DATA LIMITATIONS AND ALTERNATIVE METHODOLOGY IN ESTIMATING REGIONAL ECONOMETRIC MODELS#

DONALD RATAJCZAK*

Econometricians engaged in regional research have continually been confronted with the choice of adapting their models to the available data or constructing data to appropriately specify their models. On the one extreme, econometric techniques have been used to "explain" the historical performance of available data series. Accounting identities are not used, nor is a consistent economic system specified. [See, Roberts and Wittels (31), and Burton and Dyckman (7).] At the other extreme is the attempt to create subsystem "Keynesian" expenditures models even if the expenditures data are unavailable. [Klein (18) and L'Esperance, Nestel, and Fromm (19) provide the strongest arguments for this approach.] Some mixture of expenditures, production, economic base activity, and interregional factor flows characterize most modelling activity between these two extremes. Because of data limitations, specification problems confront the regional econometric model builder, and measurement problems also may be substantial.

The trade-off between data construction and model modification varies with the economic system that is used to describe regional economic activity. Some data, such as gross regional product, may be necessary in nearly all specifications of regional economic systems, while other information, like personal consumption expenditures, may not be missed by some model builders. This paper will discuss the data that is available, a sampling of the various descriptions of regional economic systems that have been used in regional econometric models, and the apparent trade-offs that have been made.

A. *Data Availability*

Although substantial data exists on the regional level, most information is generated on a once-only or on a census year basis. While such information may provide *a priori* material for an econometric model, such as with intercounty work-force migrations, they do not contain sufficient frequency to parameterize econometric models. The raw materials needed to construct regional models are consistent time-series of economic statistics. In some areas, such as insured labor markets, data are plentiful; in others, such as government revenues and expenditures, consistency may be a problem, but numbers are available; while in still others, such as interregional trade patterns, very little information exists.

#Presented at the Southern Regional Science Association Meeting, Washington, D.C., April 1974.
*Department of Economics, Georgia State University, Atlanta , Georgia.

Most regional models have used annual data. This frequency permits the use of manufacturing, mining, and farming data[2] that only appear annually, state and local receipts and expenditures, some which are only available annually (or on a fiscal year basis), and personal income or other statistics relying upon income tax data (although quarterly distributions of personal income are now available for states from the mid-1950's). If the regional model uses "basic" industries, dependent upon national activity, to explain nonbasic activity, and if most basic activity is concentrated in manufacturing, mining, and farming, the enriched data that is available on an annual basis may substantially increase the theoretical content of the regional models [e.g. Bell (3) and Glickman (12)].

Unfortunately, annual data do not provide many degrees of freedom for statistical estimation. Interindustry interactions, and other explanatory variables are limited, while the statistical reliability of the coefficients must be substantially less than would normally be acceptable. Furthermore, the hypothesis that regional differences in response to national stimuli are partially caused by differing speeds of adjustment cannot be effectively tested with annual data. Annual data also obscure significant short-run changes in regional economic activity which may be responsible for longer-term migratory movements. Finally, the wider range of information that is available with annual data is somewhat misleading, for the mining and manufacturing data become available two or more years after the event. The loss of degrees of freedom and the resulting restriction upon model specification may be more significant than the gains that are achieved from the increased number of available statistical series with annual frequencies.

At the state level substantial data are now available on a quarterly or even monthly basis. The most complete and consistent data are for the insured labor force. Generally, weekly earnings and hourly wage rates are available by industry as well as those employed and the level and duration of insured unemployment. Government, agricultural, and the self-employed worker may not be covered. Household surveys are not valid at the state level, so new entrants to the labor force are not counted until employed. In addition, the employee size of establishments subject to filing reports has periodically changed—most notably in 1962 and 1973. A more complete estimate of total labor force activity is also generated by many regions by using Department of Agriculture sources to establish farm employment and by using Census benchmarks augmented by tax return information to determine the non-wage and salary employment. The labor force is then derived as a residual, dependent upon the previous methods to establish employment and upon adjustments to the insured unemployment. Only rarely are household surveys used to establish labor force characteristics. Although regional Labor Department offices are attempting to develop procedures that will insure compatability between the sum of all regional employment estimates and the national estimates, comparability with national estimates is not guaranteed by the above procedures.

The quarterly personal income series which are now available at the state level can be summed to national aggregates, while the employment figures of the insured labor force provide most of the wage and salary information. The remaining personal income components ultimately are determined by tax information, but their quarterly profile largely depends upon national activity (although a Hurricane Agnes or a Mississippi flood may lead to substantial

changes in regional property incomes for a short period of time). Notwithstanding the heavy dependence of the quarterly profile of property income upon national activity, the personal income series, especially in terms of labor income, provides reasonable quarterly information on gross household income.

Since the foundations of the personal income figures are tax returns, which are on a household basis, differences between purchasing power and income generated can be caused by commuting patterns. These considerations are more relevant at a county than at a state level, but they are not absent from many states. Therefore, some knowledge of labor service exports and imports are required to effectively connect regional earnings with regional purchasing power. Currently, this information is being generated by Census commuting studies, which are then adjusted locally to reflect transportation flows, differing participation rates in terms of regional populations and labor forces in contiguous areas, or some other procedures to make intercensus adjustments.[3]

The tax return support for regional data also permits accurate household income measures only annually at the sub-state level. However, the insured labor force data are available at the county level. Therefore, the wage and salary component could be constructed quarterly by comparing changes in weekly earnings by industry locally to the changes at the state level and then using these results to alter the ratio between the local and state earnings per worker. The remaining components of local personal income could then be created quarterly in terms of their annual relationships to the state. While such a procedure would not preserve the same informational content as contained in the quarterly personal income at the state level, it would add significant information on local wage and salaries earnings and permit quarterly analysis on household activity at the local level.[4]

Of course, property income available to residents in a region has little relationship to capital earnings in that region. A large portion of rents and a significant percentage of interest may be locally generated, but returns from capital exports cannot be extracted from the property income statistics. While these data limitations may not reduce the ability to estimate gross household purchasing power (in fact, the use of national variables to explain these components may increase the predictability of the models), they reduce the ability to explain the relationship between productive activity and expenditures within a region.

In order to get disposable income, personal taxes at all government levels and gift and estate taxes at all levels must be known. Most states now have revenue data on a receipts basis and many of them also publish the annual tax payments to the federal government. (The annual census publication on revenues and expenditures for states and local governments also provides estimates.) Although quarterly estimates of federal tax payments must be calculated, reasonable estimates of disposable income can be derived.

In those states using sales taxes, taxable sales are usually available by type of retail outlet on a monthly or quarterly basis. Many times, the reporting agencies are by county, so substantial sales information can be generated at that level. However, such sales reflect tourist purchases and purchases from residents in other counties as well as local exhaustion of purchasing power. In addition, some

local purchasing power may be spent in other counties or states. Therefore, taxable sales exclude the imports of consumer goods by local residents and include the export sales to non-local consumers. Also, many states do not provide full coverage of all consumption items, so information on many expenditures may be excluded. Taxable sales may approximate retail sales in many categories, but the relationship between retail sales and consumption expenditures becomes much less clear as the region increases its interactions at the consumer level with other regions. Taxable sales therefore are more useful in providing a basis for explaining retail employment and sales tax revenues than for serving as an instrument for consumption expenditures. (Of course, if we can assume that net exports are zero, e.g. the taxable sales generated by tourists are exactly offset by the tourist expenditures of local residents elsewhere, then the data will identify regional consumption. It is unlikely, however, that this assumption would be valid for each class of retail establishment.)

Most parts of the country now require construction permits before private construction expenditures can occur. Permits are related to construction activity, which in turn is related to investment in housing and structures. Furthermore, the final site expenditures are local (although prefabricated and materials imports may be a substantial portion of the expenditures). However, permits may substantially differ from starts at any period of time (perhaps because of changes in credit or market conditions between the time of issuing permits and the planned start time or because regulatory changes may alter the desired timing between procuring permits and beginning activity). These reasons, along with changes in labor market conditions (e.g. strikes may substantially delay projects) may also alter the relationship between starts and activity. Public construction, while more expenditures oriented than the above, may be measured in terms of outlays rather than activity. Under normal circumstances, the construction permits data can be used as an instrument for construction expenditures, but the above exceptions must be considered in any analysis that uses this data. Furthermore, the expenditures may differ greatly from the value added in regional construction if substantial prefabrication or materials are imported.

Although labor markets and household income are well represented in the data, barring the earned versus received issues; government revenues and expenditures can be determined, excluding the difficulty of allocating federal expenditures to regions;[5] and taxable sales and construction permits may permit the creation of series for consumption and construction expenditures, several significant expenditures items cannot be measured. Inventory investment cannot be determined at a regional level except during that one period of the year when inventory taxes may be owed. Non-construction investment is unknown except for manufacturing, mining, and agriculture, and they are only available well after the year has passed. (Some information may be obtained on investment from the tax returns of intra-regional business, but the assignment of investment to a specific region from interregional business cannot be made from tax sources.)[6] In addition, no information is available for net exports of goods and services.

Location quotients could be used to determine exports and imports, but the assumed constancy of tastes and production techniques are unrealistic. In addition, the measuring device usually is household income (even tax differences are not considered), when corporate or government activity may be most

significant in explaining concentrations. These problems could be compounded if only certain industries are designated as export oriented.

The measurement of non-construction corporate activity is even more difficult to achieve. Generally, this sector is largely ignored (except for manufacturing) while a gross regional product measure is used in its place. [If net exports have been determined, corporate activity then can be used as a residual, as in Bell (3).] The most popular method of obtaining gross regional product uses the labor market information to specify industry wage bills and then uses national distributions by industry to convert the wage bill to value added. Where wages per hour of work are substantially different between regions, suggesting that non-wage inputs differ substantially, the value added can be adjusted by these earnings differences.[7] [See Jaycox and Kendricks (17).] Assuming all firms have the same production relationship, this technique will establish a meaningful measure of regional economic activity. However, regional differences other than those already specified by labor market information are limited to different compositions of industries having differing rates of wage bills to total value added and perhaps to different labor-capital intensities between regions (although the latter effect would only change the value added estimates if the underlying production relationship did not have a constant elasticity of substitution). The differing ages of capital and their cyclical responses to aggregate activity would not be incorporated.[8] In effect, the measurement assumes away those cases of interregional industry location where an industry will move because of different returns to capital elsewhere (or else it is assuming that such moves are done instantaneously). [See Borts (6).]

An alternative procedure is to build-up state product in terms of earnings by use of tax information. Many issues are raised by this method. The most important of these is whether regional product should be domestic (as in the above) or total. The former would measure only the production of a region while the latter would include the returns from the net export of capital and labor services. Domestic product can be used to measure different production capabilities (if correctly constructed) while the total product measures regional economic purchasing power. If the domestic product is used, household incomes must be adjusted to an earned basis. While the adjustment "only" requires commuter information for labor services, it requires knowledge that is unknown for property income. Issues concerning the assignment of corporate headquarters activity also must be resolved. No adjustments to household income are needed for total product measures, but the ownership and measurement of corporate retained earnings must be resolved.[9] More significantly, tax information is incomplete on depreciation. However, tax information provides reasonable estimates of cyclical changes in the earnings of the corporate sector, and it permits the relationship between the wage and non-wage components of value added to vary.

Little has been said about regional financial markets and regional prices. Some price statistics on consumer goods are available by SMSA and some trade journals provide price information by region for their specialties, most notably in construction. However, product prices are generally unavailable on a regional basis. This absence seriously hampers the ability to relate production with expenditures and to explain interregional migrations. Regional wholesale price indices are sorely needed.

At the present, regional flow-of-funds information is lacking. However, the Federal Reserve districts and the Federal Home Loan Bank districts generate information on bank conditions. In addition, some differences in mortgage rates by region do exist and are published. Differences may also exist in interest available to savings accounts, although most financial "prices" are determined by national markets. Currently, the financial data which is available has not been fully exploited. No major regional model has yet integrated the banking information with regional expenditures. In addition, the significant interregional capital flows are not yet measured. It is clear, however, that information is becoming increasingly available (although insurance companies, foreign capital, and financial service corporation activity may continue to cause problems), and the next extension of regional analysis may well concentrate on the importance these capital flows have upon regional development.

This brief sketch of the available data indicates that substantial difficulties still exist in constructing a statistical library to explain regional economic activity. However, some information is available historically for a significant period of time. Labor market information and household incomes are the most available data, but government revenues and expenditures, taxable sales, construction permits, and bank conditions are also accessible at the regional level. If annual data are used, inputs and outputs for manufacturing, mining, and agriculture also can be used. Unfortunately, substantial data construction is necessary to define the regional corporate sector, while undesirable procedures may be needed to determine the net exports of a region. Although total expenditures may be determined by these methods, expenditures by type of activity are still unavailable at the regional level. In addition, the range of price information by region is almost unworkably small. Thus, considerable data problems persist at the regional level, but sufficient information is available to provide some reasonable, albeit incomplete, descriptions of the regional economic system.

B. *Econometric Specifications*

Although data availability on a historical basis is important to econometric analyses, the model specification is paramount. Once a description of the regional system is constructed, the data then parameterizes the model by use of some statistical technique.[10]

Regional econometric models can usefully be considered in three groupings: an explanation of prevailing data by whatever variables are available [both Burton and Dyckman (7) and Roberts and Wittels (31) probably belong in this grouping]; strict conformity with specified theory even if substantial data must be constructed [the most significant holders of this view are Klein (18), L'Esperance, Nestel, and Fromm (19), and Moody, Puffer, and Williams (22)]; and some implicit tradeoffs between theory and data so that measurement error and specification error are jointly minimized [Glickman (12), (13), Ratajczak (28), and Crow (9) exemplify this group]. A brief outline of the methodology and shortcomings of each will be presented.

Those who permit the data to determine the structure of their model usually are seeking maximization of explained variation for each data series. Although

theory will be used to determine the acceptable structure of each equation, the model is merely a set of single equation estimations. For example, regional property income may be explained by national property income and the previous level of property income. If sufficient data are available to establish all the components of an accounting identity, the identity may then be expressed as the sum of its components, but no attempt is made to insure reasonable multipliers by fully specifying the interaction of production, expenditures, and wages.

Since little or no data is constructed, no measurement error is introduced aside from the inherent errors in the published data. However, specification error could be substantial. Personal income may be used to explain employment even though industry output is not consumed locally. As long as regional personal income is growing in conformity with the true purchasing power responsible for industrial activity, it is serving as an instrument and may explain employment very well. When these incomes differ, however, the model may create substantial errors. Furthermore, errors can easily be made in analyzing the impact of policy activity with the use of such models. For example, a model that relates coal activity in West Virginia to West Virginia personal income may perform well, but such a model would suggest that an increase in transfer payments to the region could substantially increase mining employment. Although such models may provide useful values for revenue projections or planning purposes, their results are valid only so long as the instruments maintain their relationship to the true variables which they represent. Obviously, such limitations diminish the usefulness of these models in signalling when deviations from previous experience will occur.

On the reverse extreme is the attempt to construct regional support for national expenditures models. If taxable sales are not available or are so obviously not relevant to consumption expenditures that they cannot be used, consumption expenditures may be created by assuming constant savings between regions. A measure of consumption expenditures is then constructed but at the cost of losing the ability to explain regional differences in terms of differential savings (or taxes). Net exports can also be derived by use of location quotients, but at the expense of losing explanatory power for examining differing regional tastes or production relationships. Investment activity can be evaluated as a residual after deriving gross regional product in terms of a mark-up on wage bills, but differing production techniques by region cannot be analyzed. In short, the necessary data creation is so substantial and the assumptions necessary to develop such information are so limiting that many of the factors that may cause differences in regional development cannot be analyzed.

The costs incurred in following an expenditures approach to regional economic analysis are so substantial in terms of data development that the benefits should be outlined to balance the analysis. The most substantial of these gains is the ability to use a well developed behavioral theory in regional analysis. Nearly as important is the ability to use national models as inputs into regional analysis. Once expenditures data have been developed, the investment series can be used to derive regional accelerator relationships. Thus, regions can have different investment cycles. They can also exhibit different responsiveness to national variables (although the constancy of the savings rate between regions

should not be violated). More importantly, changes in national policy can be allocated to regional changes because the regional models conform to national prototypes. As long as the national changes affect the region more than the region's internal activity, the expenditures approach will improve the degree of anticipating regional changes.[11]

Some classical regional questions cannot be confronted by the model. There exists no reason for labor and capital to migrate, for regions really only differ in terms of their investment cycles and income elasticities. These are cyclical in nature and cannot explain the secular changes that are observed. Variations in natural endowments and transportation characteristics are not examined. Also, the location quotient approach does not measure the gains to purchasing power from exported capital. Retirement areas may be growing even though their exports of merchandise and their labor incomes are unchanged. The costs incurred in using an expenditures model at the regional level may well exceed the benefits derived from such a framework.

Regional analysis has concentrated on employment rather than expenditures because of the obvious data advantages of doing so. Little is really lost by beginning the analysis with initial changes in employment rather than in expenditures and solving for total employment changes rather than total expenditures changes. (Actually, changes in expenditures that are met by inventory changes and changes in the production relationship used to meet those expenditures are excluded.) Gains include the ability to use data that are readily available. However, a theory that easily explains why employment by industry changes or how regional employment changes when national expenditures change has not been adequately developed. For example, the effects of a reduction in food expenditures upon retail employment would be difficult to explain even at the national level. Although input-output procedures can provide an approximation, the technique cannot then be used to explain how all the changes in derived employment will lead to further changes in food consumption and in other activity. Yet, some method of explaining regional labor market activity holds substantially more promise than explaining expenditures.

The demand for labor becomes the pivotal point of analysis. If capital stock is available, several production relationships under varying equilibrium conditions can be used to determine labor demand. In the more normal case where capital stock data is unknown, a CES production relationship may be used under assumed profit maximization conditions as follows:

$$1)\quad Q = Ae^{\lambda t}\{SK^{-p} + (1 - S)L^{-p}\}^{-1/p}$$

where $Ae^{\lambda t}$ is an index of technological competence at each point in time, S is a distribution weighting of income between capital and labor, $1/1 + p$ measures the degree of substitution between capital and labor, and Q, K, and L are the respective output and capital and labor inputs in each industry.

$$2)\quad \frac{dQ}{dL} = A^{-p}e^{-\lambda t}(1 - S)\{Q/L\}^{p+1},$$

assuming profit maximization,

3) $^{w}/p = {}^{dQ}/dL,$

taking logarithmic transformations and solving for labor

4) $$\text{Log } L = \frac{-p}{p+1} \log A + \frac{1}{p+1} \log (1 - S) - \frac{p}{p+1} \lambda t$$

$$- \frac{1}{p+1} \log(^{w}/p) + \log Q$$

or

5) $$\text{Log } L = a + b_1 t + b_2 \log (^{w}/p) + \log Q$$

where $b_2 < o,$
and $b_1 < o$
if $b_2 > -1.$

Actually, few models maintain this structure of the demand relationship and most include a lagged adjustment term. In addition, many exclude the substitution term either because they believe the wage-price relationship is unchanging or because data on product prices are unavailable at the regional level.

The quantity of labor and wages are known at the regional level, but product prices and industry outputs are unavailable. Several models [Crow (9) and Glickman (12)] have used the Kendricks-Jaycox method to construct outputs and then used an array of variables to explain the outputs. If output data can be used, it may increase the flexibility of estimating the production relationship, e.g. the homogeneity factor could be other than one, or provide an adjustment factor between output and labor to reflect inventory activity.[12] Therefore, a cost is borne if output estimates are not available. These costs are minor, however, and could be less than the measurement error introduced by creating the data.

The specification of output is as follows:

6) $$Q = F(E,I,A)$$

where E is some measure of regional export activity, I provides input-output relationships (industrial complexes) in a region, and A measures the effects of the regional multiplier as expressed by the influence of regional aggregate variables upon output. More specifically,

7) $$E = a_1 + b_1 (P_a/P_b) + b_2(\upsilon S)$$

where P_a and P_b are product prices prevailing in the region and its trading partners, and (υS) is some measure of economic activity of the region's trading partners;

8) $$I = a_2 + b_3 \Sigma W_i EMP_i + b_4 EMP$$

where relevant input-output relationships are explained either by the weighted activity of other sectors in terms of employment or by an overall increase in employment activity in the region;

$$9)\quad A = g(P.I., GSP \ldots)$$

where P.I. is regional personal income, GSP is regional product, and where other variables may be stimulating output. Most specifications of output do not differ greatly from the above, although they might increase the external explanation of regional activity and eliminate the relative price factor as being unavailable. In addition, some models have included the national output of the same industry as an argument. It appears unlikely that regional exports will be stimulated by trading partner activity in the same industry unless such activity implies capacity problems elsewhere. Thus, these models may be confounding a measurement problem with a specification problem.[13]

Unfortunately, relative prices are not the only unavailable data. Some activity clearly responds more closely with corporate or government activity at the regional level than with household activity. Therefore, personal income will not be sufficient to specify regional activity. Although some models perform adequately with personal income alone, a regional product variable usually aids the analysis.[14] (However, the option of using total employment in place of total output may still be fruitfully exercised without data construction.) The specification gains are significant, but other variables also are important so the information loss is not overwhelming. Whether a gross product estimate is included must depend upon the limiting assumptions necessary to create such a variable.

If employment is used in place of aggregate output the demand for labor can be specified without any data construction and only a small number of assumptions as follows:

$$10)\quad \log L = a + b_1 t + b_2 \log ({}^{w}/p) + b_3 \log E + b_4 \log I + b_5 \log A.$$

If the equation appropriately estimates the derived demand for labor, it will also contain the underlying properties of the expenditures relationship.

The supply of labor must then be estimated to derive the wages used in the labor demand equation and to generate the wage bill for the region. At least three alternative specifications have been used to reflect three differing assumptions. First, if one assumes that wages are largely determined in national markets with only minor response to local conditions, then wages regionally depend upon national wages in the industry and the unemployment rate. Second, if wages are determined in local markets, then a measure of excess demand could be used to determine wage changes (e.g. a Phillips Curve type specification). One approach to this estimation took the form:

$$11)\quad \log W_i = a + b_1 \log UNEM + b_2 \log CPI + b_3 \log SST.[15]$$

Since unemployment (UNEM), consumer prices (CPI), and social security taxes per unit of labor (SST) all are aggregate regional values, this model permits wage

differences only because of differing response to these variables. Substantial changes in sector employment needs are not reflected in the results. Third, sector supply relationships are specified as follows:

$$12)\quad \log w^i_t - \log w^i_{t-4} = \gamma \log w^{i*}_t - \gamma \log w^i_{t-4} + \epsilon^{16}_t$$

where w* is the equilibrium wage and γ is an adjustment coefficient. In some work currently in process by this author, the following specifications have been used:

$$w^{i*} = a + b_1 EM^i + b_2 LF + b_3 RDPI + b_4 CPI + b_5 SST$$

where EM reflects employment in industry i, LF is the regional labor force, RDPI is real per capita disposable income in the region, and the other variables retain their previous meaning; also

$$\gamma = \alpha \text{ UNEM}.$$

In other words, prevailing employment and wages by sector reflect a system adjusting to desired levels while the current unemployment rate alters the speed of obtaining that wage. Thus wages can be determined either from national markets, aggregate local markets, or individual sector employment needs.

Each of the above formulations is based upon disequilibrium in labor markets. This latter factor can be determined either by directly estimating the unemployment rate or by estimating the labor participation rates and then deriving unemployment. Unfortunately, some data construction on population and migration are required at this juncture.

Once these migration factors have been determined, probably in terms of relative real wages between the region and the United States and the degree of disequilibrium in labor markets, i.e. the unemployment rate or the change in employment, the specification of the labor market will be complete. The remainder of the model specification is needed to estimate those variables used in explaining labor needs.

While the above discussion may appear to end in a perfunctory fashion and seems to exclude government policy variables, this is far from true. Household incomes must be estimated by explaining property income, transfer payments, and social security payments. National and regional tax policy not only affects disposable income, but the latter may affect state and local employment and their wage bill as well. Construction employment might depend upon permit activity which in turn may depend upon bank debits. The latter may in turn respond to regional disposable income. Therefore, considerable latitude remains for specifying policy variables and utilizing the remaining regional data that is available once the labor markets are fully specified.

Some tradeoffs between data and model specification are still necessary in this middle area of regional models. Product prices are necessary to measure the demand for labor but are largely unavailable. National product prices may be used, but they reduce the ability to identify differential regional growth in specific industries. If capital stock estimates are unavailable, equilibrium conditions may

be needed to justify a given specification for labor demand. Also, any statements concerning corporate activity within a region requires substantial data construction, e.g. the generation of regional product series, although the use of tax information may reduce the informational costs of creating such a series. Nevertheless, economically justifiable models of regional activity can be constructed from existing data with a minor amount of measurement problems, if the models concentrate upon labor market activity.

C. *Summary*

Any work in regional economic analysis requires a trade-off between model specification and data availability. While the creation of a theoretical framework without regard to data factors may be useful in stimulating the search for new data sources, it provides little comfort to the regional analyst. However, a labor market approach to regional activity, while excluding the regional effects of capital flows and requiring regional price information to be correctly specified, holds substantial promise for econometric analysis with a minimum of data constructed measurement errors. Furthermore, the models themselves provide a range of flexibility that will lead to useful analysis on the fundamental issue of why regional growth rates differ.

FOOTNOTES

2. The "Farm Income Situation", a publication of the Department of Agriculture, actually maintains estimates of monthly receipts and payments by farmers, by state.
3. Sometimes the labor service exports may greatly exceed contiguous area imports. For example, Orange County in Southern California maintains a $1 billion excess of purchasing power over earned income in the county. The surrounding counties on balance generate $400 million more in earnings than they gain in personal income. The remainder is earned by consulting and other activities at the expense of the remainder of the United States.
4. Of course, this procedure only derives the earned income for a county. Commuting patterns must again be utilized to determine the residences of all wage earners.
5. The issue is how much federal government goods and services are consumed locally. Some models [e.g. Moody and Puffer (21)] assume no federal government activity is consumed locally, while others have argued that purchases should depend upon payments. Neither appears to be correct. Certainly, some federal activity e.g. post office services, are consumed locally. In addition, the sum of all regional activities should equal national activity, which is not possible if regions consume no federal services. It is equally true that heavy highway expenditures may benefit a given region by substantially more than the current contributions made by that region to the highway fund (or the reverse). Some government expenditures are people-oriented while others respond to incomes. No good theory has been devised for allocating these activities to specific regions.
6. Normally, these businesses do not file tax returns that outline the returns and costs of regional activity. Rather, their profits are allocated among states.
7. This adjustment, however, assumes other than constant elasticities of substitution in production relationships. Otherwise, the wage bill to total income would not vary regardless of the magnitude of difference between hourly earnings.
8. In unpublished comments, Michael Evans has argued that regional differences occur not only because of different price and income responsiveness to changes in overall activity but also because of differing cyclical response due to regional differences in production techniques by industry. Some regions may have developed early and are now operating inefficiently while others have the latest techniques.
9. Although a region's production depends upon the availability of headquarters activity, which might reside elsewhere, differing regional returns to capital may not be sensitive to the share of headquarters activity that they use. The ownership of corporate earnings also may not generate differing regional behavior except to the extent that incomes received or the market value of ownership have changed. The corporation itself may be the appropriate behavioral unit to generate further economic activity from the retained earnings. Thus, instead of asking what is produced and who owns, regional investigators should be asking what changes production and who spends.
10. Because the statistical procedures are not peculiar to regional analysis, they have not been considered here. However, several general statements about estimation techniques could be made. First, specification problems are still so substantial that any full information estimator, such as a three stage least squares approach, would do more harm by spreading these errors to all equations than the good that is gained by considering system interactions in the estimating techniques. Second, measurement errors may be so substantial that some data transformations to resolve econometric ills may also create biased estimates of regressors that would be unbiased without the adjustments. This may be especially true of the adjustment for autocorrelated disturbances. Third, simultaneous equation problems must be addressed by the researcher. However, Glickman (12) and Crow (9) both have used ordinary least squares and two stage least squares to estimate their

models and have generally discovered that ordinary least squares tends to lower forecast errors. The reasons for this result are unclear, and must still be interpreted with caution.

11. Although the reduced form of the regional economic system expresses regional acitvity in terms of national activity, predetermined variables, and exogenous regional activity, it is not correct to conclude that the same numbers of these variables implies a similar economic system. For example, some models may explain wages in terms of last period's national wages, regional unemployment, and last period's wages, while others might state that wages depend upon regional employment, unemployment rates, and the ratio between last period's regional and national wages. The reduced form would contain the same variables but the importance of each variable would be substantially altered by the different specifications of the wage equations.
12. A preliminary version of an econometric model for Louisiana was estimated by use of industry output estimates. This two-step procedure in explaining employment changes led to lower explained variation than the single step approach advocated in this paper. Furthermore, an attempt to determine disequilibrium behavior between output and employment via a geometric adjustment procedure led to incorrect signs for most of the production related variables. Therefore, the output measures by industry appeared to be faulty and were unable to increase the dimensions of the analysis.
13. If regions contain differing arrays of production capability in industries, an increase in overall activity may also increase activity in the least efficient industries. However, the most efficient method of expressing this effect would appear to be the national demand factors that tend to increase the profitability of the least efficient plants. Otherwise, a model may suggest that New England textiles should increase when new plants are opened in the Carolinas when the opposite will probably occur.
14. A model for Kentucky was constructed on the basis of known data alone and has continued to perform well over the three year history it has been in operation. The only regional aggregates were personal income and total regional employment. In this instance, at least, the inherent specification problems were not significant.
15. A variation of this equation was used in models for Kentucky, California, and an early version of Louisiana. The Kentucky model tended to underpredict wage changes, although the equations excluded the social security tax effects. California and Louisiana have tended to overpredict wage changes. However, recent changes in labor force estimates indicate that California has been underreporting unemployment since it discontinued household surveys in 1971. (Unemployment in January 1974 was revised upward from 5.5% to 7.3%.) Recent studies at the national level suggest that b_1 and b_2 are changing. Partially for these reasons but also because of the failure to incorporate local sector pressures, the newer version of the Louisiana model contains the form of equation 12.
16. This formulation is derived from the assumption that current wages reflect a movement toward desired wages. Since annual contract periods are assumed, the adjustment period is assumed to encompass a year (or four quarters). A geometric adjustment factor is used of the following form:

$$w_t = \left[\frac{(w_t^*)}{(w_{t-4})}\right]^{\gamma} \times w_{t-4}$$

or $$\log w_t = \gamma \log w_t^* + (1-\gamma) \log w_{t-4}$$

A direct estimation of this relationship led to substantial errors in sign for the components of w*. The biases created by a lagged dependent variable were too intense. However, no loss of information is incurred by stating the above in terms of wage changes, i.e.

$$\log w_t - \log w_{t-4} = \gamma \log w_t^* - \gamma \log w_{t-4}.$$

This specification improved the signs but the explained variation was small and the sign on the lagged term remained positive. Once the variable adjustment term was used, where $\gamma = \alpha$UNEM, the above problems were eliminated.

BIBLIOGRAPHY

1. R. J. Anderson, Jr., "A Note on Economic Base Studies and Regional Econometric Forecasting Models," *Journal of Regional Science,* 1970, 10, 325-333.
2. K. J. Arrow, H. Chenery, B. Minhas, and R. M. Solow, "Capital-Labor Substitution and Economic Efficiency," *Review of Economics and Statistics,* August 1961, 43, 225-250.
3. F. W. Bell, "An Econometric Forecasting Model for a Region," *Journal of Regional Science,* 1967, 7, 109-127.
4. C. Blanco, "The Determinants of Interstate Population Movements," *Journal of Regional Science,* 1963, 5, 77-84.
5. C. Blanco, "Prospective Unemployment and Interstate Population Movements," *Review of Economics and Statistics,* 1974, 46, 221-222.
6. G. H. Borts, "The Equalization of Returns and Regional Economic Growth," *American Economic Review,* 1960, 50, 319-347.
7. R.P. Burton and J. W. Dyckman, *A Quarterly Economic Forecasting Model for the State of California,* Center for Planning and Development, Research Institute of Urban and Regional Development, University of California, Berkeley, 1965.
8. D. Chen, "A Tabular Survey of Selected Regional Econometric Models," *Working Paper* 11, Federal Reserve Bank of San Francisco (1972).
9. R. T. Crow, *An Econometric Model of the Northeast Corridor of the United States,* Prepared for Northeast Corridor Transportation Project, U. S. Department of Transportation, March, 1969.
10. R. T. Crow, "A Nationally-Linked Regional Econometric Model," *Journal of Regional Science 13:* 187-204.
11. P. Friedly, "A Note on the Retail Trade Multiplier and Residential Mobility," *Journal of Regional Science,* 1965, 6, 57-63.
12. N. J. Glickman, "An Econometric Forecasting Model for the Philadelphia Region," *Journal of Regional Science,* 1971, 11, 15-32.

13. N. J. Glickman, "The Structure of a Large-Scale Regional Econometric Model: Towards the Simulation of Regional Development," *Discussion Paper No. 235,* University of Pennsylvania.
14. N. J. Glickman, "Son of 'Specification of Regional Econometric Models'," Paper presented at the 1973 North American Meetings of the Regional Science Association, November 11, 1973.
15. G.H. Hildebrand and A. Mace, Jr., "The Employment Multiplier in an Expanding Industrial Market: Los Angeles County, 1940-1947," *Review of Economics and Statistics,* 1950, 32, 141-149.
16. W. Isard and S. Czamanski, "Techniques for Estimating Local and Regional Multiplier Effects of Changes in the Level of Major Governmental Programs," Peace Research Society, *Papers,* III, Chicago Conference 1965, 19-45.
17. J. W. Kendrick and C. M. Jaycox, "The Concept and Estimation of Gross State Product," *Southern Economic Journal,* 32: 153-168 (1965).
18. L. R. Klein, "The Specification of Regional Econometric Models," *Papers of the Regional Science Association 23:* 105-115 (1969).
19. W. L. L'Esperence, G. Nestel, and D. Fromm, "Gross State Product and an Econometric Model of a State," *American Statistical Association Journal,* 1969, 64, 787-807.
20. J. M. Matilla, "A Metropolitan Income Determination Model and the Estimation of Metropolitan Income Multipliers," *Journal of Regional Science 13:* 1-16 (1973).
21. H. T. Moody and F. W. Puffer, "A Gross Regional Product Approach to Regional Modelbuilding," *Western Economic Journal,* 1969, 7, 391-402.
22. H. T. Moody, F. W. Puffer, and R. M. Williams, "An Eight Region Model of the United States Economy," *Growth and Change,* October 1970, 20-26.
23. H. T. Moody and R. M. Williams, "Forecasting Gross Regional Product for Southern California," *California Management Review,* 1967, 1, 83-92.
24. G. L. Perry, *Unemployment, Wage Rates, and Inflation,* Cambridge, M. I. T. Press, 1966.
25. K. Polenske, *A Multiregional Input-Output Model for the United States,* (Cambridge: Harvard Economic Research Project, 1970).
26. F. W. Puffer and R. M. Williams, "An Econometric Model for Southern California and Forecasts for 1967," *California Management Review,* 1967, 3, 89-95.
27. D. Ratajczak, "A Method for Estimating Gross State Product," in R. M. Williams, S. Rabin and D. Ratajczak, "The Use of Anticipations Surveys in Forecasting for the Southern California Economy," *CIRET Conference,* Madrid, September 1969.
28. D. Ratajczak, "Simulation of the Economic Impact of State Witholding Upon the California Economy," *Proceedings 1972 Summer Computer Simulation Conference,* San Diego, 1217-1226.
29. C. Richter, "Some Limitations of Regional Econometric Models," *The Annals of Regional Science 6:* 28-34 (1972).
30. Research Seminar in Quantitative Economics, *Econometric Model of Michigan,* Ann Arbor, Michigan: University of Michigan Press, 1965.
31. B. F. Roberts and G. Wittels, "An Econometric Model of California," Paper presented at the *Winter Meetings, Econometric Society,* December, 1971.
32. J. L. Stein, "Interregional Comparisons of the Marginal Product of Capital," *Southern Economic Journal,* 1958, 25, 24-32.
33. G. E. Thompson, "An Investigation of the Local Employment Multiplier," *Review of Economics and Statistics,* 1959, 41, 61-67.