5. The estimation results are summarized in Table 3 in terms of the geometric means across states and industries of the in-sample standard errors of regression. Larger models fit somewhat better with model 9 as the best followed by 8. These cointegrating models have 4 and 3 parameters respectively. Close behind are models 6 and 4 which are non-cointegrating models with 2 parameters each. Models 1, 3 and 7 were in the next group zero one and two parameters respectively, and finally 2 and 5 as the worst. In terms of standard errors, the cointegrating models fit the data slightly better but at the cost of additional parameters. This cost is most noticeable in forecasting.

We now turn to the forecasting ability of the nine models. We emphasize this aspect of model comparison, since much of the debate surrounding various models of shift and share concerned forecasting performance. (Again, see Stevens and Moore (1980) and James and Hughes (1973).) In each case we compare forecasts of y_{it} , but since we directly estimate RS_t , in order to convert to the variable of interest we add $(IG_t + y_{it-1})$ to the model's forecast of RS_t . This is not a problem if forecasts can be made conditional on IG_t , however unconditional forecasts will need estimates of IG_t terms. Forecasts are computed both conditionally and unconditionally, each under two settings. First, ten post-sample one step forecasts are made — at each step all available information is used. In the other scenario, one-through ten-step ahead forecasts are undertaken. These forecasts are made using actual data through the estimation period and forecasts as they are needed

through the multistep forecasts, which is the usual situation facing the forecaster. Conditional forecasts are made using actual values of $y_{US,t}$ and IG_t . Additionally, actual values of $y_{US,t}$ are also used in the calculation of s in multi-step forecasts. Unconditional forecasts are made using one-step and multi-step forecasts of $y_{US,t}$ in the two scenarios as appropriate. These latter forecasts of $y_{US,t}$ are estimated via first through fourth order autoregression model of IG_t and using the Schwarz (1978) criterion to pick the best model. The first-order model was usually selected.

Each state/industry/model combination was used in each of the four forecast regimes, with forecast errors constructed as actual y_{it} less forecasted y_{it} . Root mean square errors for the ten post-sample observations were calculated in each case, and geometric means across and states and industry were calculated in order to compare forecasts across models. The means were then normalized by dividing each by the root mean square forecast error of the best model for each state/industry combination. Thus, the figures in Table 4, which give the results for this exercise, shows how much worse one does on average by picking the indicated model rather than the one which happens to be "best." The standardization by the "best" model does not affect any comparisons across models.

While Table 3 indicates that within the estimation period, the least constrained, most parameterized models have the best fit in terms of standard error, this result definitely does not hold in the post-sample period. This is consistent with both theoretical and empirical work indicating the value of parsimony in forecasting and the tendency for a minimum standard error criterion to select over-parameterized models. The specific pattern of performance shows the value of using cointegration tests as part of an overall modeling strategy, since the results on forecasting ability now come as no surprise. Model 1, the martingale share model, does best in one step conditional forecasts even though it is among the most constrained of our models. It outperforms the more sophisticated, but equilibrium

based models (such as Model 8). Model 4 does almost as well and the geometric means of 3, 4, 6 and 7 are relatively close. Model 2, the martingale shift model does very poorly. Although martingale shift is obviously a poor assumption for the long-term, our forecast horizon is only two and a-half years so this maybe somewhat surprising.

The general flavor of these results carries over into the other three forecasting scenarios. In the one step unconditional forecasts, models 1 and 4 do very well, followed closely by 3 and 6. Again, the equilibrium models 7, 8 and 9 do somewhat worse. In the one through ten step ahead conditional scenario, model 1 seems to do substantially better than the others, while with unconditional forecasts, model 3 and model 6 do slightly better than model 1.

Similar exercises were carried out separately across states and across industries to see if individual states or individual industries were different. The states were also divided into fast growth, average growth, and slow growth categories and results computed for these separately. The results give no surprises. Models 1, 3 and 4 were always the leaders in forecasting ability, with none appearing as a clear-cut preference in all cases. In two industry cases (durables, finance) Models 6 and 7 also performed well.

The value of parsimony in selecting forecasting models is clear. The results were then reanalyzed using the AIC or Akaike Information Criterion proposed by Akaike(1974) and BIC or Bayesian Information Criterion introduced by Schwarz(1978) which provides a higher penalty for each parameter. The models selected by these criteria are better on average than a random selection of the models and the BIC is very close to selecting the best model. Further details are in Brown(1986) and for another data set in Engle and Brown(1988)and Brown and Engle(1986).

To summarize the results of this section, we find that models with

low-parameterizations, particularly the martingale share, and constant and autoregressive shift models, out-perform less constrained models in post-sample forecasting. This is not surprising for two related reasons. First, the results of our cointegration tests indicate that regional share should be modeled with a unit root in its dynamic process, and it happens that the simpler models impose this constraint. Second, examination of regional shifts across time indicate a great amount of instability in this value, and highly parameterized models often fail to forecast well in such circumstances. Thus, the martingale share, and to a lesser extent, the constant shift models, which are both low parameter, random walk constructions, forecast significantly better than the alternative models considered here.

7. Conclusions

"Traditional" models of regional share typically involve a unit root in the stochastic process which defines the share. The "martingale share" model is the prime example of this. That the share follows a random walk may be thought of as a result of the persistence of location decisions by the factors of production and the cointegration results of Section 5 bear this out. National and regional earnings do not cointegrate, hence they have no equilibrium relationship. In Section 6 we summarize the results of an exhaustive comparison of several models of shift and share, mainly via their forecasting ability. We compare the performance of these models directly. Again, as might have been expected given the non-cointegration results, those models which imposed unit root relationships on the share, even though they performed less well in-sample due to the low parameterization, proved far superior in terms of forecasting ability. In fact, the simplest one, the martingale share model performed so well overall, that due both to its simplicity, ease of computation — and its ability to forecast, it becomes highly recommended, both as

a forecasting device, where the current share is the best forecast of future share, and as a description of the temporal process of the share.

These simple conclusions may appear to contradict findings of Norrbin and Schlagenhauf(1988) for the US and Altonji and Ham(1985) for Canada, that find the bulk of the variability in regional output is due to aggregate or national shocks. In fact our very simple martingale share model can be rewritten as

$$\Delta y_{it} = \Delta y_{US,t} + \epsilon_{it}$$

which decomposes the regional changes into national and regional causes. The model makes no prediction as to which source of shocks is greater. It merely says that both permanently change regional output. If the system were cointegrated, the regional shocks would not be permanent and might then be even less important, particularly for long horizon forecasting. Thus, these results are fully consistent with ours.

REFERENCES

- Altonji, J.G, and J.C. Ham, "Employment Variation in Canada: The Role of External, National, Regional and Industrial Factors," unpublished manuscript, Department of Economics, Columbia University, 1985
- Akaike, H. (1974), "A New Look at Statistical Model Identification", *IEEE Trans. Auto. Control*,19, pp716-723
- Bhargava, A. and Sargan, J.D. (1983) "On Testing Residuals for a Gaussian Random Walk," *Econometrica*.
- Blanchard, O. and Summers L. (1986) "Hysteresis and the European Unemployment Problem," in (ed. Fischer) *Macroeconomics Annual* NewYork: National Bureau of Economic Research.
- Brown, H. (1969) "Shift-Share Projections of Regional Growth," *Journal of Regional Science*, 9, 1-18.
- Brown, S.J. (1986), "Post Sample Forecasting Comparisons and Model Selection Procedures," unpublished PhD. Dissertation, UCSD
- Brown, S.J. and R.F.Engle,(1985), "Model Selection for Forecasting", *Journal of Computation in Statistics*,
- Davidson, J. et al (1978) "Econometric Modelling of the Aggregate Time- Series Relationship Between Consumer Expenditures and Income in the United Kingdom," *Economic Journal*, 88, 661-692.
- Dickey, D. and W. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of American Statistical Association*, 74, 427-431.
- Engle, R.F. S.J.Brown,and G. Stern,(1988)"A Comparison of Adaptive Structural Forecasting Methods for Electricity Sales", *Journal of Forecasting*,pp149-172
- Engle, R. and C. Granger (1987) "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, pp251-276
- Engle, R.F., D.F. Hendry, and J.F. Richard,(1983), "Exogeneity", *Econometrica*, pp277-304
- Engle, R. and S. Yoo (1987) "Forecasting and Testing in Cointegrated Systems," *Journal of Econometrics* 35, pp 143-159.
- Glickman,N (1971), "An Econometric Forecasting Model for the Philadelphia Region", *Journal of Regional Science*, 11,1
- Granger, C. and P. Newbold (1974) "Spurious Regressions in Econometrics," *Journal of Econometrics*, 26, 1045-1066.
- Hellman, D. (1978) "Shift-Share Models as Predictive Tools," *Growth and Change*, 7, 3-8.
- James, F. and J. Hughes (1973) "A Test of Shift and Share Analysis a Predictive

Device," *Journal of Regional Science*, 223–231.

Nelson, C. and C. Plosser (1982) "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139–162.

Norrbin, S.C. and D.E. Schlagenhauf, (1987), "An Enquiry into the Sources of Macroeconomic Fluctuations", manuscript, Arizona State University

Perloff, H.S., E.S. Dunn, Jr., E.E. Lampard, and R.F. Muth, (1960), *Regions Resources and Economic Growth*, University of Nebraska Press, Lincoln Nebraska

Schwarz, G. (1978) "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Stevens, B. and C. Moore (1980) "A Critical Analysis of the Literature on Shift–Share as a Forecasting Technique," *Journal of Regional Service*, 20, 419–438.

Stock, J. (1987) "Asymptotic Properties of Least Squares Estimations of Cointegrating Vectors," *Econometrica*, 55, pp 1035–1056.

TABLE 1
TESTS FOR COINTEGRATION
DW Statistics of v from Cointegrating Regressions:

$$a) y_{it} - y_{US,t} = \alpha + v_t$$

$$b) y_{it} = \alpha + \beta y_{US,t} + v_t$$

STATE	NDUR(a)	NDUR(b)	DUR(a)	DUR(b)
AL	.092	.194	.216	.547
AZ	.019	.132	.025	.173
CA	.026	.197	.072	.159
CO	.059	.272	.026	.177
CT	.193	.195	.117	.135
FL	.040	.207	.052	.251
GA	.110	.333	.164	.186
IL	.042	.168	.062	.211
LA	.020	.228	.039	.087
MI	.189	.195	.276	.305
MO	.208	.436	.143	.258
NC	.401	.434	.062	.212
NY	.006	.110	.019	.175
OH	.105	.235	.061	.236
OR	.095	.262	.094	.245
PA	.016	.135	.060	.291
TX	.010	.158	.018	.155
VT	<u>.646</u>	.653	.046	.188
WA	.078	.309	.064	.226
WI	.114	.367	.309	.333

Note: **Boldface** are significant at 10% level
Boldface and Underscore are significant at 5%

TABLE 2

DW Statistics from Cointegrating Regressions
Other Sectors
Constrained/Unconstrained

	TPU	WT	RT	FIR	SERV	GOV	OTHR	TOT
AL	.075/.646	.043/.06	.044/.101	.091/.103	.156/.535	.112/.216	.107/.120	.051/.121
AZ	.029/.453	.031/.346	.014/.129	.055/.258	.025/.531	.020/.50	.134/.191	.016/.156
CA	.681/.688	.070/.675	.035/.112	.071/.343	.048/.273	.354/ <u>.976</u>	.175/.529	.028/.107
CO	.025/.265	.072/.332	.015/.161	.040/.379	.016/.256	.075/.428	.038/.372	.011/.138
CT	<u>.615/1.20</u>	.057/.372	.032/.166	.185/.268	.081/.606	.155/.303	.019/.107	.034/.093
FL	.052/.251	.084/.175	.023/.109	.077/.54	.027/.102	.040/.407	.075/.075	.026/.081
GA	.021/.401	.029/.341	.078/.192	.071/.072	.091/.152	.069/.104	.115/.129	.054/.245
IL	.095/.499	.067/.372	.039/.167	.260/.366	.105/.544	.218/.371	.063/.383	.029/.59
LA	.034/.170	.025/.162	.011/.126	.036/.146	.045/.481	.198/.249	.025/.186	.013/.122
MI	.143/.309	.030/.171	.056/.096	.101/.222	.157/.204	<u>.658/.799</u>	.054/.133	.136/.206
MO	<u>.786/.925</u>	.057/.284	.043/.86	.049/.282	<u>.782/1.53</u>	.104/.618	.222/.579	.046/.253
NC	.066/.279	.208/.237	.105/.623	.081/.117	.465/.466	.165/.556	.086/.138	.200/.273
NY	.029/.570	.007/.407	.008/.156	.030/.128	.008/.197	.022/.076	.011/.135	.007/.135
OH	.048/.240	.027/.155	.038/.126	.046/.210	.067/.761	.161/.415	.082/.429	.034/.183
OR	.378/ <u>1.24</u>	.051/.260	.017/.201	.053/.192	.023/.121	.025/.300	.163/.311	.028/.133
PA	.033/.258	.016/.316	.014/.091	.057/.576	.043/.686	.148/.267	.076/.585	.011/.173
TX	.012/.186	.010/.136	.007/.113	.023/.142	.011/.140	.041/.145	.013/.174	.009/.117
VT	.136/.428	.159/.643	.130/.256	.137/.261	.160/ <u>1.55</u>	.259/.279	.059/.171	.160/.260
WA	.108/.124	.026/.103	.014/.087	.058/.127	.02/.645	.060/.064	.067/.342	.019/.096
WI	.458/ <u>.886</u>	.139/.211	.068/.185	.171/.248	.095/.308	<u>1.10/1.10</u>	.094/.159	.295/.296

Note: **Boldface** are significant at 10% level

Boldface and Underscore are significant at 5%

TABLE 3
GEOMETRIC MEANS OF IN - SAMPLE RMSE's

(over states and industries)

MODEL	GEO.MEAN
1	1.35%
2	1.88%
3	1.30%
4	1.25%
5	1.74%
6	1.27%
7	1.30%
8	1.23%
9	1.18%

TABLE 4
OUT-OF-SAMPLE RMSE's
GEOMETRIC MEANS OVER STATES AND INDUSTRIES

MODEL	One-Step Unconditional	One-Step Conditional	One-through Ten Step Conditional	One through Ten Step Unconditional
1	1.1119	1.0968	1.5202	1.5680
2	1.5619	1.4147	3.0638	2.3853
3	1.1352	1.1055	1.6508	1.5353
4	1.1132	1.0944	1.6359	1.5848
5	1.5690	1.2261	2.7244	1.6798
6	1.1435	1.1083	1.7426	1.5412
7	1.1855	1.1417	1.7934	1.6570
8	1.2075	1.1484	1.9819	1.6654
9	1.2754	1.2063	1.9311	1.7154
ave	1.3160	1.1967	1.5530	1.9790