Building US National and Regional Forecasting and Simulation Models

GANG SHAO & GEORGE I. TREYZ

ABSTRACT In this paper, we describe how Regional Economic Models, Inc. constructs US national and regional forecasting and simulation models, using the US Bureau of Labor Statistics Outlook-2000 forecast. The building procedure extends a traditional input-output model to a dynamic and structural forecasting and simulation model. While much emphasis is on building a consistent US model, we also discuss the linkages and structural differences between the national and regional models. The procedure we present is generally applicable to any regional modelling undertaking. It extracts changing relationships from the national level for use in model building and forecasting at the regional level, where these relationships are not directly observable.

1. Introduction

Regional modelling has developed rapidly over the last three decades. Despite data limitations and other empirical difficulties, many single- and multiregional economic models have been constructed and implemented for forecasting and policy analysis. Notable examples are the 1963 and 1977 US multiregional input-output models (Polenske, 1980; US Department of Health and Human Services, 1984); the econometric models for Philadelphia (Glickman, 1971), Michigan (Shapiro & Fulton, 1985) and Ohio (Baird, 1983); and the regional computable general equilibrium models (Jones & Whalley, 1988; Harrigan & McGregor, 1989; Morgan et al., 1989).

Since 1980, starting with a core model developed for the National Academy of Sciences (Treyz et al., 1981), Regional Economic Models, Inc. (REMI) has produced regional economic and demographic forecasting and simulation models at the state and county level of the USA (see also Treyz & Stevens, 1985; Treyz et al., 1992). The models are used for regional economic forecasting and policy planning, some examples of which include estimating the economic effects

Gang Shao was Director of Development at Regional Economic Models, Inc., when this paper was written. He currently is Senior Consultant for the Policy Economics Group at KPMG Peat Marwick, 2001 M Street, NW, Washington, DC 20036, USA. George I. Treyz is Professor of Economics, University of Massachusetts at Amherst, and President of Regional Economic Models, Inc., 300 Lincoln Avenue, Amherst, MA 01002, USA. The authors wish to acknowledge helpful comments on an earlier draft by Charles Bowman, Ronald Miller and Dan Rickman. Responsibility for the contents of the article, of course, remain with the authors.
of air pollution control, construction of new factories, new transportation projects, tax changes, and utility rate increases in local areas.

In building a forecasting model for a subnational region, a US economic model plays an important role. The US model provides key structural parameters for the regional models and it also provides control totals in the building and use of regional models. For instance, when we build single-region models for each of 51 regions (i.e., 50 states plus Washington, DC), consistency checks for the regional models are performed to assure that the regional forecasts add up to the national total.

This paper describes how we build US and subnational economic models, using the recent Bureau of Labor Statistics (BLS) Outlook-2000 forecast, which includes national input-output accounts and national forecasts for other variables up to the year 2000. The BLS forecast provides fundamental information for REMI to construct the US national and regional models for forecast and policy simulation purposes. It is also important to us to extend the input-output model that is the main part of the US model in various ways; for example, by closing it with respect to final demand. The primary task is to construct internally consistent US national and regional models. Our discussion concentrates on the theoretical and empirical problems we face in the model-building process and the methods we use to solve them. We introduce an inventory residual approach for creating year-by-year technical coefficient matrices to replicate BLS historical and projected output. An approach to determine dynamically labor productivity in the forecast period is also presented, as well as a method for creating a new national forecast based on current final demand and personal income forecasts. We also discuss the linkages and structural differences between the US national and regional models.

The paper is organized in the following way. Section 2 discusses the process of building and projecting annual national input-output models. The subject in Section 3 is how to create a baseline forecast for any given final demand and to formulate a complete US forecasting and simulation model. Here, the REMI model is a model closed with respect to final demand that explicitly recognizes the components of real disposable income and also is used in regional models. Section 4 introduces the structure of the REMI regional model and reviews some important issues in the construction of a regional model. We offer some concluding remarks in Section 5.

2. National Input-Output Models

The BLS releases a national forecast biannually. The latest release was 'Outlook-2000' (BLS Monthly Labor Review, November 1989; BLS Office of Employment Projections, January 1990). The data include the US national input-output accounts (for 1977, 1982, 1986 and 2000), all the accounts are expressed in constant (1982) dollars), employment by industry, output by industry and gross national product (GNP) by final demand components. The 1977 input-output accounts are inflated to 1982 dollars from the 1977 Bureau of Economic Analysis (BEA) survey-based input-output table, and the 1982, 1986 and 2000 input-output tables are estimated or projected. The BLS projected the final demands for the years 1989-2000, by applying judgments and various economic assumptions. Then, it projected the employment and other economic variables for 1989-2000 consistent with its exogenous final demand projections. The input-output ac-

2.1. US Input-Output Models for Benchmark Years

The input-output accounts of the BLS follow the convention of a commodity-by-industry format. The commodities and industries are classified separately. This enables one to account for the multiple outputs that each industry may potentially produce; the input and output structures of each industry are recorded in different matrices.

The 'make' matrix \( V \) is used to represent the output structure of industries; element \( v_{ij} \) in the 'make' matrix records the amount of primary \( i \) or secondary \( (i+j) \) product (commodity) \( j \) produced by industry \( i \) \( (i = 1, \ldots, n) \) produced by industry \( i \). \( (i = 1, \ldots, n) \). The input structure for industries is represented in the 'use' matrix \( U \), in which the \( u_{ij} \) is the amount of intermediate use (input) of commodity \( j \) by industry \( i \).

Define the input proportion matrix as

\[
B = U (X)^{-1}
\]

where \( B = [b_{ij}] \) is the dollars' worth of commodity \( j \) used to produce one dollar's worth of the output of industry \( i \), \( X \) represents the vector of total industry output and the 'hat' over the vector creates a diagonal matrix from that vector.

For the industry-based technology assumption, define

\[
D = V (Q)^{-1}
\]

where \( Q \) is the vector total commodity output and \( D = [d_{ij}] \) designates the fraction of total commodity \( j \) that is produced by industry \( i \). These functions are referred to as commodity output proportions and use of the industry-based technology framework embodies the assumption that the output of each commodity produced by industries is in the fixed proportions given by the elements of \( D \). The matrix \( D \) is sometimes called the market-share matrix. Jansen & ten Raa (1990) have shown that the input-output models under the commodity-based technology assumption have desirable theoretical properties in some circumstances. However, because input-output models with commodity-based technology may yield negative technical coefficients or industry (commodity) output, we choose industry-based technology as a basic assumption. The input-output model under industry-based technology can be derived as

\[
X = (I - DB)^{-1} DHY
\]

where \( I \) denotes the identity matrix, \( H \) is the bridge (or final demand coefficients) matrix and \( Y \) is the vector of commodity final demand. The bridge matrix \( H = [h_{kj}] \) is defined to represent the dollars' worth of commodity \( j \) \( (i = 1, \ldots, n) \) used to produce one dollar's worth of purchases made by final demand consumer \( k \) \( (i = 1, \ldots, n) \).

Using equation (3), we can create the input-output models for those benchmark years for which the BLS provides input-output accounts. The corresponding
technical coefficients matrix $A (= DB)$ can be used to represent the production relationship between industries.

2.2. Technical Coefficients Matrix and Bridge Matrix for Non-benchmark Years

The previous section described how we created input-output models for the benchmark years. However, we want a technical coefficient matrix for non-benchmark years, because the REMI model needs to reproduce history from 1969 to 1988 and forecast from 1989 to 2000. This is the subject of this subsection.

In input-output analysis, the structural change of the US economy has been studied by many people (Leontief, 1953; Vaccara, 1970; Carter, 1970; Blair & Wyckoff, 1989). Traditionally, the structural change is examined as the changes in intermediate demand, which measures the demand for goods and services used as ingredients in the production process for final products. The most important finding from this research is that the structure of the US economy has changed from the 1970s to the 1980s. The increase in services relative to manufacturing and natural resource-intensive industries is an observable fact. However, the structural changes are relatively gradual over time, which is a typical characteristic of industrialized countries. The relatively gradual change in technical coefficients suggests that one can model the change in them using some relatively simple functions.

In the absence of annual, survey-based input-output tables, and in line with our needs, we use a non-survey method. In the literature, non-survey techniques to update input-output models have received much attention. Among them, several methods, such as RAS, an econometric approach, and linear interpolation, have been suggested and used in practice. However, the RAS technique depends heavily on the quantity and quality of the available information in both the 'base' and 'target' data sets. The technical coefficients estimated econometrically are based on conventional statistical procedures to determine cell entries (Jensen, 1980). There are an inadequate number of data points to support this approach. Therefore, we use the linear interpolation method to update the technical coefficients. Unlike the RAS method, which will not change coefficients with a value of zero, the linear interpolation will ensure that the changes take place if technical coefficient $a_{ij} = 0$ (for some $i$ and $j$) at one benchmark year and $a_{ij} = 0$ at another benchmark year.

The updating of technical coefficient matrices is achieved in the following manner. The technical coefficient matrices for 1969–1981 are interpolated linearly for each year by using the 1982 values and the technological change rates implied, using the 1982 and 2000 matrices as endpoints. The coefficient matrices for the years between two benchmark years (1982–1986 and 1986–2000) are interpolated simply by using the matrices of the two endpoint benchmark years. Although the technical coefficients estimated by linear interpolation can reflect smooth changes in technology and consumption, they are not guaranteed to produce the same industry output as that provided by the BLS data. This may be caused by many economic and non-economic factors. A few possible reasons are the difference between the estimated technical coefficients and the 'true' coefficients, unmodelled inventory changes, changes in technologies and changes in market share among industries.

To solve this problem, we adopt a two-step approach. First, we bring as much information as we can into the bridge matrix. For instance, we incorporate commodity-specific international trade data, which are part of 'Outlook-2000' from the BLS and cover the historical years from 1977 to 1987, into the bridge matrix. For 1969–1976, we used an estimated distribution for the major categories of exports and imports based on the 1977 data. Exports and imports data are vital in the REMI model, because they are used to determine the regional purchase coefficients (RPC) for the USA. These, in turn, are used as one of the determinants of the RPCs for each state and county (see Treya et al., 1992, for details). The RPC represents the proportion of local demand supplied locally by industries in the region. When used in the US model, they produce a model where industry production is to fulfill domestic demand and export requirements. Specifically, let $R = \text{diag}(R_1, R_2, \ldots, R_n)$ represent the RPCs for industries and define it as

$$ R = [X - D (EX)] [X - D (EX) + D (IM)]^{-1} $$

where $EX$ and $IM$ denote the vectors of commodity exports and imports of the USA. It is easy to show that equation (3) can be rewritten as

$$ X = (I - RD) \hat{Y}^{-1} [RD \hat{Y} + D (EX)] $$

where $\hat{Y}$ and $\hat{Y}$ are the bridge matrix and the final demand vector that exclude international trade. Instead of equation (3), equation (5) is used in the REMI model, because equation (5) represents domestic technology coefficients rather than total (domestic plus imports) coefficients.

We also use industry-specific inventory–sale ratios from the Survey of Current Business (1986) as a proxy to distribute the change in business inventory for the non-benchmark years. Here, we interpolate linearly other coefficients in the bridge matrix as we did for the technical coefficient matrix. Once we have updated the technical coefficient and bridge matrices for each year, we can compute the industry output from equation (3) (or equation (5)) for each industry for 1969–1988 (history) and for 1989–2000 (projections).

The next step is to calculate the difference between the industrial output (historical and projected) from the BLS and the output results just obtained, treat these differences as the amount of change in the business inventory, and combine them with the previous change in inventory in the bridge matrix. A second run is then performed, generating the industry output, which exactly replicates the BLS data. This replaces REMI’s previous approach, which was to use a multiplicative adjustment method to adjust the rows of the technical coefficient matrix rather than the inventory approach which we use here. The row modification weights (MULTS) for the technical coefficient matrix are defined as the ratio of the actual industry output to the estimated output. The inventory approach only affects the final demand, while the MULTS change the intermediate and final demands uniformly for every industry.

To examine the difference between the estimated output without using the inventory residual method and the output reported by the BLS, we compute a mean absolute normalized deviation (MAND) over all the industries for each year. The MAND for year $t$ ($t = 1969, \ldots, 2000$) is

$$ \text{MAND}_t = \frac{\sum_{i=1}^{n} \left| X^t_0 - X^t_{BLS} \right|}{\sum_{i=1}^{n} X^t_{BLS}} \times 100 $$

(6)
where the superscripts 'e' and 'BLS' on the industry output $X_n$ denote the estimated output and the output from BLS, respectively, and $n$ is the total number of industries. Thus, $X_n^e - X_n^{BLS}$ represents the residual change in inventory that we need to add to yield the $X_n^{BLS}$ values. We also compute a mean normalized deviation (MND) without taking the absolute value for the residual each year. Owing to cancellations of the positive and negative residuals, the values of the MND should be smaller than those of the MANID.

Table 1 shows the MAND and MND values for all the industries studied. First, the MAND and MND values are approximately zero for three benchmark years, namely 1982, 1986 and 2000. From 1980 to 2000, the MAND for non-benchmark years is also small, with a maximum MAND of 4.45% in 1993. For the years before 1982, the MAND increases gradually. Although the distribution of residuals for the industries may cause some industries, such as the service sector that traditionally have no inventory, to have non-zero inventory, we argue that, unlike the MULTS adjustment approach, the residuals will not be buried in coefficients in technical and bridge matrices.

3. Incorporating the US National Input-Output Model into the REMI Model and Formulating a Forecasting and Simulation Model

The US national input-output model that primarily represents the inter-industry relations forms a core for the REMI US model. In general, for any given final demand, the input-output model can tell us how much intermediate demand will be needed in the production process and how much total output will be produced by industries. However, the REMI model goes beyond this point. We create a complete structural and dynamic model to reflect changes not only in industry outputs but also in employment, personal income and real disposable income. In a broad sense, the traditional input-output model is an 'open' model with respect to final demand components, such as personal consumption, investment and government purchases. The REMI US model extends the input-output model to form a forecast and simulation model that is 'closed' with respect to most final demand components. Both the intermediate demand and final demand (with the exception of exports) are endogenously and simultaneously determined in the model. In this section, we first discuss a way to form a baseline forecast and then discuss the closed policy simulation model with respect to the final demand.

3.1. Baseline Forecast

To make a prediction, a model should be able to incorporate any estimations of the final demand and create a baseline forecast for other variables. The baseline forecast, which can then be used for policy simulations in national and regional models, reflects a prediction of model builders from current economic status. The estimation of the final demand may change when new economic data become available. Then, the baseline forecast from the US model should change as well. There are at least three alternative forecasts of the final demand that can be introduced into the REMI model. These are the BLS 'Outlook-2000', a trended forecast or the US prediction from other vendors. For the BLS 'Outlook-2000', some recessions have been implicitly built into the forecast data. For instance, the BLS predicts that the US economy will decline in 1993 and 1998 respectively. Even though these recessions are short and are of relatively small scale, they may not fit the needs of some analysts. Therefore, the trended prediction of final demand with a constant growth rate, taking the values of the 1988 and 2000 final demand (from the BLS) as two endpoints, may be more desirable.

Assuming exports are given and a national RPC is calculated, the industry output $X_i$ can be obtained by substituting the estimated final demand $Y$ into equation (5). Using the estimated employment per unit of output (EPV (or labor productivity$^2$)), the employment by industry then can be calculated from

$$E_i = EPV_i X_i,$$

where $E_i$ is the employment for industry $i$.

There are different ways to obtain the EPV. Suppose that we rely only on the BLS 'Outlook-2000' in forecast years, where the EPV$^2$ is calculated for each industry and for each year before putting them into the REMI model, and then is assumed not to be changed. However, as Verdoorn's law or Kaldor's second law indicates, there is a relationship between the growth of output and the growth of productivity for manufacturing industries (see McCombie (1986) and Whiteman...
(1987) for discussions). Therefore, the change in industry output should have an effect on the EPV.

In addition, labor productivity in the national economy has shown a substantial time trend in the past. Many macroeconomists studied this issue and found that the labor productivity increased dramatically before the 1970s and slowed down after that. Although there are various interpretations to account for the slower growth in labor productivity, such as oil crises and environmental regulations, most researchers are not satisfied by these limited explanations. However, it is beyond the scope of our paper to continue these discussions further. To reflect the trend and the effect of output on the EPV, we estimate the following equation for each of 49 private non-farm industries:

\[ EPV_t = \alpha_0 e^{\beta_t} (X_t/X_{t-1})^\gamma \quad (8) \]

where \( X_t \) represents output for industry \( i = 1, \ldots, 49 \) at time \( t = 1970, \ldots, 1988 \) and \( X_t/X_{t-1} \) is a cyclical measurement.

Assuming that a random disturbance is multiplicative, and taking a logarithmic form for equation (8), we estimate the equations by using the ordinary least-squares method. The expected signs for \( \beta \) and \( \gamma \) are negative. From the regression results, most industries fulfill this expectation. Except the lumber sector, all the manufacturing industries have a relatively high adjusted \( R^2 \), which shows a strong relationship between the EPV and the other variables in equation (8). The values of the t-statistics are also significant for all the coefficients of the manufacturing industries. For the non-manufacturing industries, 20 out of 28 sectors show a strong relationship between these variables as well. The weak results appear in those industries for which the data quality is always in serious doubt. Examples of these industries are local transit sector, air transportation sector, and brokers, credit and other investment sector.

Regarding the sign of the parameters, the manufacturing industries show a consistent result as expected, where \( \beta \) is in the range of \(-0.0347\) to \(-0.0036\) and \( \gamma \) is in the range \(-0.7788\) to \(-0.2230\). For the non-manufacturing industries, 15 out of 26 sectors show a pattern of increasing labor productivity (note that the estimated coefficient for time trend variables has a negative sign). For these 15 non-manufacturing industries, \( \beta \) is in the range \(-0.0532\) to \(-0.0004\) and \( \gamma \) is in the range \(-1.3381\) to \(-0.0364\).

A significant role of equation (8) is that the EPVs in a region will be determined by national time trend and local industry outputs when we apply the equation to the regional model. Of course, the labor intensity and total factor productivity relative to the USA are also taken into account for each industry in the region. Nevertheless, the EPVs in the local area that are determined endogenously (by equation (8)) in the regional model reduce the reliance on the US values.

To illustrate how we form the baseline forecast, we use the equations that determine personal income and real disposable income as an example. From the estimated employment (from equation (7)), we can compute the corresponding personal income and real disposable income. The components of personal income and real disposable income are personal contributions to social insurance, labor and proprietor's income, residence adjusted income, property income, transfer payments, taxes, and personal consumption expenditure deflator. We present the equations that determine these components of personal income below. These equations are only a subset of the equations used in the REMI model. They are not only used for the US national model, but also for the subnational model. It should be noted that, to determine personal income, we need initial values for some variables (such as those with a superscript 'us' in equations (9)-(15) below).

Somewhat similarly to our treatment of final demand, we have three ways to set up the values. One way is to use directly the data from the BLS 'Outlook-2000', the second is to use the trended values from the 1988 and 2000 BLS data as two endpoints, and the third way is to take a forecast from other sources. These values provide the initial estimates and the changing directions for the variables of personal income.

**Personal contributions to social insurance (TWPER)**

\[ TWPER = \lambda_{TWPER} \text{WSD} (\text{TWPER}^w/\text{WSD}^w) \quad (9) \]

where \( \lambda_{TWPER} \) is an area-specific adjustment coefficient that is determined from the value of the last history year, \( \text{WSD} \) is the total wage bill for all the industries in the local area and \( \text{TWPER}^w \) and \( \text{WSD}^w \) contain the initial values for the USA, as we mentioned above. It should be noted that the terms \( \text{TWPER}^w \) and \( \text{WSD}^w \) can be viewed as shift parameters in the creation of a baseline forecast, as is also the case for the US variables in equations (10)-(15) below.

**Labor and proprietor's income (YLP)**

\[ YLP = \text{WSD} + \text{YOL} = \sum_i E_i \omega_i + \sum_i \lambda_{YOL} E_i (\text{YOL}^i/\text{B}^i) \quad (10) \]

where \( \text{YOL} \) is other labor income in the local area, \( \omega_i \) is the local wage rate, \( \lambda_{YOL} \) is a region-specific coefficient and is determined from the value of the last history year, and \( \text{YOL}^i \) and \( \text{B}^i \) are other labor income and employment for the major industry \( i \) where \( i \in I \) in the USA.

**Property income (YPROP)**

\[ YPROP = \lambda_{YPROP} N (\text{YPROP}^w/N^w) = \lambda_{YPROP} \text{YPROP}^w \quad (11) \]

where \( N \) is the population.\(^{11}\)

**Transfer payments (V)**

\[ V = \lambda_V \text{V}^w (N - E (1 + \text{RA}^w/\text{YLP}^w))/(N^w - E^w) \quad (12) \]

where \( E \) is the local total employment, \( \text{RA}^w \) is the residential adjustment, and the term \((1 + \text{RA}^w/\text{YLP}^w)\) converts place-of-work employment into place-of-residence employment.

**Personal income (YP)**

\[ YP = YLP - TWPER + \text{RA}^w + YPROP + V \quad (13) \]

**Taxes (TAXES)**

\[ TAXES = \lambda_{TAXES} (YP - V) [\text{TAXES}^w/(YP^w - V^w)] \quad (14) \]
Real disposable income (RYD)

\[
\text{RYD} = (\text{YP} - \text{TAXES})/\text{CP}^u
\]

(15)

where \(\text{CP}^u\) is the personal consumption expenditure deflator\(^{12}\).

After the model is run as an open model with new final demand data, the new estimated values for the variables above will replace the old values. For example, the value of \(\text{YP}^u\) will be replaced by the new value of \(\text{YP}\) calculated from equation (13). Thus, the baseline forecast is formed, which has consistent estimates for all the variables calculated from the equation above. In other words, the baseline forecast eliminates possible inconsistency between the final demand and personal income. As an additional check for internal consistency, the equality of the total value-added and total final demand also can be checked out at this stage.

Until now, we have described the process of creating a baseline forecast with the REMI model. It is worth noting that the REMI model operates as a typical input–output model at this stage and generates a new set of values for industry employment, output and personal income, corresponding to any new final demand data. The new forecast values, which we call the baseline forecasts, are used in the US national and subnational models for forecast and simulation.

3.2. Formulating a Forecast and Simulation Model

We now have a complete US model, with a technical coefficients matrix, bridge matrix and the baseline forecast, which is ready for use either as a forecast and simulation US national model or as a driver\(^{13}\) for the construction of a subnational model.

To present how the US model can be formed for policy simulation, we first show the equations that determine endogenously the final demand components.

Personal consumption equations

\[
C_i = \sum_{k=1}^{13} \text{PCE}_{i,k} C_k = \sum_{k=1}^{13} \text{PCE}_{i,k} \left[\lambda_{\text{CONSP},i,k} C_k \left(\text{RYD}/\text{RYD}^w\right) \text{PV}_{\text{CONSP},i,k}\right]
\]

(16)

where \(C_i\) represents the total purchase of the product of industry \(i\) by 13 personal consumption categories, \(\text{PCE}_{i,k}\) is a consumption coefficient from the bridge matrix, \(C_k\) is the total consumption (which is represented by the \(i\)' subscript in the model for the \(k\)th consumption category, \(\lambda_{\text{CONSP},i,k}\) is a coefficient that represents an area's consumption of the \(k\)th category as a proportion of \(\text{RYD}\) relative to that for the USA\(^{14}\) (and is 1 for the US model), and \(\text{PV}_{\text{CONSP},i,k}\) is a multiplicative policy variable\(^{15}\) for the \(k\)th category. The superscript 'us' in equation (16) indicates the value from the baseline forecast. Equation (16) shows the difference between an 'open' model and a 'closed' model. When we create the baseline forecast and treat the model as an open model, the term RYD/RYD\(^w\) is ignored and, therefore, \(C_k = C_k^u\) in which \(C_k^u\) can come from any source. A similar observation can be seen from equations (17)–(19) below.

Investment equation for non-residential structure and equipment

\[
\text{INV}_{i} = \sum_{k=1}^{8} \text{INV}_{i,k} \lambda_{\text{INV},i} \left[K^d_i (K^d_i)^* \text{PV}_{\text{INV},i}\right]
\]

(17)

where \(\text{INV}\) is the total demand for industry \(i\), \(\lambda_{\text{INV},i}\) is an investment coefficient from

Investment equation for residential investment

\[
\text{INV}_{r} = \text{INV}_{r} \lambda_{\text{INV},r} \left[\lambda_{\text{INV},r} \left[\text{ARYD}/\text{ARYD}^w\right] \text{PV}_{\text{INV},r}\right]
\]

(18)

where the subscript \(r\) simply denotes residential investment and ARYD is the expected permanent real disposable income, which is determined as a geometrically moving average of real disposable income.

Government purchase equation

In the REMI model, the purchases by state and local governments are determined endogenously by changes in the population. There are four components of state and local government expenditures being considered: education, health and welfare; safety; others. The equation is as follows:

\[
\text{GOV}_{i} = \sum_{k=1}^{4} \text{GOV}_{i,k} \lambda_{\text{GOV},i,k} \left[N^w_i (N^w_i)^* \text{PV}_{\text{GOV},i}\right]
\]

(19)

where \(\text{GOV}_{i}\) is the total purchase of the product of industry \(i\) product by state and local governments, \(\lambda_{\text{GOV},i,k}\) denotes the total purchases of state and local governments and \(N^w_i\) is a coefficient from the bridge matrix.

The equations (16)–(19) clearly present how the REMI model determines endogenously the three major components of final demand, which in turn affect directly the purchases of industry outputs. As we mentioned earlier, the REMI model can reproduce the same industry output as that with the BLS data if we take the BLS 'Outlook-2000' as the baseline forecast and if the proportion terms in equations (16)–(19) are equal to one; i.e. RYD = RYD\(^w\) in equation (16), \(K^d_i = K^d_i^w\) in equation (17), ARYD = ARYD\(^w\) in equation (18) and \(N^w_i = N^w_i^w\) in equation (19). For a policy simulation, a simple example can be used for demonstration.

Let us suppose that personal consumption in the automobiles and parts category, which is the first need of personal consumption, would increase by 5% in the first forecast year and by 7% in the second forecast year. To implement this hypothesis, we need to do is simply to put these percentage changes into the policy variable \(\text{PV}_{\text{CONSP},i}\). The change introduced in one final demand may increase or decrease other final demand, because of changes in real disposable income (see equations (9)–(15)) and other variables. Taking the example we just discussed, when the personal consumption of automobiles and parts changes, increases in industry output and employment result. The increase in employment would create more purchasing power (more real disposable income) and more personal consumption. Therefore, as implied in equation (16), other categories of personal consumption will be affected. This feedback effect is a characteristic of a closed model.
4. Constructing a Regional Forecasting Model

4.1. General Outline of the Structure of a Regional Model

Overall, the REMI US model is designed to have a similar structure to that of the REMI regional models. However, many structural details are different. For instance, in the regional models, industries in a region are classified into two groups, namely national industries and regional industries\(^4\), which are not distinguished as such in the US model. National industries include all the manufacturing industries (except stone, clay and glass products, printing, and petroleum and coal products) and the hotels sector, while regional industries cover the other two-digit SIC industries. The basic assumption behind this classification is that national industries are competing on the national market and their products must be priced at the average national prices. Any changes in the production costs affect the profitability of national industries, which in turn affects the potential market share reflected in the RPC and export share. However, a change in the production costs for regional industries affects directly their competitiveness in the export market and in the local market with imports.

Migration is another component that is treated differently in national and regional models. For the USA as a whole, international migration is the only migration that needs to be considered. However, for a local region, other migration, such as economic migration, retiree migration and former military personnel and their dependents re-entering the civilian population, also should be considered. Interregional migration caused by economic factors is determined by employment opportunity, real wage rate and wage mix for industries in the area relative to the nation, and by a region-specific amenity effect that is calculated based on past observed economic migration behavior (see Greenwood et al. (1991) and Treyz et al. (forthcoming)).

When we build the US model, we apply the RPC to each industry to obtain the domestic industry output. When we construct a regional model, we want to obtain the industry output produced by local industries in the area. In other words, we need to have a regionalized input–output model to bridge between local demand and local output. The local demand is the total demand required in the region and is fulfilled by imports and local output, while the local output satisfies both exports and local requirements. There are few cases where regional input–output models have had survey-based technical coefficients. Even so, most of them are at the state level and not at the county level, although the county is a fundamental unit for policy planners. Some methods that regionalize national technical coefficients from different perspectives are seen in the literature. To name a few, location quotients, cross-industry quotients, supply-demand pool approaches, fabrication effects and regional purchase coefficients are suggested in the literature (see Miller & Blair (1985) for a summary). Again, REMI uses the RPC method to regionalize the national technical coefficients to express the proportion of local demand supplied locally by industries in the region. Hence, if we use the RPC for a local area and substitute it into equation (5), the local industry output can be determined. As we discussed above, the RPC for the national industries is determined by each industry's profitability relative to the nation and industry mix for the national industries. However, the RPC for the regional industries is determined by relative production costs and relative value added.

Each region's economy has its own unique features. To build a model for each region, we calibrate the parameters (such as \(\lambda_{\text{TPPER}}\) in equation (9), \(\lambda_{\text{VOL}}\) in equation (10) etc.) in the model to reflect the characteristics of the regional economy. The calibration is done by solving those equations where there is a parameter based on the value of the last historical year (LHYR). For example, \(\lambda_{\text{TPPER}}\) can be solved as

\[
\lambda_{\text{TPPER}} = (\text{TPPER/\text{WSD})} \times \text{LHYR)/(TPPER/WSD)} \times \text{LHYR}
\]

4.2. Reconciliation of Regional and National Variables

The consistency between the national and regional models is an important issue for regional modelling. Even though we use the national variables as shares in various equations to estimate regional variables, it does not always guarantee that the sum of regional variables across all regions will be exactly the same as the national aggregates. The difference between the national and regional variables, if there is any, may be for many reasons. One possible reason is implied by the non-linearity of the model structure. Another example of a reason that regional results might not add up exactly to the national total could be an effect by shifts in production among regions that have different factor productivities.

However, we have found that the sum of the results for the 51 regional models in long-term forecasting is close to the US national forecast. For our most recent forecasts, by the year 2000 (i.e., the eleventh forecast year), the sum of the forecast values for 51 regions for total employment, personal income and population differed from the corresponding US predictions by only 0.116%, -0.001%\(^7\) and -0.311%\(^7\) respectively.

5. Concluding Remarks

In this paper, we present a modelling procedure that is applied at REMI for building the national and regional forecast and simulation models. Much emphasis is on building the REMI US model and facilitating the generation of a new US forecast based on new final demand and personal income data. The building procedure extends a traditional input–output model to a dynamic and structural model that can be used for forecasting and policy simulation. Furthermore, it allows the introduction of features that apply to regions, while using as parameters some national values that result from the US model. This procedure is generally applicable to any regional modelling undertaking. It extracts changing relationships from the national level for use in model building and forecasting at the regional level, where these relationships are not directly observable.

Notes

1. This list is not meant to be exhaustive but rather illustrative of the types of regional models that have been built. For a survey of regional econometric models, see Bolton (1985), for input–output models, see Richardson (1985) or Rose & Niemnyk (1989) and, for computable general equilibrium models, see Shoven & Whalley (1984).

2. In general, the benchmark input–output models refer to full-survey-based input–output models. Here, we use it to distinguish those input–output accounts provided by the BLS with those generated by REMI.

3. In the BLS input–output accounts, \(n = 226\) and the number of commodities is the same as the number of industries.

4. The other types of structure changes have also been studied in the input–output analysis, such as changes in the final demand, output and value-added multipliers (see Blair & Wyckoff, 1989).
A study conducted by Szymery (1989) examines the trade-off between error and information in the RAS procedure. Because we find that some of value added in the 1977 input-output table (in 1982 constant dollars) have negative values, which is inappropriate for our analysis, we decide not to use the BLS 1977 table (in 1982 dollars). Throughout this paper, the subscript representing the time period used only if it is necessary for clarity. Conventionally, labor productivity is defined as the ratio of employees to output. The employment per unit of output is the reciprocal of labor productivity. The BLS provides industrial employment and output for 1989-2000. From these data, EPV values can be calculated. Some researchers use an input-output model to examine labor productivity (see, for example, Ochoa (1986) and Cowey & Miller (1987)). For other studies, see, for instance, Hall (1986). Since the population for the USA is determined outside the model, N = N0.

To simplify our discussion, we assume that CIP is given. However, in the real model, it is determined from the industry sales prices weighted by the coefficients of personal consumption. In a sense, the driver simply means that the values of the national variables can be taken as a base for incorporating changes such as shifts in consumption preferences, into regional forecasts and can be used to generate external demand for each industry in a regional model.

This is determined from the BLS Consumer Expenditure Survey (1983).

An additive policy variable also can be added into the equation. Based on a procedure for calculating interregional and international exports in 1977, an industry is classified as a national (regional) industry if the sum of exports from all states is 50% or more (less) of the total production in that industry.

The negative percentage means that the sum of regional variables is smaller than the corresponding US variable.

References