# APPENDIX A
# REMI ECONOMIC MODELING BACKGROUND MATERIALS

# TABLE OF CONTENTS

# ECONOMIC DOCUMENTATION FOR REMI POLICY INSIGHT

The basis for the socioeconomic data used as growth surrogates in this study was a REMI Policy Insight model of the 58 California counties. This Appendix presents a summary of the REMI Policy Insight model, including the model's structure and major data sources. The major features embedded in REMI's Policy Insight model are described in *The REMI Economic-Demographic Forecasting and Simulation Model* by Treyz, Rickman & Shao, International Regional Science Review, Vol. 14, No. 3, pp. 221-253, 1992, which is enclosed. The major changes to the material in this article are as follows:

1. The reformulation of the consumption sectors of the model to incorporate price responses, high and low income elasticities, different responses to increased population and changes in per capita income, fixed effects for each area and responses of medical expenditure to the age structure of the population (See the enclosed paper, *Consumption Equations for a Multi-Regional Forecasting and Policy Analysis Model* by George Treyz and Lisa Petraglia, November, 1996).

2. Endogenous participation rates by age, gender, race, cohort, and population by age, gender, and race (See the attached paper, *Regional Labor Force Participation Rates* by George I. Treyz, Chris G. Christopher, Jr., and Chengfeng Lou, March 3, 1996).

3. Chain weighted price index (See the enclosed paper).

4. Calibration of trade flows taking distance as well as demand and supply into account.

Aspects of the model that are not fully documented in the above sources include the following, which are also included in this appendix:

1. Demographic/migration module.

2. Data sources and estimation/calibration procedures.

1992 PRICES & CHAIN WEIGHTED INDEX

JULY, 1996

REMI recently converted its real dollar concepts from 1987$ to chained 1992$. This process involved several steps. The first consideration involved the U.S. input-output tables, as the most recent release available from the BLS is still based on 1987$. After much consideration, it was decided that the best approach would be to continue using the industry-by-industry coefficient matrices derived from the BLS 1987$ industry and commodity flow matrices. Any attempt to convert from 1987$ flows to chained 1992$ flows would likely lead to inconsistent tables, and therefore introduce more new errors than those we would be attempting to correct. For this reason, we decided not to change the technology, and instead to just change the base year of the dollars. It is likely that the next BLS release (in November 1997) will be based on chained 1992$.

In order to generate a consistent industry-specific historical output series in chained 1992$, we inverted the I-O matrices (year-by-year for 1969-1994) and fed in the new chained 1992$ final demand data published by the BEA (Survey of Current Business). For conversion of the 2005 BLS data, we applied the 1994 92$/87$ CPI ratio. We then applied the industry-specific value-added share of output coefficients from BLS based on their 1987$ I-O tables to this generated chained 1992$ output series to get value-added in chained 1992$. The GSP data (from BEA) is still only available in 1987$, so we reconciled it by industry to the new value-added series (by adding up all the states and comparing to the U.S.) to convert to chained 1992$.

To generate industry-specific deflators in chained 1992$ over history, we applied the 92$/87$ CPI ratio each year to the 1987$ series, and then normalized to make sure each industry's deflator equaled 1.0 in 1992.

As was previously stated, all real dollar variables in 2005 were converted using the 1994 92$/87$ CPI ratio. We then used the same interpolation and extension methodologies to generate the full forecast. We were granted permission by R.S.Q.E. to use their new chained 1992$ short-term forecast to capture predicted business cycles.

# TABLE OF CONTENTS

**For the**
**Reprints Volume**

**SECTION**

# SECTION I

## SELECTED ARTICLES ABOUT MODEL STRUCTURE



---

## EARLIER ARTICLES THAT HAVE BEEN SUPERCEDED

# SECTION I (cont.)

## EARLIER ARTICLES THAT HAVE BEEN SUPERCEDED

| | | *Page No.* |
|---|---|---|
| 4. | Stevens, Benjamin H., G.I. Treyz, D.J. Ehrlich, and J.R. Bower. "A New Technique for the Construction of Non-Survey Regional Input-Output Models." *International Regional Science Review,* 8(3) 1983, pp. 271-286. | Available From REMI |
| 5. | Treyz, George I., and B.H. Stevens. "The TFS Regional Modeling Methodology." *Regional Studies,* 19(6), 1985, pp. 547-562. | Available From REMI |
| 6. | ---------. "Fundamentals of Regional Macroeconomic Modeling: Part I." *The Survey of Regional Economic Literature,* Pilot Issue, November, 1986, pp. 11-34. | Available From REMI |
| 7. | ---------. "Fundamentals of Regional Macroeconomic Modeling: Part II." *The Survey of Regional Economic Literature,* (2), June, 1987, pp. 1-28. | Available From REMI |
| 8. | ---------, M.J. Greenwood, G.L. Hunt, and B.H. Stevens. "A Multi-Regional Economic-Demographic Model for Regions in the United States." in *Recent Advances in Regional Economic Modeling.* Edited by F. Harrigan and P.G. McGregor. London: London Papers in Regional Science #19, Pion Limited, 1988, pp. 66-82. | Available From REMI |

# SECTION II

## ARTICLES ABOUT USING THE REMI MODEL

(see also "Unpublished Articles and Excerpts" - Section V and "Directory of Studies Using the REMI Model" - Section VI)

# SECTION II (cont.)

## ARTICLES ABOUT USING THE REMI MODEL

# SECTION III

## ARTICLES REVIEWING THE MODEL & BOOK REVIEWS



* Another version of this article is available upon request.

## SECTION III (cont'd)

# SECTION IV

## REPRINTS FROM BLS U.S. FORECAST

# SECTION V

## UNPUBLISHED ARTICLES AND EXCERPTS

# SECTION VI

## REMI BROCHURES, UPDATES, CLIENT LISTS, ETC.

© *International Regional Science Review*, Vol. 14, No. 3, pp. 221–253, 1992

***Editor's Note:*** *In May 1980 the* Journal of Regional Science (20, 2) *published a symposium on multiregional econometric models. Roger Bolton wrote in the introduction that these models are "a major new development in regional economics which has significant implications for theory, policy analysis, and data development." This issue of the* International Regional Science Review *revisits the topic 12 years later. It contains articles describing the current state of two of the four models featured in the original symposium. NRIES II and REMI are extraordinary successes in the history of regional modeling. They have been used repeatedly for policy studies, impact analyses, and forecasts, and they have been updated and improved on a regular basis. They represent the state of the art of multiregional econometric modeling practice. Future issues of the* Review *will feature other methods that now claim the mantle of "major new development."*

# *The REMI Economic-Demographic Forecasting and Simulation Model*

**George I. Treyz**

*Department of Economics*
*University of Massachusetts*
*Amherst, Massachusetts 01003 USA*

**Dan S. Rickman**
**Gang Shao**

*Regional Economic Models, Inc. (REMI)*
*306 Lincoln Avenue*
*Amherst, Massachusetts 01002 USA*

**ABSTRACT** This article presents the Regional Economic Models, Inc. (REMI) Economic-Demographic Forecasting and Simulation (EDFS) model, which is used for regional forecasting and policy simulation in both the private and public sectors in the United States. The detailed structure of the model is presented. To illustrate the dynamic simulation properties of the model, results of two sample simulations for a REMI multi-area model of a region in Southern California are presented. Post-sample historical forecasts for all U.S. states are provided to evaluate the forecasting capabilities of the model.

## 1. Introduction

The Regional Economic Models, Inc. (REMI) Economic-Demographic Forecasting and Simulation (EDFS) model is designed with

Frank Giarratani and Andrew Isserman were instrumental in the initiation and development of this paper. Three anonymous referees provided useful comments, Sherri Pierce and Laura Corcoran played an important part in building REMI models, and Erin O'Toole prepared the manuscript with care. Dan Rickman is now at the University of Nevada, Las Vegas, and Gang Shao is at KPMG Peat Marwick in Washington, D.C.

Received 12 April 1990; in revised form 12 March 1991.

the objective of improving the quality of research-based decision making in the public and private sectors. It is calibrated to many subnational areas for forecasting and policy analysis by government agencies, consulting firms, nonprofit institutions, universities, and public utilities throughout the United States. Simulations with the model are used to estimate the economic and demographic effects of economic development programs, transportation infrastructure investments, environmental improvement regulations and projects, energy and natural resource conservation programs, state and local tax changes, and other policy initiatives.

The structure of the model incorporates interindustry transactions and endogenous final demand feedbacks. In addition, the model includes: substitution among factors of production in response to changes in relative factor costs, migration response to changes in expected income, wage response to changes in labor market conditions, and changes in the share of local and export markets in response to changes in regional profitability and production costs. The essence of the REMI model is the extent that theoretical structural restrictions are used instead of individual econometric estimates based on single time-series observations for each region. The explicit structure of the model makes it easier to use policy variables that represent a wide range of policy options and to track policy effects on all the variables in the model.

Inclusion of price responsive product and factor demands and supplies gives the REMI model much in common with Computable General Equilibrium (CGE) models, which have been widely used in economic development (Dervis, DeMelo, and Robinson 1982), public finance and international trade (Shoven and Whalley 1984), and have been applied more recently in regional settings (Jones and Whalley 1988; Harrigan and McGregor 1989; Morgan, Mutti, and Partridge 1989). The models differ, however, in that static CGE models usually invoke market clearing in all product and factor markets, and dynamic CGE models typically assume a perfect foresight intertemporal clearing of markets or temporary market clearing if expectations are imperfect; whereas, the REMI EDFS model does not require product and factor markets to clear continuously. The time paths of responses between variables are determined by combining a priori model structure with econometrically estimated parameters.

The REMI EDFS model's basic structure was first implemented as the Massachusetts Economic Policy Analysis Model (Treyz, Friedlaender, and Stevens 1980; Treyz 1981). A core version of this model for every state was developed for the National Cooperative Highway Research Project (NCHRP) and became publicly available in 1980 (Treyz, Stevens, and Ehrlich 1981). Treyz (1980) proposed a multiregional generalization, and REMI refined and developed the NCHRP model in response to new research methods and client requirements.

Early results of using the methodology were reported by Treyz and Stevens (1985) and Treyz et al. (1988).

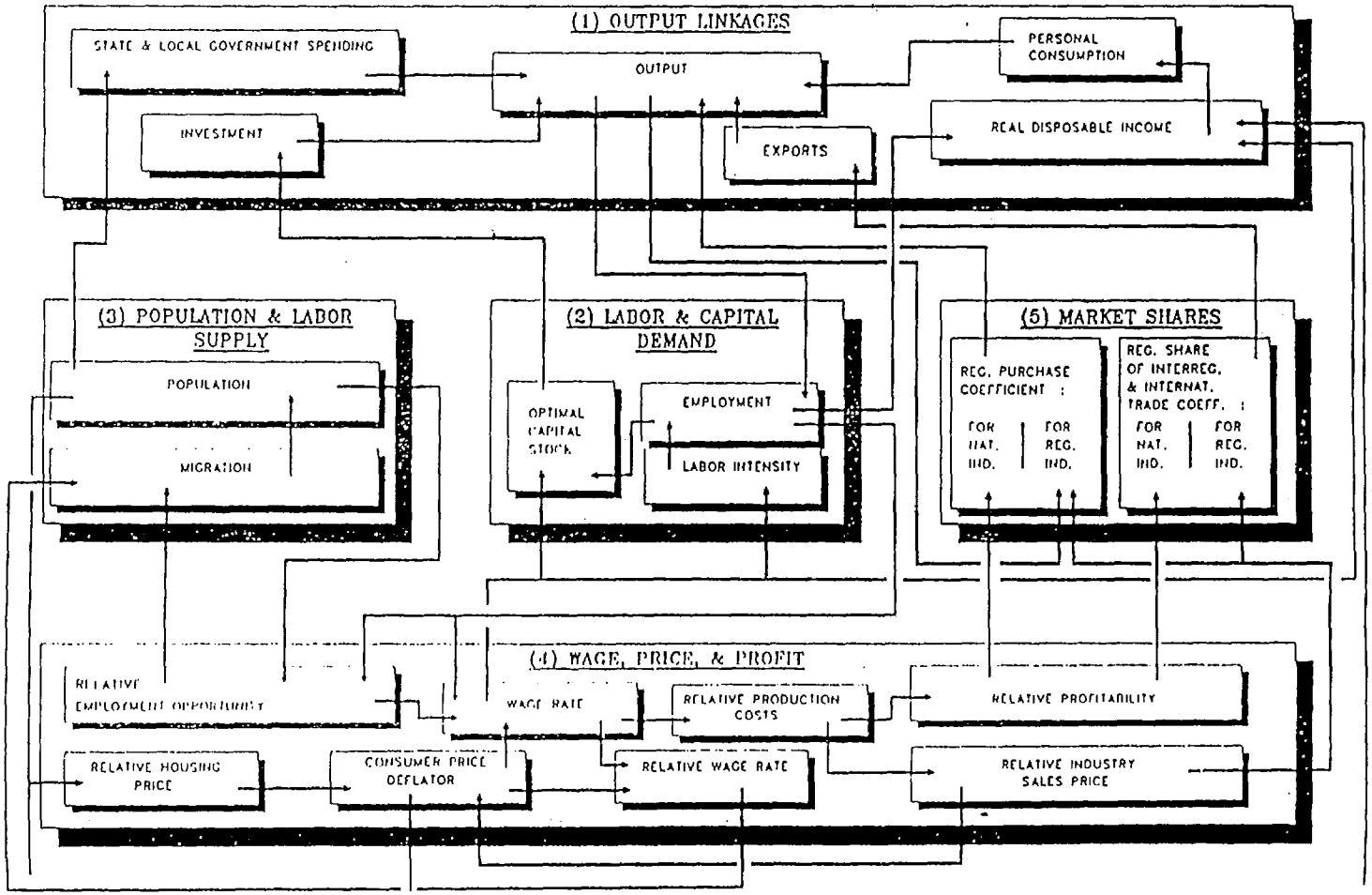
Since these earlier published accounts, the theoretical structure of the model has been further developed. A complete demographic module has been added that integrates economic and demographic processes. Research based on new data sources and econometric formulations has continued on the determinants of human migration and the location of firms, and new estimation techniques for regional wage determination have been developed. Besides documenting the new structural features of the model, this article presents the entire model in much more detail than has been previously published. Also, the dynamic simulation properties are examined and results of post-sample period forecasts for all U.S. states are presented. The article concludes with a few summary comments.

## 2. Model Structure

Although the model contains numerous equations, the five blocks in figure 1 illustrate the underlying structure of the REMI model. Each block contains several components (shown in rectangular boxes), and the lines and arrows represent interactions of key components both within and between blocks. Most interactions flow both ways, indicating a highly simultaneous structure. Block 1, output linkages, forms the core of the model. An input-output structure represents the interindustry linkages and final-demand linkages by industry. Interaction between block 1 and the rest of the model is extensive. Predicted outputs from block 1 drive labor demand in block 2. Labor demand interacts with labor supply from block 3 to determine wages. Combined with other factor costs, wages determine relative production costs and relative profitability in block 4, which affects the market shares in block 5. The market shares are the proportions of local demand in the region in block 1 and exogenous export demand that local production fulfills.

The endogenous final demands include consumption, investment, and state and local government demand. Real disposable income drives consumption demands and is calculated by using the regional consumer price deflator from block 4 to deflate nominal disposable income. An accounting identity defines nominal disposable income as wage income from blocks 2 and 4, plus property income related to population and the cohort distribution of population calculated in block 3, plus transfer income related to population less employment and retirement population, minus taxes. Optimal capital stock calculated in block 2 drives investment, and population in block 3 drives state and local government final demand. The endogenous final demands combined with exports drive the output block.

**FIGURE 1**
**Endogenous Linkages in the REMI Model**



Detailed descriptions of the components in each major block of the model follow and include theoretical underpinnings, empirical methodologies, and linkages to other components.

## Block 1 — Output Linkages

*Output equations.* Outputs for 53 sectors (49 private nonfarm industries plus 3 government sectors and the farm sector) are calculated from a regionalized input-output model. For industry i (i = 1, . . . , 49) the output equation is

$$Q_i = \sum_{j=1}^{49} R_i\, a_{ij}\, Q_j + R_i\,(C_i + I_i + G_i) + X_i, \tag{1}$$

where $Q_i$ is output in industry i in the region, $R_i$ is the regional purchase coefficient, $a_{ij}$ is the technical coefficient from the national input-output model, and $C_i$, $I_i$, $G_i$, and $X_i$ are the regional final demand components denoting personal consumption, investment, government spending, and exports by the region.

The regional purchase coefficient, $R_i$, represents the proportion of local demand supplied locally by industry i. By incorporating it into the model, the national input-output technical coefficients are regionalized and the need for an explicit import component is eliminated. Historical and projected U.S. national input-output tables produced by the Bureau of Labor Statistics provide the technical coefficients for benchmark years.[1] Linear interpolation of the technical coefficients between benchmark years gives the technical coefficients for each year.

If the region has more than one area, exports out of the area will be divided into two parts. The first, $X_i^{kr}$, is shipments from area k to other areas within the region (r), and the second, $X_i^{ku}$, is sales from area k outside the multi-area region (u). Equation (1) can then be rewritten for the multi-area model as

$$Q_i = R_i\left(\sum_{j=1}^{49} a_{ij}\, Q_j + C_i + I_i + G_i\right) + X_i^{kr} + X_i^{ku}. \tag{2}$$

Exports to the rest of the multi-area region are proportional to total imports of other areas (h = 1, . . . , m and h ≠ k) in the region. They are determined as

$$\begin{aligned} X_i^{kr} &= S_i^{kr}\left(\sum_{h=1,h\neq k}^{m} M_i^h\right) \\ &= S_i^{kr}\left[\sum_{h=1,h\neq k}^{m} (1 - R_i^h)\left(\sum_{j=1}^{49} a_{ij}\, Q_j^h + C_i^h + I_i^h + G_i^h\right)\right], \end{aligned} \tag{3}$$

[1] Since 1972, the U.S. input-output accounts have adopted a commodity-by-

where $S_i^{kr}$ is the share of exports into other areas from area k, $M_i^h$ are the imports of area h, and $R_i^h$ is the regional purchase coefficient for industry i in area h.

Exports to the rest of the country and the world can be written as

$$X_i^{ku} = S_i^{ku} X_i^u,$$

where $S_i^{ku}$ is the regional share of interregional and international trade of the United States.

*Consumption equations.* Real disposable income received by consumers translates into consumption demand by industry, $C_i$, according to

$$C_i = \sum_{j=1}^{13} PCE_{ij} \, Concol_j \, C_j^u \, (RYD/RYD^u), \tag{4}$$

where $PCE_{ij}$ is a coefficient denoting the proportion of consumption category j satisfied by industry i,[2] $Concol_j$ is a location-specific differential consumption measure derived from a consumer expenditure survey, RYD represents real disposable income in the region, and $C_j^u$ is consumption of good j in the United States.

*Real disposable income equations.* Real disposable income in the region equals personal income adjusted for taxes and the cost of living. Total personal income depends on wages and salaries, other labor income, property income, personal contributions to social insurance, transfer payments, and an adjustment to account for the difference between place-of-work and place-of-residence earnings.

Wage and salary disbursements, WSD, are an aggregation of individual industry wages and salaries. Thus,

$$WSD = \sum_{i=1}^{53} E_i \, w_i,$$

where $E_i$ is employment of industry i, and $w_i$ is the wage rate of industry i.

Employees also receive other labor income, such as fringe benefits, and the self-employed generate proprietors' income. The Bureau of Economic Analysis reports the total of these two income sources by

industry input-output framework in which commodities and industries are classified separately. Under the assumptions of industry-based, commodity-based, and mixed technology, different input-output models that connect relations between commodity (industry) output and commodity (industry) final demand can be constructed (see Miller and Blair 1985). The REMI EDFS model is an industry-based, industry-by-industry input-output model.

[2] The matrix PCE = $[PCE_{ij}]$ is often called a bridge matrix. The coefficients are projected forward based on the Bureau of Labor Statistics projected values for the United States.

major sector, so the local change in these totals on a per employee basis is predicted by their national change using the following equation:

$$YOL_I = \lambda_{YOL}\, E_I\, (YOL_I^u/E_I^u),$$

where $YOL_I$ is other labor income and proprietors' income for major sector I (to which industry i belongs), $\lambda_{YOL}$ is a region-specific coefficient and is greater than 1 if the region has a higher ratio than the nation, and $E_I$ is employment in sector I. Total labor and proprietors' income, YLP, for all industries in the region can be calculated as

$$YLP = WSD + \sum_I YOL_I.$$

Property income, YPROP, depends on both population and its age distribution as well as historical regional differences in property income received.

$$YPROP = \lambda_{YPROP}\, NP\, (YPROP^u/NP^u) \tag{5}$$

and

$$NP = L65 + m65 \times G65, \tag{6}$$

where m65 is the U.S. ratio of per capita property income received for persons 65 years and older (G65) relative to property received by persons younger than 65 (L65), and $\lambda_{YPROP}$ adjusts for regional differences and is calculated in the last historical year by solving equations (5) and (6).

Personal contributions to social insurance, TWPER, are predicted as

$$TWPER = \lambda_{TWPER}\, WSD\, (TWPER^u/WSD^u),$$

where $\lambda_{TWPER}$ is a coefficient calculated historically to adjust for regional differences in TWPER.

The residence adjustment, RA, is used to convert place-of-work income (wage and salaries, other labor income, and personal contributions for social insurance) to place-of-residence income. It is calculated as a fixed share of the income on which it depends. A positive residence adjustment in the last history year ($RA_T > 0$)[3] implies that some residents living in the area work outside the area, so RA will depend on income outside the area. For a one-area model, income outside the area is given by U.S. income. For a multi-area model, income outside the area is given by national income plus negative residential adjustments in other areas, LOSTINC. If $RA_T$ is negative, RA depends on local income, i.e., wages, proprietors' income, and

[3] Subscripts for time (t) are used only in equations where they are necessary for clarity. The capital T is used to indicate the last year of history.

other labor income minus social security contributions. The equation for RA is

$$RA = \begin{cases} \lambda_{RAMR}\ LOSTINC + \lambda_{RAU}\ (WSDNF^u + YOLNF^u - TWPER^u), & \text{if } RA_T \geq 0 \\ \lambda_{RA}\ (WSDNF + YOLNF - TWPER), & \text{if } RA_T < 0 \end{cases} \quad (7)$$

where $\lambda_{RAMR}$, $\lambda_{RAU}$, and $\lambda_{RA}$ are coefficients (for a one-area model, $\lambda_{RAMR} = 0$), LOSTINC is the sum of the income in the rest of the multi-area region generated by negative residential adjustments, and WSDNF ($WSDNF^u$) and YOLNF ($YOLNF^u$) are nonfarm wages and other labor income.

Transfer payments, V, depend on the number of persons in each of three groups: persons 65 years and older, persons younger than 65 who are not working, and all persons not working. Transfer payments also are adjusted for historical regional differences. More formally stated

$$V = \lambda_V\ NV\ (V^u/NV^u) \quad (8)$$

and

$$NV = VG\ (G65) + VL\ [L65 - E\ (1 + RA/YLP)] + [N - E\ (1 + RA/YLP)], \quad (9)$$

where VG are per capita transfer payments for persons over 65 years old relative to per capita transfer payments for all persons not working, VL are per capita transfer payments for persons younger than 65 who are not working relative to per capita transfers for all persons not working, $\lambda_V$ adjusts for regional differences and is calculated in the last historical year by solving equations (8) and (9), and E and N are, respectively, total employment and population in the region.

Finally, personal income, YP, in the region is

$$YP \equiv YLP - TWPER + RA + YPROP + V,$$

and real disposable income, RYD, for the region is then calculated as

$$RYD = (YP - TAX)/CP, \quad (10)$$

where TAX denotes the rest of the deductions (primarily local, state, and federal income taxes) from personal income, and CP is the personal consumption expenditure deflator.

The variable TAX depends on net income after subtracting transfer income. It is adjusted for regional differences by $\lambda_{TAX}$ and changes as U.S. tax rates change.

$$TAX = \lambda_{TAX}\ (YP - V)\ [TAX^u/(YP^u - V^u)].$$

*Investment equations.* There are three types of investment: resi-

dential, nonresidential, and equipment investment. In a multi-industry model, it is necessary to allocate investment to the industries supplying the investment goods and carrying out the construction. Thus,

$$I_{ij} = INV_{ij}\ I_j, \tag{11}$$

where $INV_{ij}$ is a coefficient denoting the proportion of investment category j supplied by industry i, and $I_{ij}$ is type j investment demand for output from industry i.

In the version of the model used for this article, the equation for nonresidential and equipment investment is

$$I_j = \lambda_j\ (K_j^*/K_j^{u*})\ I_j^u, \tag{12}$$

where the coefficient $\lambda_j$ reflects higher or lower nonresidential and equipment investment in the region in the last historical year than indicated by the rest of the equation and partially compensates for the absence of a stock adjustment process when the growth differential from the United States is constant. $I_j^u$ is total investment of type j capital in the nation, and $K_j^*/K_j^{u*}$ is a ratio of local to national optimal capital stock for type j capital, whose calculation will be described in the section on factor demands.

The equation for residential investment is

$$I_r = \lambda_r\ (ARYD/ARYD^u)\ I_r^u, \tag{13}$$

where $\lambda_r$ is a coefficient to reflect regional differences. ARYD is expected permanent real disposable income,

$$ARYD_t = ARYD_{t-1} + \lambda_{ARYD}\ (RYD_t - ARYD_{t-1}),$$

where $\lambda_{ARYD} = 0.25$, t denotes the current period, and t − 1 denotes the previous period.

In the alternative version of the model, released in 1991, investment is determined from a stock adjustment process (Rickman, Shao, and Treyz, Forthcoming). The investment equation is

$$I_{jt} = \alpha_j\ [K_{jt}^* - K_{j0} \prod_{i=1}^{t}(1 - d_i) - \sum_{i=1}^{t-1} I_{ji} \prod_{k=i+1}^{t} (1 - d_k)], \tag{14}$$

where $I_{jt}$ is type j investment at year t, $\alpha_j$ is an adjustment coefficient for capital stock estimated over all states (the value of $\alpha_j$ is .127 for residential and .061 for nonresidential investment), $K_{j0}$ is the initial capital stock, and $d_i$ is the depreciation rate. $K_{j0}$ and a capital preference parameter are region-specific estimates. The optimal capital stock for the region is calculated as a share of the U.S. optimal capital stock, $K_{jt}^{u*}$, which in turn is determined from

$$K_{jt}^{u*} = (I_{jt}^u/\alpha_j) + (1 - d_{jt})\ K_{j(t-1)}^u. \tag{15}$$

The regional share for optimal nonresidential capital stock, $K_j^*/K_j^{u*}$, is calculated in the factor demand section below. The share for residential investment is the region's share of real disposable income, RYD, and equipment investment is predicted to be in the same ratio to nonresidential investment in the region as it is in the nation.

*Government spending equations.* Government spending is predicted for six components: federal civilian, federal military, and state and local expenditures for education, health and welfare, safety, and miscellaneous. Federal government civilian and military employment in the region are exogenous and maintained at a fixed share of the corresponding total U.S. values. Federal military procurement is allocated according to each region's representation in the industries for which the federal military spending occurs.

The expenditures of state and local government depend on relative population, nationally predicted state and local government expenditures, and regional differences ($\lambda_G$).

$$G_i = \lambda_G \, (N/N^u) \sum_{j=1}^{4} GOV_{ij} \, G_j^u, \tag{16}$$

where $GOV_{ij}$ is the proportion of component j of government demand for industry output i. State and local governments maintain per capita services and follow national per capita government expenditure trends.

## Block 2 — Factor Demands

Industries demand profit-maximizing levels of factor inputs. The optimal choice of inputs involves two stages. First, industries demand fixed shares of composite value added and composite intermediate inputs. Second, industries choose optimal levels of the components of the composite factors. The shares of the individual intermediate inputs of the composite intermediate input are fixed. A Cobb-Douglas constant-returns-to-scale function relates the value added components to the composite value added factor as follows:

$$VA_i = A_i \, (E_i)^{\alpha_i} \, (K_i)^{\beta_i} \, (F_i)^{\gamma_i},$$

where $A_i$ is total factor productivity, $E_i$ is labor, $K_i$ is a composite of capital factors, $F_i$ is a composite fuel factor, and $\alpha + \beta + \gamma = 1$.[4]

Though more flexible forms than Cobb-Douglas exist, Griffen and Gregory (1976) show that the Cobb-Douglas production function

[4] E and K are used interchangeably for employment and capital and for demands for labor and capital because wages and the cost of land are determined by reduced-form equations, and because the costs of capital and fuel are not related to the levels of usage. Thus, for given factor prices, actual factor usages are calculated from the factor demand equations.

provides a good characterization of substitutability between fuel and other inputs. Also, since regional production technology is assumed to be identical to national technology, the historical changes in U.S. value added shares and projected changes during the forecast period can be directly incorporated into the regional Cobb-Douglas functional form.

*Labor demand.* Homotheticity of production yields profit maximizing factor intensities that are independent of output and can be obtained by cost minimization techniques. Demand for labor can be expressed as

$$E_i = (1/A_i)\,(w_i/\alpha_i)^{\alpha_i - 1}\,(c_i/\beta_i)^{\beta_i}\,(f_i/\gamma_i)^{\gamma_i}\,VA_i, \tag{17}$$

where $w_i$ is the wage rate, $c_i$ is the cost of capital, and $f_i$ the cost of fuel. Deriving a similar expression for the United States and dividing it into equation (17) yields

$$E_i = (1/RFPROD_i)\,(RLC_i)^{\alpha_i - 1}\,(RCC_i)^{\beta_i}\,(RFC_i)^{\gamma_i}\,(E_i^u/VA_i^u)\,VA_i, \tag{18}$$

where $RFPROD_i$ is the ratio of total factor productivities $(A_i/A_i^u)$, $RLC_i$ represents relative labor cost (defined as the regional wage rate divided by the national wage rate) adjusted for four-digit industry mix differences within the two-digit industry, $RCC_i$ represents relative capital cost, and $RFC_i$ is relative fuel cost.

Due to the costs of adjusting factor intensities, the demand for labor may not be the optimum demand. From equation (18), optimum labor intensity, $h_i$, is defined as

$$h_i \equiv \frac{E_i/E_i^u}{VA_i/VA_i^u}\,(RFPROD_i) = (RLC_i)^{\alpha_i - 1}\,(RCC_i)^{\beta_i}\,(RFC_i)^{\gamma_i}. \tag{19}$$

Labor intensity, $\ell_i$, is defined as a geometrically declining weighted average of current and past adjustments to deviations of the factor intensities from optimum factor intensities. Thus,

$$\ell_{it} = \ell_{i(t-1)} + \lambda_\ell\,(h_{it} - \ell_{i(t-1)}), \tag{20}$$

where $\lambda_\ell$ is the partial adjustment coefficient and equals 0.0625 on the assumption that the average lifetime of equipment is 13 years (see Coen 1975). Substituting $\ell_i$ for $h_i$ as defined in equation (19) into equation (18) gives

$$E_i = (1/RFPROD_i)\,\ell_i\,(E_i^u/VA_i^u)\,VA_i. \tag{21}$$

*Capital demands.* Optimal capital stock, which drives investment in Block 1, is calculated for the aggregate of nonresidential structures and equipment. Substituting actual aggregate employment into the Cobb-Douglas expansion path between capital and labor gives

$$K^* = \left(\frac{\Sigma\,KW_iRLC_i}{\Sigma\,KW_iRCC_i}\right)\left(\frac{AEF}{AEF^u}\right)K^{u*},$$

where K* denotes optimum capital stock, $KW_i$ reflects the capital intensity of i for the United States, and AEF denotes anticipated employment, which is also weighted by $KW_i$. The weighting produces greater capital demands when capital intensive firms are stimulated than if labor intensive firms are stimulated.

The optimal capital stock becomes forward-looking by specifying anticipated employment according to an adaptive expectations framework.

$$AEF_t = AE_t + CAAE_t$$

and

$$CAAE_t = CAAE_{t-1} + \lambda_{CAAE}\,[(AE_t - AE_{t-1}) - CAAE_{t-1}],$$

where $CAAE_t$ is the expected change in employment, $AE_t$ is current employment weighted by $KW_i$, and $\lambda_{CAAE}$ denotes the rate of adjustment and equals 0.33. For the alternative version of the model that includes the stock adjustment equation, $AEF_t$ is set equal to $AE_t$.

*Demand for fuel.* While demand for fuel is not explicit in the model, the cost of fuel enters the demands for labor and capital in equations (17) through (21) and plays an important role in the model. The treatment of fuel is unique in that the detailed intermediate outputs for coal mining, crude petroleum and natural gas mining, petroleum refining, and electric and natural gas utilities are excluded from the intermediate industry transactions and treated as value added. As a value added factor, fuel becomes an input that is a substitute for capital and labor.

## Block 3 — Population and Labor Supply

Regional population and labor supply determine the interaction of the demographic section of the model with the economic section. The demographic section applies a cohort algorithm to the population for single-year age cohorts for both males and females. The cohort algorithm predicts the number of births and deaths that occur and, by taking the difference between them, gives the natural change in population. Combined with potential labor force participation rates, the labor force is built up from the cohorts. The indigenous labor force competes with potential migrants for jobs. Migrants in turn alter the long-term demographic structure of the population. The demographic-economic interaction becomes pronounced for rapidly growing or declining regions, and models that ignore this interaction may be misleading. The demographic section also calculates migration exogenous to regional economic conditions.

*Cohort algorithm.* The cohort algorithm applies fertility and survival rates to the appropriate cohorts while adding births to and

subtracting deaths from each corresponding cohort. Two simple equations summarize the process:

$$BTH_t = \sum_{k \in F} COH_{k(t-1)} \, FER_{k(t-1)} \tag{22}$$

$$COH_{kt} = COH_{k(t-1)} \, SRV_{k(t-1)} \qquad \text{for all } k, \tag{23}$$

where BTH is the total number of births, COH is the size of the age and sex cohort k, FER is the probability of giving birth, SRV is the probability of surviving from one specific age cohort to the next, and F denotes the sample of women in child-bearing years. For k equal to one, SRV captures both the probability of a successful birth and the probability of reaching one year of age.

Fertility and survival rates are state-specific for 1980 and were trended forward and backward based on national trends observed and projected by the Bureau of the Census. Historically predicted births and deaths by cohort also are proportionately adjusted to obtain regional total birth and death rates. The forecast period retains the birth and death rate adjustments calculated in the last year of history. Also, since military personnel and their dependents are treated as a special population, the cohort algorithm applies only to the civilian population.

The potential labor force, NLF, is calculated by

$$NLF = \sum_{k} NPR_k \, COH_k, \tag{24}$$

where $NPR_k$ is the potential participation rate for the cohort. To determine $NPR_k$, a cross-section regression was run across 51 regions for each of 40 aggregated cohorts (20 for males and 20 for females) for 1980. The most relevant finding was that the estimated coefficient for the unemployment rate was always negative, which implies that high unemployment rates cause discouraged potential workers to withdraw from or stay out of the labor force. When the overall national unemployment rate is near the natural rate, replacing the local unemployment rate for a particular age and sex group with a national rate for the same group generates a measure of the underlying potential labor force. Predicted 1980 participation rates were trended backward and forward based on historical and projected U.S. labor force participation rate trends for each cohort.

*Migrants.* International migrants, retired migrants, former military personnel and their dependents reentering the civilian population, and economic migrants are the four components of net migration. All but economic migrants are exogenous to the economic sectors of the model.

International migration is determined by calculating a fixed regional share of U.S. immigration, which is projected by the Bureau of the Census based on its reports of international migration patterns

by state from 1975 to 1980. The general state migrant distribution serves as the distribution of international migrants. The distributed migrants become part of the receiving region's population, to which the region's cohort algorithm applies. Survival and fertility rates apply to one-half of the migrants to reflect an average of one-half year of residency for each migrant during the year of migration.

Retired migrants are defined as migrants over age 65 and include international immigrants. Retired migrants respond to noneconomic factors such as differential regional amenity levels. Rates of migration for retirees are calculated for single-year age cohorts for both males and females and are defined as the sum of the cohort's residually calculated migrants between the two most recent censuses (1970 and 1980) divided by the sum of the cohort in the region if total migrants are negative, and the sum of the cohort in the United States if total migrants are positive.

$$RTMG = \sum_k rm_k \, [(1 - RTDUM_k) \, COH_k + RTDUM_k \, COH_k^u], \quad (25)$$

where RTMG denotes total retired migrants, $rm_k$ is the rate of migration for cohort k, $RTDUM_k$ equals one if $rm_k$ is positive and zero if $rm_k$ is negative, and COH is the size of the cohort. Survival and fertility rates apply to one-half of the migrants to reflect an average of one-half year of residence for each migrant.

Because military personnel and their dependents are treated as a special population that includes overseas personnel, a link is required between regional changes in the civilian population and the changes in the size of the military. The share of civilian population that a region receives when the size of the military decreases, or loses when military size increases, is determined by the regional proportion of total domestic population in the previous year. For personnel returning from domestic duty, associated dependents are also included. Again, the cohort algorithm applies to one-half of the returning personnel and their dependents.

Economic migrants are migrants under age 65 who were part of the civilian population in the United States the preceding year. Economic migration, ECMG, responds to both economic and amenity factors and may be expressed as

$$ECMG = f\left(\frac{EY}{EY^u}, \frac{\mathcal{A}}{\mathcal{A}^u}\right), \quad (26)$$

where EY is expected income, and $\mathcal{A}$ denotes the level of amenities.

Expected income is the probability-weighted sum of wages,

$$EY = \sum_i P(E_i) \, w_i, \quad (27)$$

where $P(E_i)$ denotes probability of employment. This formulation is

consistent with the theoretical model by Harris and Todaro (1970). Breaking the probability into two parts so that

$$P(E_i) = P(E) \times P(E_i | E),$$

and specifying forecasts for the two parts as

$$P(E) = ER/NLF$$

and

$$P(E_i | E) = E_i/E,$$

equation (27) can be rewritten as

$$EY = (ER/NLF) \left[ \sum_i (E_i/E)\, w_i \right],$$

where ER is civilian employment adjusted for place-of-residence, NLF is the potential labor force from equation (24), and E is total nonfarm private employment by place of work.

Adjusting $w_i$ for regional cost-of-living and tax differentials, dividing by a similar expression for $EY^u$, and multiplying the numerator and denominator by $\Sigma_i(E_i/E)w_i^u$ yields

$$\begin{aligned} \frac{EY}{EY^u} &= \frac{ER/NLF}{ER^u/NLF^u} \times \frac{\sum_i (E_i/E)\, w_i\, (RYD/YP)}{\sum_i (E_i/E)\, w_i^u\, (RYD^u/YP^u)} \times \frac{\sum_i (E_i/E)\, w_i^u}{\sum_i (E_i^u/E^u)\, w_i^u} \\ &= \quad REO \quad \times \quad RWR \quad \times \quad RWM, \end{aligned} \tag{28}$$

where REO, RWR, and RWM represent relative employment opportunity, relative wage rate, and relative wage mix, respectively. Thus, REO reflects the probability of getting a job, RWR the real after-tax wage independent of regional industry mix differences, and RWM the regional mix effect on average pay.

Substituting equation (28) into equation (26) gives

$$ECMG = f\,(REO \times RWR \times RWM,\ \mathcal{A}/\mathcal{A}^u). \tag{29}$$

Equation (29) is estimated by assuming a semi-log functional form, normalizing ECMG to the previous year's potential labor force, incorporating lagged migration responses of up to two periods, and adding a random disturbance term. The estimated equation is

$$\begin{aligned} NECM_t = {} & \ln(\lambda_k) + \delta_1 \ln(REO_t) + \delta_2 \ln(REO_{t-1}) \\ & + \delta_3 \ln(REO_{t-2}) + \delta_4 \ln(RWR_t) + \delta_5 \ln(RWR_{t-1}) \\ & + \delta_6 \ln(RWR_{t-2}) + \delta_7 \ln(RWM_t) + \delta_8 \ln(RWM_{t-1}) \\ & + \delta_9 \ln(RWM_{t-2}) + \mu_t, \end{aligned} \tag{30}$$

where $\lambda_k$ denotes the region-specific relative amenity value obtained

by specifying a dummy variable for each region, and $NECM_t = ECMG_t/NLF_{t-1}$.[5] Because equation (30) is part of a larger system of equations that specifies REO as an explanatory variable of the wage rate, and ECMG as a component of REO, the equation was estimated with two-stage least squares.[6] Also, a polynomial-distributed lag of degree one was imposed on the lag weights for each variable to get desired dynamic model properties. The results of the estimation, with corresponding t-statistics in parentheses, for the 1973–88 period for 50 states plus Washington, D.C., are

$$\begin{array}{lll} \delta_1 = 0.296\ (8.8) & \delta_2 = 0.121\ (12.0) & \delta_3 = -0.054\ (-2.5) \\ \delta_4 = 0.320\ (7.1) & \delta_5 = 0.117\ (8.0) & \delta_6 = -0.086\ (-3.4) \\ \delta_7 = 0.196\ (2.2) & \delta_8 = -0.004\ (-0.2) & \delta_9 = -0.206\ (-2.6) \\ R^2 = 0.63 & & \end{array}$$

The results confirm the importance of each economic factor. For substate areas, equation (30) and the estimates for $\delta_1$ through $\delta_9$ are used to calibrate a region-specific intercept.

## Block 4 — Production Costs, Prices, and Profits

*Production costs.* Total cost of production, $TC_i$, is the sum of the input costs weighted by their usage,

$$TC_i \equiv w_i E_i + c_i K_i + f_i F_i + \sum_j sp_j Q_{ij}, \tag{31}$$

where $sp_j$ is the price of material input j. Substituting the expansion path expressions for labor with capital and labor with fuel into equation (31) and then substituting equation (17) for $E_i$ gives the cost function written as

$$TC_i = (VA_i/A_i)\,(w_i/\alpha_i)^{\alpha_i}\,(c_i/\beta_i)^{\beta_i}\,(f_i/\gamma_i)^{\gamma_i} + \sum_j sp_j\, Q_{ij}. \tag{32}$$

Dividing equation (32) by $Q_i$, deriving a similar expression for the United States, normalizing $A_i^u = (w_i^u/\alpha_i)^{\alpha_i}(c_i^u/\beta_i)^{\beta_i}(f_i^u/\gamma_i)^{\gamma_i}$ and $sp_j^u = 1$, and then multiplying the first term in equation (32) by $A_i^u/A_i^u$ (using the normalized expression for $A_i^u$ in the numerator) gives

$$P_i = AC_i/AC_i^u = af_i\,(1/RFPROD_i)\,(RLC_i)^{\alpha_i}\,(RCC_i)^{\beta_i}\,(RFC_i)^{\gamma_i} + \sum_j a_{ij}\, sp_j, \tag{33}$$

where $P_i$ denotes relative production costs, $AC_i$ is average cost, $af_i$ is $VA_i/Q_i$, $a_{ij}$ is $Q_{ij}/Q_i$, and relative factor productivity ($RFPROD_i$) is

[5] For a comparison of calculated amenity effects from this equation with those of hedonic models, see Greenwood et al. (1991).

[6] A more thorough description of the construction of the variables, tests of alternative lag lengths, and the use of exogenous variables in the first stage of the estimation can be found in Treyz et al. (Forthcoming).

$A_i/A_i^u$. The $VA_i$, $Q_{ij}$, and $Q_i$ values are proxied by their U.S. counterparts. Due to constant returns to scale, $AC_i/AC_i^u$ can be replaced by marginal costs, $MC_i/MC_i^u$. The relative component costs of labor ($RLC_i$), capital ($RCC_i$), fuel ($RFC_i$), and intermediate inputs ($sp_j$) are described below in more detail.

*Relative cost of labor.* The relative cost of labor is

$$RLC_i = w_i/(w_i^u \; WINDX_i), \tag{34}$$

where $WINDX_i$ is an index to control for the effect of area-specific differences in the proportions of four-digit industries within two-digit industry i. An area concentrated with high-wage jobs in industry i will not show a high relative labor cost if the wages for the four-digit industries are at the national average.

Prediction of $w_i$ is given by

$$w_{it} = \left(\sum_{j=1}^{94} d_{ij} \, \Delta WD_j + \tau \, \Delta CP_{t-1} + k^u\right) w_{i(t-1)}, \tag{35}$$

where $d_{ij}$ is the proportion of occupation j in industry i projected by the Bureau of Labor Statistics, $\Delta WD_j$ is the change in wage rate due to changes in demand for occupation j and changes in labor market conditions, $\Delta CP_{t-1}$ is the change in the relative consumer expenditure deflator from time t−2 to t−1, $\tau$ is a parameter that is estimated as described below, and $k^u$ is the change in U.S. wages not explained by the rest of equation (35) and industry mix changes.

Wage rate changes due to occupational wage rate changes are predicted by

$$\begin{aligned}\Delta WD_j = {} & \theta_1 \, (EO/EOA) \, (1 - DHSK) + \theta_2 (EO/EOA) DHSK \\ & + \theta_3 \, (OD_j/ODA_j) \, DHSK + \mu,\end{aligned} \tag{36}$$

where EO is employment in the region divided by the potential labor force, DHSK is a dummy variable equal to one for high-skill occupations and zero for low-skill occupations, $OD_j$ is demand for occupation j originating from all firms in the region and $\mu$ is a random disturbance term. EOA and $ODA_j$ are geometrically declining weighted averages of EO and $OD_j$.

$$EOA_t = EOA_{t-1} + \lambda_{EO} \, (EO_{t-1} - EOA_{t-1}),$$

and

$$ODA_{jt} = ODA_{j(t-1)} + \lambda_{OD} \, (OD_{j(t-1)} - ODA_{j(t-1)}), \tag{37}$$

where $\lambda_{EO}$ and $\lambda_{OD}$ are coefficients of adjustment.

The general methodology for estimating $\theta_1$, $\theta_2$, and $\theta_3$ follows that of Treyz and Stevens (1985), which combined data from the 1980–81 *Current Population Survey* with data from REMI. Because

more recent research uses 1986–87 *Current Population Survey* data, the estimated coefficient values differ from those reported earlier. The independent variables EO and OD are generated with the model and linked with the dependent variable from the *Current Population Survey* (U.S. Bureau of the Census 1982, 1988). Variables that reflect traits specific to individuals (including race, gender, level of education, age, marital status, union membership, and whether the person was the head of a household) are added to the regression to control for their influence on wage rate changes. Also, the 1986 wage rate is added to control for the influence of wage levels on wage rate changes. The most significant relationships were found with $\lambda_{OD}$ = 0.2, $\lambda_{EO}$ = 1.0 for high-skill occupations, and $\lambda_{EO}$ = 0.2 for low-skill occupations. The estimated coefficients for the economic variables (with corresponding t-statistics) are $\theta_1$ = 0.393 (2.767), $\theta_2$ = 1.461 (3.251), and $\theta_3$ = 0.136 (1.803).

Wages of both high- and low-skill occupations are sensitive to general market conditions. The impact of market conditions is more immediate for highly skilled workers because the most significant relationship was found when the coefficient of adjustment equaled one, as compared with one-fifth for low-skilled workers, for whom the effect is less immediate. Occupational demand is not a significant factor in explaining wage changes for low-skilled workers but is significant for highly skilled workers.

A final step in the estimation required substituting equation (36) into the model to predict wage changes. New estimates were obtained by minimizing the sum of squared errors of predicted wages to actual wages with the errors normalized to predicted wages. $\theta_1$ and $\theta_2$ were uniformly increased to twice their initial values and decreased to zero in increments of one-eighth their initial values for each trial. $\theta_3$ was similarly estimated given the final estimates of $\theta_1$ and $\theta_2$. The sample included all fifty states plus Washington, D.C., for 1970 to 1987. Final estimates for $\theta_1$ and $\theta_2$ were one-half their initial values, while no improvement could be found for $\theta_3$. A joint F-test on the three coefficients rejected that they were all zero with a level of significance of 0.05.

Given the final estimates of $\theta_1$, $\theta_2$, and $\theta_3$ and using the approach described above, $\tau$ is estimated to equal 0.25. Based on an F-test, $\tau$ equal to zero is rejected at the 0.05 level of significance. Thus, 25 percent of changes in the relative regional cost of living are passed on into regional wages.

*Relative cost of capital.* The derivation of the cost of capital is based on the implicit rental cost of capital approach and draws on the work of many researchers, especially Hall and Jorgenson (1967). The form used here is given in Treyz and Stevens (1983). Three types of capital are considered, and their costs are related to aggregate capital cost according to a Cobb-Douglas cost function.

$$c_i = (CSTR_i\ PSTR)^{CS_i}\ (CEQP_i\ PEQP)^{CE_i}\ CINV_i^{CI_i}\ CLAND^{CL_i}, \quad (38)$$

where $CSTR_i$, $CEQP_i$, and $CINV_i$ are the relative costs of structures, equipment, and inventory; PSTR and PEQP are the relative prices of structures and equipment; CLAND is the cost of land, which is proxied by the relative price of housing; and $CS_i$, $CE_i$, $CI_i$, and $CL_i$ (which sum to equal one) are the proportions of capital accounted for by structures, equipment, inventory, and land, respectively.

Since equipment and structures are not explicit in the model, their relative prices require separate calculation. The equations for PEQP and CEQP are

$$PEQP = \sum_i sp_i\ CWEQP_i, \quad (39)$$

where $sp_i$ is the price paid by a local firm for the *i*th input into equipment, and $CWEQP_i$ is the proportion of all inputs to equipment represented by the input from the *i*th industry as obtained from the U.S. input-output table, and

$$CEQP_i = \frac{r + CERR}{1 - UM_i}\left[1 - TIC^u - (1 - TCP^u)\ TIC - DDFE - DDME_i - RDME + \frac{WM}{(r + CERR)}\right] \Big/ CEQP_i^u, \quad (40)$$

where r is the interest rate and is assumed equal across regions, CERR is the rate of replacement of equipment capital, $UM_i$ is the combined national and local corporate profit tax rate for industry i, TIC ($TIC^u$) is local (national) investment tax credit, $TCP^u$ is national corporate profit tax rate, DDFE is present value of federal tax savings from depreciation deductions on equipment, $DDME_i$ is present value of local tax savings from depreciation deductions on equipment, RDME is present value of local tax savings from interest deductions on equipment, and WM is equipment tax rate less the federal deduction. $CEQP_i^u$ is calculated the same as $CEQP_i$ (without the division by $CEQP_i^u$) except local variables are replaced with national averages. To calculate PSTR and CSTR, all terms specific to equipment can be replaced by corresponding terms for structures.

The present value of federal tax savings from depreciation deductions can be written as

$$DDFE = (1 - XSYDE)\frac{TCP^u}{TEL^u}\frac{1 - \exp(-TEL^u\ r)}{r} + XSYDE\frac{2\ TCP^u}{TEL^u\ r}\left[1 - \frac{1 - \exp(-TEL^u\ r)}{TEL^u\ r}\right], \quad (41)$$

where XSYDE is the proportion of firms using sum-of-the-years digits

depreciation for equipment, and $TEL^u$ is equipment lifetime according to national tax laws.

Likewise, the present value of local tax savings from depreciation deductions is

$$DDME_i = (1 - XSYDE)\frac{VM_i}{TELM}\frac{1 - \exp(-TELM\ r)}{r} + XSYDE\frac{2\ VM_i}{TELM\ r}\left[1 - \frac{1 - \exp(-TELM\ r)}{TELM\ r}\right], \tag{42}$$

where TELM is equipment lifetime allowed by the state, and $VM_i$ is the state corporate tax rate net the effect of the federal deduction of state taxes.

$$VM_i = TCP_i - TCP^u\ TCP_i, \tag{43}$$

$$TCP_i = \lambda_{TCP}\ Z_{TCP(i)}, \tag{44}$$

$$UM_i = TCP^u + VM_i, \tag{45}$$

$$WM = TEQP\ (1 - TCP^u), \tag{46}$$

where $TCP_i$ is local corporate profit tax rate for industry i, $\lambda_{TCP}$ is a coefficient that reflects the corporate profit tax rate from the last year of history, $Z_{TCP(i)}$ is an adjustment for industries whose state corporate profit tax rate differs from the average rate for all industries, and TEQP is the tax rate on equipment. Again, all terms specific to equipment can be replaced by those specific to structures.

Inventory costs are calculated as

$$CINV_i = \frac{[r/(1 - UM_i)]\ (1 - UM_i\ B)}{CINV_i^u},$$

where B is the proportion of business capital financed by bonds and loans. $CINV_i^u$ can be calculated by replacing $UM_i$ with a national average and eliminating the denominator.

*Relative fuel costs.* Relative fuel costs, $RFC_i$, are based on the cost of fuel for commercial or industrial users as appropriate for industry i. The functional form for fuel substitutability is Cobb-Douglas and is written as

$$RFC_i = \prod_{j=1}^{3} (RFE_j)^{\eta_j}, \tag{47}$$

where $RFE_j$ is the exogenously determined relative fuel cost for electricity, residual fuel oil, and natural gas; and $\eta_j$ is the share of fuel category j in industry i.

*Prices.* Goods produced by industries are assumed to be sold

primarily in either national markets or local regional markets.[7] Goods and services in industries that serve national markets are also assumed to be priced at the average national price to be competitive, therefore $sp_i = 1$ for these industries. The relative marginal costs of production determine the relative prices of goods and services that are regional in nature. Thus, $sp_j = P_j$ from equation (33) for these industries. The price of housing is predicted according to changes in the regional share of U.S. population, which is a proxy for changes in population density.

The elasticity of response of housing prices to population is estimated by

$$\text{DPHOUS} = \varepsilon \text{ DPOPSH}, \tag{48}$$

where DPHOUS and DPOPSH are one-period growth rates in the relative price of housing and in a region's share of the U.S. population; $\varepsilon$ was estimated to equal 0.445 with a t-statistic equal to 4.079.

Both national and regional industries produce for export and local demand. For regional industries, the delivered price of imports equals the price charged by local producers. The regional personal consumption deflator can be constructed as

$$CP = CP^u \sum_i pc_i \, sp_i, \tag{49}$$

where $pc_i$ is the weight of industry i in the personal consumption column of the input-output matrix.

*Profits.* Due to constant returns to scale, profits can be defined as profit per unit of output. Recalling that for national industries a uniform price exists across regions ($P_i = P_i^u = 1$), normalizing average profits for the nation to equal zero ($P_i^u = AC_i^u$), and subtracting equation (33) from $P_i^u$ gives relative profit per unit of output, $\pi_i$, as

$$\begin{aligned} \pi_i = \Pi_i/Q_i &= AC_i^u - AC_i \\ &= 1 - \frac{af_i \, (RLC_i)^{\alpha_i} \, (RCC_i)^{\beta_i} \, (RFC_i)^{\gamma_i}}{RFPROD_i} \\ &\quad + \sum_j a_{ij} \, (1 - sp_j). \end{aligned} \tag{50}$$

This formulation does not imply that profits are zero for industry i in the United States, only that they can be normalized to zero when relative profits are calculated. Relative profits are thus directly related to relative factor productivity and inversely related to relative input costs.

[7] Based on the REMI procedure for calculating exports in 1977, an industry is categorized as national if the sum of exports from all states equals one-half or more of total production in that industry. National industries include all manufacturing industries except stone, clay, and glass; printing; and petroleum production. Hotels, in the service industry, are also included.

The only term in equation (50) yet to be calculated is $RFPROD_i$. Rearranging terms in equation (21) gives

$$RFPROD_i = \frac{VA_i/E_i}{VA_i^u/E_i^u}\,\ell_i.$$

The term $\ell_i$ incorporates an often neglected aspect of the productivity calculation. Moomaw (1981) showed the importance of incorporating differences in factor intensities into productivity calculations, and Malhotra and Garofalo (1988) reported that Hulten and Schwab's (1984) findings for regional patterns of productivity growth in the United States are reversed if capital is treated as a quasi-fixed factor in production because actual and optimum intensities diverge.

Profitability is not calculated for regional industries because productivity differentials reflect location advantages that do not affect production at the margin. Productivity calculations for regional industries also may be more suspect at the two-digit level because of the greater heterogeneity of regional industries.

Block 5 — Market Shares

For a one area model, equation (2) can be written more succinctly as

$$Q_i = R_i\,D_i + S_i^{ku}\,X_i^u, \tag{51}$$

where $D_i$ is the sum of local intermediate demands and local final demands, $X_i^u$ is interregional trade in the United States, including U.S. exports. The market shares $R_i$ and $S_i^{ku}$ reflect the relative competitiveness of the industry in the region.

Most factors that influence a region's competitiveness are difficult to quantify. Therefore, the approach here uses cross-sectional methodologies to calibrate equation (51) to a base year (the last historical year for which the required data is available). The calibration produces initial estimates of the market shares that presumably reflect cross-sectional variation in location advantages. Changes in the market shares reflect changes in economic conditions that affect the competitiveness of industries producing in the region and, at the margin, induce firm migration or market expansion by existing firms.

The 1977 *Census of Transportation* and generated supply-to-demand ratios provide the necessary inputs to estimate the regional purchase coefficient, $R_{jT}$ at the 466-sector level.[8] The equation is given by

$$R_{jT} = PS_{j77}\,\frac{Q_{jT}}{D_{jT}}, \tag{52}$$

[8] A complete description of the technique and a report of its comparability to survey techniques can be found in Stevens et al. (1983).

where $PS_{j77}$ is the percent of output at the four-digit level shipped to destinations within the region. It is reported by the 1977 *Census of Transportation* (U.S. Department of Commerce 1982) for manufacturing industries and subjectively estimated for nonmanufacturing industries. The REMI 466-sector input-output model generates $D_j$ in base year T, and j denotes the industry at the four-digit level. A cross-section equation (Treyz and Stevens 1985) was estimated with 1977 data at the four-digit level and aggregated to the two-digit level. Any errors at the two-digit level for which observed values existed are maintained when the equation in the base year is applied to convert regional purchase coefficients from four digits, $R_{jT}$, to two, $R_{iT}$.

Given $R_{iT}$, exports are calculated as a residual, and the base year export share, $S^{ku}_{iT}$, is given by

$$S^{ku}_{iT} = \frac{Q_{iT} - R_{iT}\, D_{iT}}{X^{u}_{iT}} = \frac{X^{ku}_{iT}}{X^{u}_{iT}} . \tag{53}$$

For multi-area models, $S^{kr}_i$ can be obtained by rearranging terms in equation (3) such that

$$S^{kr}_i = X^{kr}_i \Big/ \left( \sum_{h=1, h\neq k}^{m} M_{ih} \right).$$

Multi-area exports, $X^{kr}_i$, in the base year are

$$X^{kr}_{iT} = PS^{mr}_{iT}\, Q^{k}_{iT} - R^{k}_{iT}\, D^{k}_{iT},$$

where superscript mr indicates the multi-area region and k, the area. The behavioral responses for $S^{kr}_i$ are based on those for $S^{ku}_i$.

Equation (53) exactly predicts output in year T. Outputs might be predicted in other years using equation (51) and the calibrated market shares. For given values of local and export demands, however, predicted changes in market shares may reduce the differences between measured and predicted output. Changes in market shares should in turn be induced by changes in economic conditions, but the variables that explain changes in market shares differ for national and regional industries. A complete presentation of the approach is in Rickman and Treyz (1992).

*National industries.* If a national industry becomes more profitable in a region, local firms expand or firms from other regions relocate to the region to serve both local and export markets. The local share of local demand can change as the national propensity to import changes, and market shares at the two-digit level can change because

of a changing composition of three-digit industries. Equations reflecting the above hypotheses are

$$R_{it} = R_{iT} \left(\frac{RPROFA_{it}}{RPROFA_{iT}}\right)^{\phi_1} \left(\frac{R^u_{it}}{R^u_{iT}}\right)^{\phi_2} \left(\frac{IMIX_{it}}{IMIX_{iT}}\right)^{\phi_3} \tag{54}$$

and

$$S^{ku}_{it} = S^{ku}_{iT} \left(\frac{RPROFA_{it}}{RPROFA_{iT}}\right)^{\psi_1} \left(\frac{IMIX_{it}}{IMIX_{iT}}\right)^{\psi_2}, \tag{55}$$

where RPROFA is a geometrically declining weighted moving average of an index of relative profitability ($RPROF_i = \pi_i + 1$), $R^u_i$ denotes the U.S. regional purchase coefficient, and IMIX denotes industrial mix that captures differential regional representation in slow and fast growing three-digit industries within the two-digit industry. The equations for RPROFA and IMIX are

$$RPROFA_{it} = RPROFA_{i(t-1)} + 0.2\,(RPROF_{i(t-1)} - RPROFA_{i(t-1)})$$

and

$$IMIX_{it} = \frac{\sum_{j \in i} (E_{jT}/E_{iT})\,(E^u_{jt}/E^u_{jT})}{\sum_{j \in i} (E^u_{jT}/E^u_{iT})\,(E^u_{jt}/E^u_{jT})},$$

where j denotes the three-digit element within two-digit industry i.

Substituting equations (54) and (55) into equation (51) gives a time-series equation for industry i. The time-series equations for all national industries, for all states plus Washington, D.C., and for the period 1969–87 were pooled together. Based on prior beliefs, the following restrictions were imposed: $\phi_1 = \psi_1$, and $\phi_2 = \phi_3 = \psi_2 = 1$. Thus, changing profitability affects the market shares proportionately. A 1-percent increase in the national propensity to import increases all the regions' propensities to import by 1 percent, and a 1-percent increase in industrial mix increases the market shares by 1 percent. Given the restrictions, taking the natural logs of both sides of the equation, and adding a random disturbance term gives

$$\ln (Q_{it}/RHS_{it}) = \phi_1 \ln (RPROFA_{it}/RPROFA_{iT}) + \varepsilon_t,$$

where

$$RHS_{it} = R_{iT} \left(\frac{R^u_{it}}{R^u_{iT}}\right) \left(\frac{IMIX_{it}}{IMIX_{iT}}\right) D_{it} + S^{ku}_{iT} \left(\frac{IMIX_{it}}{IMIX_{iT}}\right) X^u_{it}.$$

The transformed dependent variable, $Q_{it}/RHS_{it}$, can be interpreted as the change in output not explained by changes in total demand, international competitiveness, or industrial mix. An ordinary least squares regression on the pooled data gave $\phi_1 = 1.83$ with a t-statistic of 28.0 and 15,245 degrees of freedom.

*Regional industries.* For regional markets, any profit differentials resulting from productivity differentials apply solely to pre-marginal production. Productivity differentials reflect location advantages for existing firms. New firms migrating to the region share in the location advantage but their migration reduces profits of existing firms. Changes in market shares result from existing or new firms expanding their individual markets.

Two avenues for market expansion are hypothesized. First, as the cost of production decreases, firms become more competitive in the export market and with imports. Second, markets are assumed to expand as a region's economy grows. Growth facilitates expansion of specialized parts of regional sectors precluded from local production until the market reaches a certain size. The hypotheses are formulated as

$$R_{it} = R_{iT}\, SVA_t^{\mu_1}\, SVA_{t-1}^{\mu_2}\, (SPA_{it}/SPA_{iT})^{\mu_3}, \tag{56}$$

$$S_{it}^{ku} = S_{iT}^{ku}\, (SPA_{it}/SPA_{iT})^{\xi}, \tag{57}$$

and

$$SVA_t = SVA_{t-1}/SVA_T,$$

$$SPA_{it} = SPA_{i(t-1)} + \lambda_{SP}\,(sp_{it} - SPA_{i(t-1)}),$$

where SVA represents the region's share of U.S. value added, $SPA_i$ is a moving geometric average of $sp_i$ (see equation (33)) and $\lambda_{SP}$ is the coefficient of adjustment.

Equations (56) and (57) are substituted into equation (51) and estimated with nonlinear least squares. Because production cost changes are assumed to affect both market shares proportionately $\mu_3$ was restricted to equal $\xi$. The results (and corresponding t-statistics) are $\mu_1 = 0.296$ (115.4), $\mu_2 = -0.070$ (−3.4), and $\mu_3 = -0.563$ (−18.3) with 24,466 degrees of freedom. The most significant estimate for $\mu_3$ was obtained when $\lambda_{SP} = 0.2$. The results confirm expectations that market shares increase both when a region grows and when industries become more cost competitive. A two-period lag for SVA also was added to the equation but was found to be insignificant and did not influence the other results.

## 3. Policy Simulation

Policy variables are not shown in the equations in section 2 but are built into the model to allow the user to exogenously perturb the desired variables in the model. The explicit, detailed structure of the model readily lends itself to a wide range of policy scenarios. For an example, see Treyz (1981). The REMI model allows policy variables to be substituted directly into the equations. Reduced form econometric models, on the other hand, preclude the use of policy variables if the variables do not change over the historical period,

and would lack statistical validity if the variables changed infrequently (Marschak 1953). Table 1 presents the key policy variables of the REMI model and where they enter it.[9]

To examine the dynamic simulation properties of the model, two sample simulations were run. One involved a demand shock to the economy and the other, a supply shock. To calculate the effects of the exogenous shocks to the economy, an unperturbed forecast, or control forecast, was run first. Then alternative forecasts corresponding to each exogenous shock were run. Comparison of the control forecast with each simulation showed the effects of the exogenous shocks. The specific model used for the analysis was a four-area model constructed for the South Coast Air Quality Management District, where the areas corresponded to Los Angeles, Orange, Riverside, and San Bernardino counties in California. The last year of history for the model is 1988, and the forecast period goes from 1989 to 2035.

To calculate the demand shock, an amount of exports of fabricated metals that would require five thousand employees in Los Angeles County to produce was added for each year in the forecast period. This represented a 7-percent increase in production of fabricated metals and an increase of 0.1 percent of private nonfarm production in Los Angeles County.

As figure 2 shows, the total residence-adjusted employment effect in Los Angeles County is greatest in the first year. The initial increase in employment tightens the labor market, which increases both nominal and real wages. In turn, increased nominal wages reduce the competitiveness of both national and regional industries, which causes a loss of exports. The reduced competitiveness also increases import substitution, but the positive effect of growth in value added on self-sufficiency for regional industries dominates the negative competitive effect. Migration increases, which mitigates the initial increase in wages, and housing prices and sales prices of regional goods increase, which further deflates nominal wages.

The responses over time are gradual. Wage increases lag behind the tightened labor market, and export losses resulting from increased production costs lag behind wage increases. The labor force gradually increases in size as net in-migration occurs in response to increases in real wages and employment rates.

The endogenous response by the labor force causes the real wage and the employment rate to move downward toward their control forecast levels. Concurrently, nominal wages and export losses stabilize. The increase in economic migration ceases in 1997 with the employment rate above its control forecast level and the real

[9] Vectors of disturbances for specialized sectors and components of final demand can be calculated from the 466-sector national input-output table. For a detailed simulation, an option also is available to conjoin an input-output model with the REMI model (Stevens, Treyz, and Kindahl 1981).

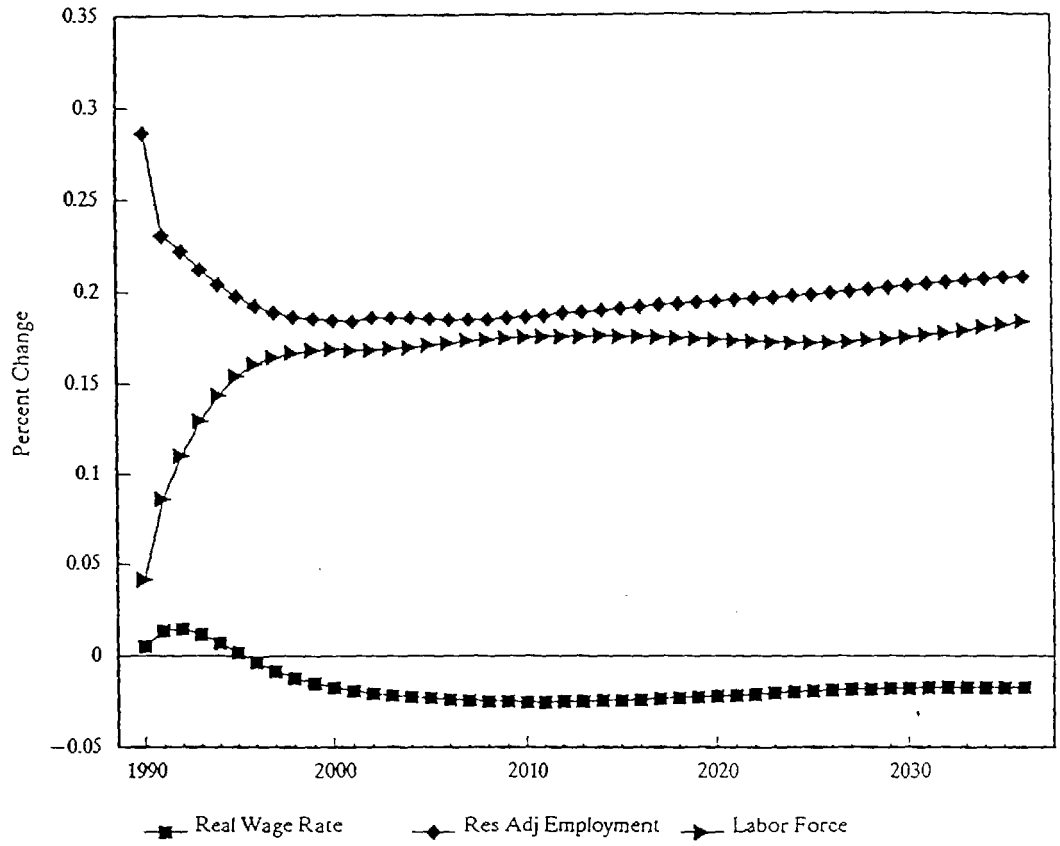
**TABLE 1**
**Policy Variables Used for Simulation**

| Block | Category | Equations | Elements directly affected |
|---|---|---|---|
| 1 | Consumer spending | 4 | Within-state final demand |
| | Investment spending | 11 – 15 | |
| | Government spending | 16 | |
| | Industry demand | 1 | Within-state demand |
| | Industry output | 1 | Industry output |
| | Dividends, interest, & rent | 5, 6 | Disposable income |
| | Residence adjustment | 7 | |
| | Transfer payments | 8, 9 | |
| | Personal taxes & deductions | 10 | |
| 2 | Factor productivity | 21, 33, 50 | Factor demands, profits, prices |
| | Industry employment | 21 | Industry employment |
| 3 | Birth rates | 22 | Births |
| | Survival rates | 23 | Deaths |
| | Labor force participation rates | 24 | Labor force |
| | Retired migrant rates | 25 | Retired migrants |
| | Economic migrants | 30 | Economic migrants |
| | Amenity attractiveness | 30 | |
| 4 | Production costs | 33 | Production costs |
| | Labor costs (other than wage rates) | 34 | |
| | Wage rates | 35 | Wage rates |
| | Occupational supply | 37 | |
| | Business taxes & credits | 38 – 46 | Production costs |
| | Fuel costs | 47 | |
| | Housing prices | 48 | Consumer price deflator |
| | Consumer prices | 49 | |
| 5 | Regional purchase coefficient | 52, 54, 56 | Self supply |
| | Export share | 53, 55, 57 | Exports |

wage below its control forecast level. Because of the Harris-Todaro (1970) specification that both the probability of being employed and the real wage rate influence migration, the effect of the employment rate being above its control level offsets the effect of the real wage rate being below its control level. The real wage rate multiplied by the relative wage mix is above its control level, however, because fabricated metals is a high-wage industry. Because of this and because increased production costs and housing costs are partially transmitted into nominal wages, the nominal wage rate remains above its control level. The ratio of Los Angeles County's total change in output to

**FIGURE 2**

**Effects of Demand Shock**

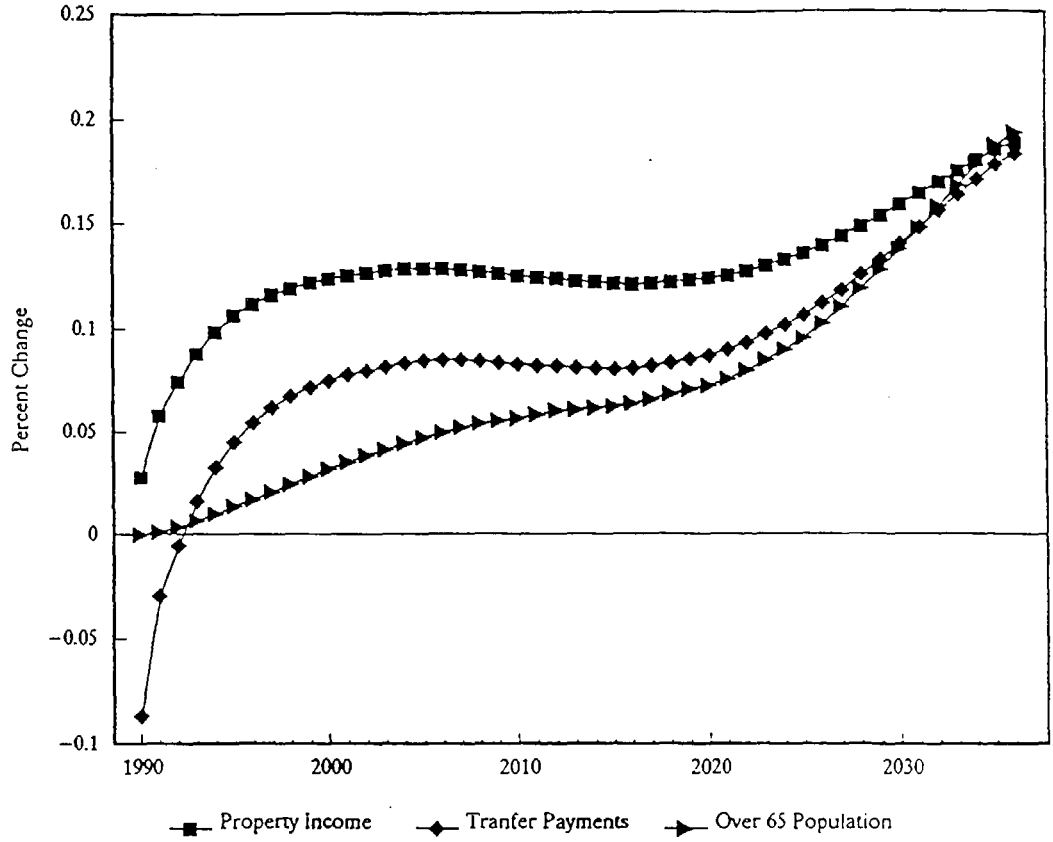



the increase in exports of fabricated metals declined from 2.10 in 1989 to 1.53 in 1997. For the entire four-county region the ratios were 2.26 in 1989 and 1.61 in 1997. The larger ratios for the total region reflect the stimulus to the other counties through the multi-area residence-adjusted income flows and trade flows.

Additional reverberations in the economy continue after 1997 and become most noticeable by 2020. Large numbers of migrants who arrived in the 1990s become retirement age around 2020. As figure 3 shows, the population over age 65 increases, which causes proportionate increases in dividends, interest, rent, and transfer payments due to the significantly greater per capita values of these factors for persons over 65 years of age. The region does not suffer labor force shortages when these migrants retire because many of them brought young children, or had children once they arrived, that will have become work force age by 2020.

To calculate the supply shock, the amenity attractiveness of Los Angeles County was increased by an income equivalent of 1 percent. Though many attributes of the amenity attractiveness are fixed, attributes such as air pollution, traffic congestion, and noise pollution can be directly affected by governmental policies. Willingness-to-pay

**FIGURE 3**
**Effects of Demand Shock**

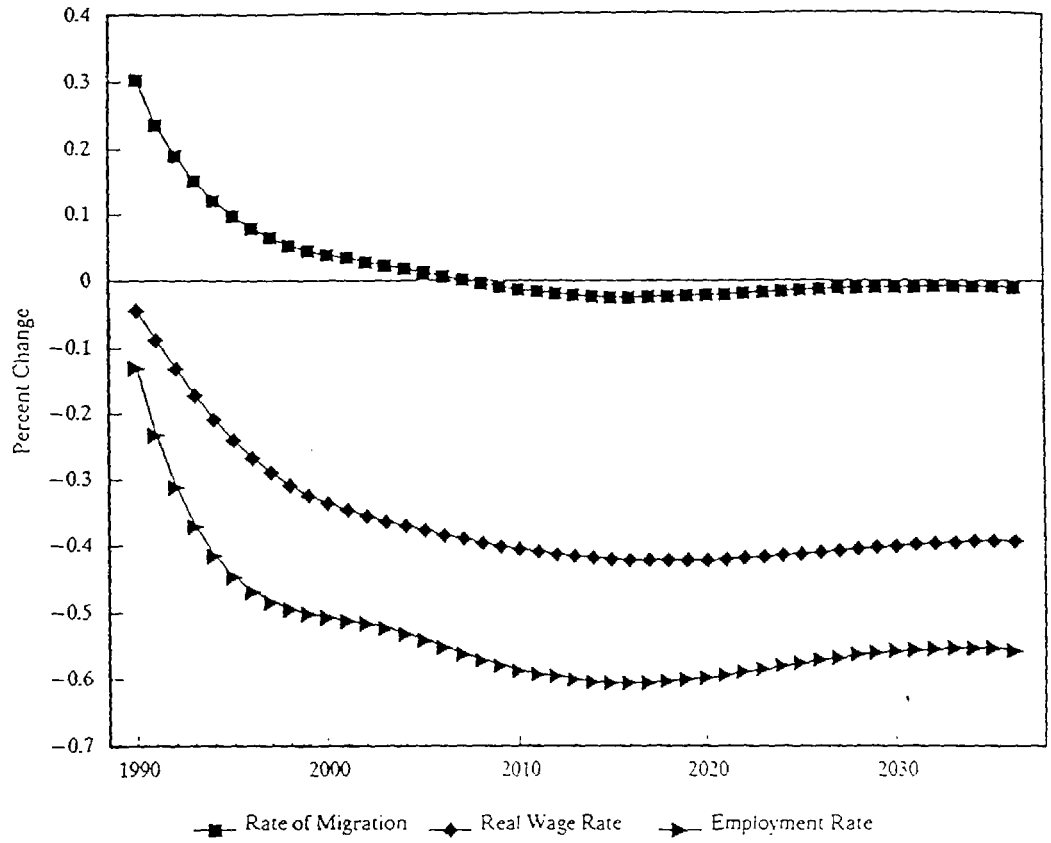



measures for such amenity features are often available from other studies.

A change in amenity attractiveness directly affects the rate of net migration. The income measure can be translated into a change in the amenity factor by using equation (30). By summing the coefficients for relative employment opportunity (REO) and relative wage rate (RWR) and averaging the sums, the direct effect on the rate of migration of a 1-percent increase in income is found to equal 0.357. Consequently, the corresponding amenity attractiveness of the area is changed to reflect a 0.357-percent direct increase in the rate of net migration.

Figure 4 shows that the increase in amenity attractiveness increases the rate of net migration above the control level. The first year increase is less than 0.357 because increased net migration reduces relative employment opportunity and relative wage rate, which mitigates the initial increase in migration. The real wage decreases because nominal wages decrease due to the increased labor supply and because increased land costs and housing prices outweigh reduced labor costs to increase the personal consumption expenditure deflator. Reduced production costs increase the production share of

FIGURE 4
Effects of Supply Shock




both the export and local market, which stimulates employment growth.

By 2007 the migration response is complete and even becomes negative as children of earlier migrants become labor force age. At this point the increase in the attractiveness of the region for migrants due to improvements in the quality of life has been exactly offset by the effects of drops in the probability of being employed and in the real wage rate. Alternatively stated, a 1-percent income-equivalent increase in the amenity attractiveness of the area leads to a 1-percent decrease in the value of the components of income to net migrants. The increased level of amenities become partially capitalized into land and housing prices, which increase 0.34 percent. Employment also increases by 0.45 percent. Employment growth in the total four-county region in 2007 is 5.7 percent greater than the growth in Los Angeles county alone.

## 4. Forecast Evaluation

To evaluate the forecasting potential of the model, post-sample period forecasts for all fifty states plus Washington, D.C., were run.

Each state was treated as a single region for the purpose of the evaluation. To determine the error in the forecast caused by the regional model, actual U.S. historical values were used in the post-sample period. Exogenous regional variables were maintained at their last sample period values. The mean absolute percentage error (MAPE) over all regions for each period is calculated as

$$\text{MAPE} = \frac{\sum_{i=1}^{51} |A_i - P_i|/A_i}{51}$$

where $A_i$ and $P_i$ denote actual and predicted values for region i. Table 2 contains MAPEs for several key variables for post-sample periods 1981–89 and 1985–89. Each year's MAPE represents the cumulative error from the beginning of the forecast.

The MAPEs for first-year forecasts in the two post-sample periods are of similar magnitudes, except for the consumer price deflator. The prediction error of the consumer price deflator in the first-year forecast of the 1985–89 post-sample period arises primarily from errors in forecasting the dramatic housing price increases in the New England states. Overall, the annual forecast error for employment declines with the length of the forecasting period. Therefore, the model appears to be comparatively more successful in long-term forecasting.

The absolute percentage errors are not uniform across states. The first-year forecasts of total employment in the 1981–89 post-sample period range from .01 percent to 5.38 percent with a standard deviation of 1.25 percent. For the ninth year of the forecast, the errors range from .11 percent to 18.3 percent with a standard deviation of 3.86 percent. First-year forecasts of total employment in the second post-sample period, 1985 to 1989, range from .003

TABLE 2
MAPE Statistics

| Selected Variables | 1981 – 1989 | | | 1985 – 1989 | | |
|---|---|---|---|---|---|---|
| | 1981 | 1982 | 1989 | 1985 | 1986 | 1989 |
| Total employment | 1.57 | 2.67 | 5.48 | 1.36 | 2.50 | 4.27 |
| Personal income | 1.68 | 2.84 | 7.21 | 1.20 | 2.48 | 5.30 |
| Real disposable income | 1.51 | 2.49 | 6.32 | 1.54 | 2.81 | 5.12 |
| Consumer price deflator | 0.55 | 0.99 | 6.60 | 1.28 | 2.63 | 5.70 |
| Gross state product (1982 dollars) | 2.00 | 3.52 | 6.71 | 1.51 | 2.61 | 4.83 |
| Wage rate | 0.85 | 1.44 | 6.59 | 0.84 | 1.85 | 3.97 |
| Population | 0.41 | 1.04 | 3.79 | 0.59 | 1.14 | 2.98 |

percent to 2.98 percent with a standard deviation of .22 percent. For the fifth year of the forecast, the errors range from .70 percent to 10.4 percent with a standard deviation of 3.23 percent.

## 5. Conclusion

The REMI EDFS model possesses an explicit structure, detailed in this article, and is a useful tool for both regional forecasting and policy simulation. Much of its current structure reflects implementation of research conducted since earlier versions of the model were published. Sample policy simulations illustrate the dynamic properties of the model, and validation of the model with post-sample forecasts for the 1980s shows its usefulness for regional forecasting.

An advantage of structural models is that they can be updated as new theories emerge, new data sources are provided, and innovations in specifying and estimating econometric relationships develop. Ongoing research and development should continue to improve both the regional forecasting and regional policy simulation capabilities of the model.

## References

Coen, R. M. 1975. Investment behavior, the measurement of depreciation, and tax policy. *American Economic Review* 65: 59–74.

Dervis, K., J. DeMelo, and S. Robinson. 1982. *General equilibrium models for development policy.* Cambridge: Cambridge University Press.

Greenwood, M. J., G. L. Hunt, D. S. Rickman, and G. I. Treyz. 1991. Migration, regional equilibrium, and the estimation of compensating differentials. *American Economic Review* 81: 1382–90.

Griffen, J. M., and P. R. Gregory. 1976. An intercountry translog model of energy substitution responses. *American Economic Review* 66: 845–57.

Hall, R. I., and D. W. Jorgenson. 1967. Tax policy and investment behavior. *American Economic Review* 57: 391–414.

Harrigan, F., and P. McGregor. 1989. Neoclassical and Keynesian perspectives on the regional macro-economy: A computable general equilibrium approach. *Journal of Regional Science* 29: 555–73.

Harris, J. R., and M. P. Todaro. 1970. Migration, unemployment, and development: A two-sector analysis. *American Economic Review* 60: 126–42.

Hulten, C. R., and R. M. Schwab. 1984. Regional productivity growth in U.S. manufacturing: 1951–78. *American Economic Review* 74: 152–61.

Jones, R., and J. Whalley, 1988. Regional effects of taxes in Canada: An applied general equilibrium approach. *Journal of Public Economics* 37: 1–28.

Malhotra, D. M., and G. A. Garofalo. 1988. Analysis of regional productivity with capital as a quasi-fixed factor. *Regional Science and Urban Economics* 18: 533–47.

Marschak, J. 1953. Economic measurements for policy and prediction. In *Studies in econometric methods,* eds. T. C. Koopmans and W. C. Hood. New York: John Wiley and Sons.

Miller, R. E., and P. D. Blair. 1985. Input-output analysis: Foundations and extensions. Englewood Cliffs, N.J.: Prentice-Hall.

Moomaw, R. L. 1981. Productive efficiency and region. *Southern Economic Journal* 48: 344–57.

Morgan, W., J. Mutti, and M. Partridge. 1989. A regional general equilibrium model

of the United States: Tax effects on factor movements and regional production. *Review of Economics and Statistics* 71: 625–35.

Rickman, D. S., and G. I. Treyz. 1992. Industry-level estimates of location response to state differentials in profitability in the United States. Amherst, Mass.: Regional Economic Models, Inc. (REMI), photocopied.

Rickman, D. S., G. Shao, and G. Treyz. Forthcoming. Multiregional stock adjustment equations of residential and nonresidential investment in structures. *Journal of Regional Science.*

Shoven, J. B., and J. Whalley. 1984. Applied general equilibrium models of taxation and international trade: An introduction and survey. *Journal of Economic Literature* 22: 1007–51.

Stevens, B. H., G. I. Treyz, and J. K. Kindahl. 1981. Conjoining an input-output model and a policy analysis model: A case study of the economic effects of expanding a port facility. *Environment and Planning A* 13: 1029–38.

Stevens, B. H., G. I. Treyz, D. J. Ehrlich, and J. R. Bower. 1983. A new technique for the construction of non-survey regional input-output models. *International Regional Science Review* 8: 271–86.

Treyz, G. I. 1980. Design of a multi-regional policy analysis model. *Journal of Regional Science* 20: 191–206.

Treyz, G. I. 1981. Predicting the economic effects of state policy initiatives. *Growth and Change* 12: 2–9.

Treyz, G. I., and B. H. Stevens. 1983. The derivation of the REMI cost of capital equation. Amherst, Mass.: Regional Economic Models, Inc. (REMI), mimeographed.

Treyz, G. I., and B. H. Stevens. 1985. The TFS regional modeling methodology. *Regional Studies* 19: 547–62.

Treyz, G. I., A. F. Friedlaender, and B. H. Stevens. 1980. The employment sector of a regional economic policy simulation model. *Review of Economics and Statistics* 62: 63–73.

Treyz, G. I., B. H. Stevens, and D. J. Ehrlich. 1981. A state core forecasting and policy simulation model. Project 8-15A, National Cooperative Highway Research Program. Washington, D.C.: National Academy of Science.

Treyz, G. I., M. J. Greenwood, G. L. Hunt, and B. H. Stevens. 1988. A multiregional economic-demographic model for regions in the United States. In *Recent advances in regional economic modelling,* eds. F. Harrigan and P. G. McGregor. London: London Papers in Regional Science #19, Pion Limited.

Treyz, G. I., D. S. Rickman, G. L. Hunt, and M. J. Greenwood. Forthcoming. The dynamics of U.S. internal migration. *Review of Economics and Statistics.*

U.S. Bureau of the Census. 1982. *Current Population Survey* matched file by the Bureau of Labor Statistics 1980/81, Data Users Services Division. Washington, D.C.

U.S. Bureau of the Census. 1988. *Current Population Survey* matched file by the Bureau of Labor Statistics 1986/87, Data Users Services Division, Washington, D.C.

U.S. Department of Commerce. 1982. *1977 Census of Transportation.* Washington, D.C.

# Consumption Equations for a Multi-Regional Forecasting and Policy Analysis Model

George Treyz

and

Lisa Petraglia[1]

November, 1996

ABSTRACT: *The lack of adequate regional data limits the development of an econometrically-based macroeconomic specification of regional consumption expenditures. This paper aims to improve upon non-econometric specifications by examining panel data from the Consumer Expenditure Survey (CES) for 13 broad commodities and adhering to Stone's expenditure function. Commodity-specific income and price elasticity estimates are grouped based on a priori restrictions to allow for final estimation of two possible income elasticities and a single-price elasticity using a minimum Chi-squared (MCS) estimator. The final specification satisfies desirable properties within the consumer theory literature as well as key regional forecasting and simulation requirements.*

KEYWORDS: *Consumption expenditures, regional model, consumption equations, income elasticity, price elasticity*

## 1. Introduction

---

[1] George Treyz & Lisa Petraglia, Regional Economic Models, Inc. 306 Lincoln Avenue, Amherst, MA 01002, USA. The authors wish to acknowledge contributions by Chris Christopher, Dale Ballou and Frederick Treyz to this paper. The authors of course assume full responsibility for any errors or omissions. An earlier version of this paper was presented at the Regional Science Association International, Arlington, VA, November 1996.

difficult for two reasons. First, regional macro data is very limited and of poor quality, and second, the resulting set of equations has to fulfill a number of requirements if it is to be used in a regional model.

In this paper we make maximum use of the Consumer Expenditure Survey time series panel of regional data to estimate individual consumption equations in a form suitable for incorporation into a regional macroeconomic model. We then use a minimum Chi-squared technique that utilizes these single equation estimates and our *a priori* restrictions to estimate the parameter set for our final set of equations.

In the following sections, we review past consumption literature, establish the desired properties that are necessary for a set of consumption equations within the context of a multi-regional model, define the functional form of the equations to be estimated, discuss the data, present the estimated equations, and implement a method to generate the final equation parameters that meet our restrictions. Subsequently, we show that the set of equations conform to our requirements and improve on the methodologies that are commonly used for regional structural modeling. Finally we demonstrate how these equations perform in a regional policy analysis model before concluding.

## 2. Literature Review

Consumption equations in the simplest regional input-output models allocate total household income to consumption commodity expenditures using the consumption vector in the national input-output table. Tiebout (1969) developed a methodology for splitting income between old and new residents and using a different consumption vector for each. Blackwell (1978)

expanded these categories to include a redistributive category for those who were previously unemployed. Another approach is to designate different consumption vectors by income class (Miyazawa, 1976). However, the link between industry incomes and household income classes is empirically difficult to model. Subsequent development extends the use of different consumption vectors to incorporate demographic information (Oosterhaven and Dewhurst, 1990).

A consumption sector equation set that captures shifting national consumption patterns in a forecast period and local differences in consumption preferences is specified in Treyz et al.. (1992) as follows:

$$C_j^i = K_j^i \; C_j^u \; (RYD^i/RYD^u), \tag{1}$$

where

$K_j^i$ is a location-specific differential consumption measure for commodity j in area i derived from the Bureau of Labor Statistics consumer expenditure survey (CES) in a recent year. $RYD^i$ and $RYD^u$ represents real disposable income in the region and U.S., respectively, and $C_j^u$ is consumption of good j in the United States.

While these above approaches are able to account for some regional differences, they assume linear homogeneous changes in response to income and lack an econometric foundation. Furthermore, these methods do not reflect price responsiveness among the components of total consumption.

A direct econometric approach is difficult due to the lack of adequate data. Conway (1990) circumvented this problem by using national data directly. West (1994) estimated consumption

equations for Queensland based on 11 years of data for 8 commodities. First, he estimated an equation for total consumption using current and lagged-disposable income, and a lagged-dependent variable with a constraint insuring that the long-run income elasticity was unity. He also estimated similar equations for each commodity using total consumption in place of disposable income. West used Zellner's (1962) seemingly unrelated regression (SUR) technique and reports results that he characterizes as fairly reasonable. In his system, there is no response to prices, and nominal consumption shares vary in the short run but remain fixed in the long run.

A Cobb-Douglas form is used by Berick et al. (1996) in a California CGE model. Here, nominal shares are constant and real shares by commodity shift as individual prices change. A CES structure might be used but presents particular problems in estimation and in meeting desirable modeling criteria. The almost ideal demand system (AIDS) specification, (Deaton and Muellbauer, 1980) is a feasible alternative but requires relative price and income elasticities to be linked and proved unworkable with our data set.

An alternative to macro-specification is the use of micro-economic household survey information. While this approach has a certain appeal, our major interest is in the response of a regional set of households that are all exposed to similar regional economic change. This response may differ from an aggregation of estimated individual household responses to a change in individual household income. See Blundell et al. (1993) for a discussion of the issues involved.

In this study we have chosen to use a macro-specification directly. We have also expanded the set of desirable properties to insure an equation set that conforms to accepted economic

theory and produces accurate forecast and policy simulation results.

## 3. Desirable Properties for Consumption Equations in a Multi-Regional Forecasting and Policy Analysis Model

When choosing the specification of any set of consumption equations it is desirable to have the following standard properties:

| | | |
|---|---|---|
| Criterion 1 | Adding up. | The conventional expression of the adding-up condition is that expenditures by commodity group add up to total income. In our framework, we allow for a residual category, savings. The share of real disposable income saved is assumed to be invariant with respect to prices and income. Thus, our modified adding-up condition requires that expenditure by commodity group add up to a rescaled income variable, where the scalar represents the average propensity to consume. |
| Criterion 2 | Homogeneity. | Increasing all prices and incomes by the same proportion leaves consumption of each commodity unchanged. |
| Criterion 3 | Symmetry. | Cross-price elasticities of income-compensated demand functions are equal, implying consistent choices. |
| Criterion 4 | Negativity. | An increase in the price level reduces income-compensated demand. |

See Deaton and Muellbauer (1980), pp. 43-45 for a formal statement of these properties. In addition to these standard properties, five other properties are important for multi-regional

modeling as follows:

Criterion 5 Selected explanatory variable set. Since regional consumption data has limitations, it is desirable to use a methodology that can predict consumption expenditures by incorporating available regional variables such as income, prices, and age distribution.

Criterion 6 Regional aggregation consistency. This insures that the sum of regional consumption in all the areas of a nation in the baseline forecast will equal the national forecast for that commodity.

Criterion 7 Separate effects for income per capita and population change. Following earlier work mentioned in the literature review, a regional model simulation should show different consumption effects for changes in income due to migration than for changes in income per capita.

Criterion 8 Age composition effects. Any age cohort-specific effects on consumption of particular commodities should be reflected in simulations with an economic/demographic regional model.

Criterion 9 Maximum use of regional data available. Use of all regional data available to estimate response parameters and to estimate unexplained consumption differences (individual effects) by region and by commodity.

The above criteria will be used to evaluate the final equations that are developed here.

## 4. Functional Form

The functional form that we choose for our consumption equation follows that proposed by Stone (1954).

$$\frac{C_{j,t}^{k}}{N_t^k} = e^{\alpha_j^k}\left(\frac{RYD_t^k}{N_t^k}\right)^{\beta_j}\left(\frac{P_{j,t}^k}{\bar{P}_t^k}\right)^{\gamma_j} e^{\psi_j Z_t^k + \theta_{j,t} + \epsilon_{j,k,t}^k} \tag{2}$$

where

$$\bar{P}_t^k = \frac{\prod_{\ell=1}^{13} P_{\ell,t}^{w_{\ell,t-1}}}{\prod_{\ell=1}^{13} P_{\ell,t-1}^{w_{\ell,t-1}}} \bullet \bar{P}_{t-1}^k \qquad t > 1 \tag{3}$$

and

$$\sum_{\ell=1}^{13} w_{\ell,t-1} = 1$$

$C^k_{j,t}$= consumption of commodity j in area k in time period t

$N^k_t$= the population in area k in time period t

$RYD^k_t$= real disposable income in area k in time period t

$P^k_{j,t}$= the price of commodity j in area k in time period t

$w_{\ell,t}$= the proportion of commodity $\ell$ in time t in total nominal U.S. consumption

$Z^k_t$ = a demographic variable specified by Fair (1991) and as follows:

$$Z^k_t = \sum_{i=1}^{18} S_i \cdot j_i$$

where

$S_i$ = the share of the population represented by the five year cohort j

$j_i$ = the number of each cohort starting with 1 for the first cohort

We also experimented with a second order $Z^k_t$ term but found that its inclusion was not supported by the data.

## 5. Data

There are thirteen major components of consumption for which time series are available in the national income and product accounts (NIPA). No data exactly match these series for regions of the U.S. The data that most closely matches the U.S. NIPA time series come from the Consumer Expenditure Survey (CES) and are available from 1987-1993 for 26 metro regions in the U.S. This metro data is grouped in two-year overlapping groups by CES due to the inadequate sample size for single-year cohorts. We used the four groups ('86-'87, '88-'89, '90-'91, '92-'93) that did not include an overlap in years. We also used similar data which are available annually for four major U.S. regions from 1984 to 1993. These two data sets were used to apportion U.S. Consumption by commodity to local areas for 11 out of the 13

consumption series.

The CES data for health services account for only out-of-pocket medical expenses and are therefore inadequate for our purposes. However, the medical industry, for which time series data is available for all states from 1970-1993, supplies 93% of the health services consumption category (the other 7% comes from the insurance industry). Since the medical industry also delivers most of its output to private final consumption, it is possible to calculate a series that serves as a proxy for personal health care consumption by subtracting intermediate industry and other final demand from industry output.

The personal health care consumption series is constructed by subtracting all intermediate and other final-demand uses for the medical industry's output from its total output, which gives a proxy for the amount of medical output that is supplied to consumers. This proxy includes changes in the exports and imports of these services. However if we assume that the exports and imports of these services remain a fixed proportion over time, we can then assume that the rate-of-change in the industry within a state is a good indication of the rate-of-change of local use of the output of that industry. The assumed fixed share that is imported or exported will be captured as a fixed or random effect in our panel estimation if this proportion is not correlated with one of the explanatory variables. The proxy data across all states is normalized using NIPA health care expenditures, to insure that it adds up to the appropriate national consumption series. It is then used as a measure of state-level health care consumption.

A housing services consumption proxy for each state was calculated using value added estimates by state for the real estate industry, and the same procedure that was used for health care. A large proportion of housing services is imputed rent.

The price series used for this analysis comes from a simultaneous solution for industry prices, relative to the U.S., assuming that regional industries (those that typically sell more than 50% of their output in state) base their prices on local factor and intermediate input costs (Treyz et al. 1992). A bridge matrix which indicates the proportion of each industry in the composition of the consumption commodity is applied to the relative industry prices to generate state specific relative consumption commodity prices. These prices, once multiplied by the corresponding U.S. commodity price, are used as commodity prices in the model.

The aggregate consumption price index is a chain weighted index as shown above. The U.S. nominal weights of the previous year are used in each case to establish the change in the price index from year to year.

Real disposable income (RYD) is calculated by dividing disposable income by the price index. Disposable income and total population (N) is reported by the U.S. Department of Commerce, Bureau of Economic Analysis (BEA) (1994, 1995) for each area. The population by five-year cohorts is from the U.S. Department of Commerce, Bureau of the Census (1992) and extended through 1994 using the REMI demographic model calibrated to BEA totals.

## 6. Econometric Estimation and Regression Results

The equations were estimated using the time series panel data described above. Each commodity was specified in a logarithmic form of equation (2) above. All commodities except housing and health care had samples comprised of the 25 (we omit Anchorage) selective MSA's and 4 major regions that C.E.S. tracks (nt=140) . Each of these 11 commodities define an unbalanced panel. Since the samples for housing and health care expenditures were constructed

across 51 states for the period 1970 through 1992 (23 years), the sample size for each of these commodities is 1173. These form balanced panels.

Since time is specified linearly, there is no need to specify a period-effect. An O.L.S. estimation with region-specific dummy variables was performed on each commodity. The decision was made to implement the fixed-effects (FE) estimator over the random-effects (RE) estimator on the basis of greater consistency while foregoing some efficiency gains in the estimates. These 13 one-way fixed-effect models were each compared with the results of their respective O.L.S. estimations (without group dummies). F-test statistics indicate the FE model is appropriate for each commodity.

Consideration was given to possible simultaneity bias in the estimates on both $[RYD/N]^k$ and $[P_j/\overline{P}]_i^k$ since predictions of income, prices and consumer expenditures in the regional model structure are fairly simultaneous. Both RHS variables were instrumented using an exogenous variable set drawn from the REMI EDFS modeling system.[2] The one-way FE model was re-estimated with instruments for the income-per-capita and compensated commodity price concepts.

Insert Table 1 here

A Hausman test statistic was calculated between the non-instrumented and instrumented estimations for each commodity. Based on this, statistic, we rejected the use of non-instrumented results. These previous estimates are shown on Table 1.

[2] Available from author

## 6. Restricted Estimation Set For The Consumption Equations

The individual equation estimates above indicate a preponderance of positive income elasticities and of negative own price elasticities. Three commodities have income elasticity estimates of less than unity. From budget studies of spending by households in different income categories, these three commodities show below average increases in budget share as we move from low to higher income families. The income elasticities greater than one fall into two groups, those with elasticities between 1.26 and 2.16 with negative price elasticities and those with income elasticities above 3.00 and with unacceptable positive own price elasticities.

The estimates as given, if put into a model, would not meet many of the criterion that we have set forth above. Therefore, we follow Blundell et al. (1993) and use a Minimum Chi-Square (MCS) technique. We designate two income-elasticity groups. The first is a low-income elasticity composed of those commodities with elasticities below one, and secondly, a high-income elasticity group, including all of those commodities with single-equation estimated income elasticities above one. We also restrict own-price elasticities to a single-elasticity estimate. More formally, the problem can be stated as:

$$\begin{aligned} &\underset{\gamma,\beta_L,\beta_H}{\textit{Minimize}} \quad \sum_{j=1}^{13}\left(\frac{\hat{\gamma}_j-\gamma}{\hat{\sigma}_j}\right)^2+\sum_{i\in L}\left(\frac{\hat{\beta}_i-\beta_L}{\hat{\sigma}_i}\right)^2+\sum_{i\in H}\left(\frac{\hat{\beta}_i-\beta_H}{\hat{\sigma}_i}\right)^2 \\ &\textit{subject to} \quad \beta_L W_L+\beta_H W_H=1 \end{aligned} \tag{4}$$

The resulting price elasticity is $\gamma$ = -.86 and the resulting low and high-income elasticities are $\beta_L$ = .241 and $\beta_H$ = 1.38 respectively.

In an ideal world, one would wish that the restrictions imposed would meet a strict pooling test or be included in the 99% confidence interval of all the single-equation estimates. Yet, this is not the case. Nevertheless, more than 80% of the coefficients would meet this later test. Another way to view the choice presented here is to compare the closeness of the estimates obtained to the single equation results with the default elasticities of zero-price elasticities and unitary income elasticities for all commodities. In this comparison, the restricted estimates that we obtain are closer to the single equation estimated parameters than zero-price elasticities for almost 70 percent of the price elasticities and unitary-income elasticity for more than 90 percent of the income elasticities. With this evidence, it would appear to be a gain to incorporate these new estimates instead of zero and unitary elasticities until such time when more data is available.

The next step is to estimate the individual effect for each metropolitan area and each of the four major regions of the country using our panel data. This is done for the 11 series where appropriate consumption data is available. The individual effects for metro areas are attributed to each county in that area and the non-metro area counties are assigned their corresponding major region value.

Finally, to complete the specification of the equation to be used in the model, we divide by a similar national consumption equation. The intercept of this equation is chosen in such a way that the consumption predictions across all states add up to the total consumption for the U.S. by commodity. The final equation is as follows:

$$C_{jt}^{k} = \left(\frac{e^{\alpha_j^k}}{e^{\alpha_j^u}}\right)\left(\frac{\frac{RYD_t^k}{N_t^k}}{\frac{RYD_t^u}{N_t^u}}\right)^{\beta_j}\left(\frac{\frac{P_{jt}^k}{\bar{P}_t^k}}{\frac{P_{jt}^u}{\bar{P}_t^u}}\right)^{\gamma}\left(\frac{e^{\psi_j \cdot Z_t^k}}{e^{\psi_j \cdot Z_t^u}}\right)\left(\frac{C_{jt}^u}{N_t^u}\right) \cdot N_t^k \qquad (5)$$

where J = L, H

In this equation the values with a superscript u stand for U.S. values from a baseline forecast. They act as fixed parameters within the framework of this model and, therefore, they only effect the calibration of the baseline forecast. Thus, equation 2 is preserved by setting $\Theta_{jt}$ equal to the extracted value of all the variables with u superscripts. In the equation represented in (5), the regional aggregation of all areas(k) very closely approximates the U.S. consumption of that commodity.

## 8. Meeting the Modeling Criteria

| | | |
|---|---|---|
| Criterion 1 | Adding up. | The adding up properties with our consumption equation, in combination with the restriction imposed on equation (4) which is included in the model structure, are met for price changes and approximately met for income changes. (A one percent change leads to less than .01% discrepancy). |
| Criterion 2 | Homogeneity. | The homogeneity property is met by construction. |
| Criterion 3 | Symmetry. | This property would be met if weights were invariant, income elasticity was equal to 1, and price elasticity was |

equal to minus one. We only meet these conditions approximately. However, given that we are modeling aggregate consumption, there is no reason to expect that it should have this property in any case.

Criterion 4 Negativity. Our equation set and parameter estimates meet the negativity criterion.

Criterion 5 Endogenous effects of income, prices, population and age structure. This condition is also met. All of the key endogenous variables that influence consumption are represented in the consumption equation so that simulation results with the model will not be under-or-over stated due to misspecification of the model.

Criterion 6 Regional aggregation. Both theoretically and empirically, this equation structure yields forecasts by state that very closely approximate total consumption in the U.S. In the absence of area-specific individual effects, as represented by the area-specific estimates, shifts of people or income among areas would only affect U.S. consumption to the extent that it leads to a change in the inequality of income distribution. The inclusion of individual effects means that the model will also capture shifts in consumption by commodity due to different spending patterns in different

parts of the country (e.g. more fuel oil in New England).

Criterion 7 Different effects for income per capita and population changes. This condition is met by our equation. For example, an increase in income per capita will lead to a higher proportionate increase in automobile than medical consumption while an increase in the size of the population due to migration will lead to equal proportionate increases for all income categories.

Criterion 8 Age composition effect. Age composition is an important part of the medical expenditure equation.

Criterion 9 Use of regional data. Is met both for its use in estimating the equations and in its use for estimating individual effects for all areas.

## 9. Demonstration of Use

To demonstrate the performance of these equations in the model, we have introduced a 10% investment tax credit for ten years starting in 1995 as one of the components of a hypothetical fiscal policy change in Pennsylvania. For simplicity, we do not include any tax increase or government spending reduction that would offset the revenue loss from the tax credit.

In figure 1, we show the population and income effects of this policy as a proportion of the control forecast. Real disposable income increases immediately due to price reductions in regional industries, in which lower business costs are passed on to consumers. Additionally,

prices in industries benefiting from the tax credit are reduced and are transmitted to the commodities using inputs from these industries.

Population increases as higher real wages and employment demand attracts migrants to the state. Employment gains are caused by improved competitive conditions, reflected in lower prices or higher profits for firms located in Pennsylvania that lead to increases in local and external market shares. The higher self-supply and exports in turn lead to further output increases over time. Higher wages mitigate the competitive improvement, and the in-flow of migrants erodes the per capita income gains.

Figure 1 follows here

Against this background, figure 2 shows the expenditure effects on three consumption commodities. In the initial year, consumption of other non-durables (mainly tobacco products) bearly increases despite the income increase. Consumption of this commodity changes little since it has a low income elasticity and because its price has a relative increase. The investment tax credit affect on relative prices is in proportion to capital intensity. In this case, other non-durables is the least capital intensive, furniture is next, and household operations (mainly electricity and communications) is the most capital intensive. The increase in both furniture and household operations show gains due to their high income elasticity and relative price changes.

Figure 2 follows here

Over the next ten years of the simulation, the increase in these two high income responsive industries is very small. This is the net effect of decreases in the per capita income gain and increased population due to inward migration. The steady increase in expenditure for other non-durables shows that the effect of the population increase dominates the decrease in income per capita on this commodity because it has a low-income elasticity.

## 10. Conclusion

The consumption equation developed, estimated and used here meets six and approximately meets the other three out of the nine criteria necessary for a structural model that is based ultimately on assumptions of utility maximizing behavior. This equation increases the realism and accuracy of policy analysis uses for a regional or multi-regional model. Its empirical estimation is based on the regional data set currently available and will be enhanced as additional years of data are collected and released.

## References


Berck, P., E. Golan, and B. Smith, with J. Barnhart and A. Dabalen (1996), *Dynamic Revenue Analysis for California*, Mimo, California Department of Finance, Sacramento, California.

Blackwell, J. (1978) Disaggregation of the household sector in regional input-output analysis: some models specifying previous residence of workers, *Regional Studies*, 12, pp. 367-377.

Blundell, R., P. Pashardes, and G. Weber (1993), What Do We Learn About Consumer Demand Patterns from Micro Data?, *American Economic Review*, Vol. 83, No. 3, pp. 570-597.

Conway, Jr, R.S. (1990) The Washington projection and simulation model: a regional interindustry econometric model, *International Regional Science Review*, 13, pp. 141-165.

Deaton, A. and Muellbauer, J. (1980) An almost ideal demand system, *American Economic Review*, 70, pp. 312-326.

Fair, R. C. and K. Dominguez (1991) Effects of the Changing U.S.Age Distribution on Macroeconomic Equations, *American Economic Review*, Vol 81, No.5, pp. 1276-1294.

Miyazawa, K. (1976) *Input-Output Analysis and the Structure of Income Distribution* (Berlin, Springer).

Oosterhaven, J. and Dewhurst, J. H. L. (1990) A prototype demo-economic model with an application to Queensland, *International Regional Science Review*, 13, pp. 51-64.

Stone, R. (1954) Linear expenditure systems and demand analysis: an application to the pattern of British demand, *Economic Journal*, 64, pp. 511-527.

Tiebout, C.M. (1969) An empirical region input-output projection model: the state of Washington, 1980, *Review of Economics and Statistics*, 51, pp. 334-340.

Treyz, G. I. (1993), *Regional Economic Modeling: A Systematic Approach to Economic Forecasting and Policy Analysis* (Boston, Kluwer).

Treyz, G. I., D. Rickman, and G. Shao (1992), The REMI Economic Demographic Forecasting and Simulation Model, *International Regional Science Review*, Vol. 14, No. 3, pp. 221-253, 1992.

U.S. Department of Commerce, Bureau of the Census, *Population Division,* County Estimates by Age, Sex, Race, and Hispanic Origin (1992).

U.S. Department of Commerce, Bureau of Economic Analysis, *Regional Economic Information Systems,* (1994, 1995).

U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Data,* Internet, (Jan. 1996).

U. S. Department of Labor, Bureau of Labor Statistics, *Consumer Expenditure Survey*, (1995).

West, G. R. (1994) The Queensland impact and projection model: the household sector, *Economic Systems Research*, 6, pp. 363-383.

**Table 1.** Fixed Effects Instrumental Variable Estimates

| | $\beta$ (income) | Standard Error | $\gamma$ (price) | Standard Error | Z (age) | Standard Error | adj $R^2$ |
|---|---|---|---|---|---|---|---|
| Non-Durables | -.07 | 1.53 | -2.7 | 2.9 | | | .517 |
| Health Care | -.12 | .16 | -1.3 | 0.1 | .254 | .028 | .995 |
| Housing | .92 | .23 | -0.8 | 0.1 | | | .988 |
| Food | 1.26 | .35 | -1.0 | 0.6 | | | .931 |
| Transp. | 1.36 | .44 | -2.0 | 0.9 | | | .901 |
| Household Operation | 1.44 | .35 | -0.8 | 0.4 | | | .934 |
| Durables | 1.53 | 1.01 | -5.2 | 2.9 | | | .578 |
| Services | 1.62 | .44 | -1.8 | 0.7 | | | .913 |
| Gasoline | 1.85 | .56 | -0.4 | 0.1 | | | .919 |
| Furniture | 2.16 | .69 | -1.0 | 1.3 | | | .865 |
| Clothing | 3.02 | .65 | 2.3 | 1.5 | | | .856 |
| Fuel | 4.18 | 1.46 | 0.4 | 0.4 | | | .943 |
| Autos | 4.65 | .92 | 1.6 | 1.1 | | | .711 |

Figure 1: Pennsylvania Investment Tax Credit Effects on Population and Income

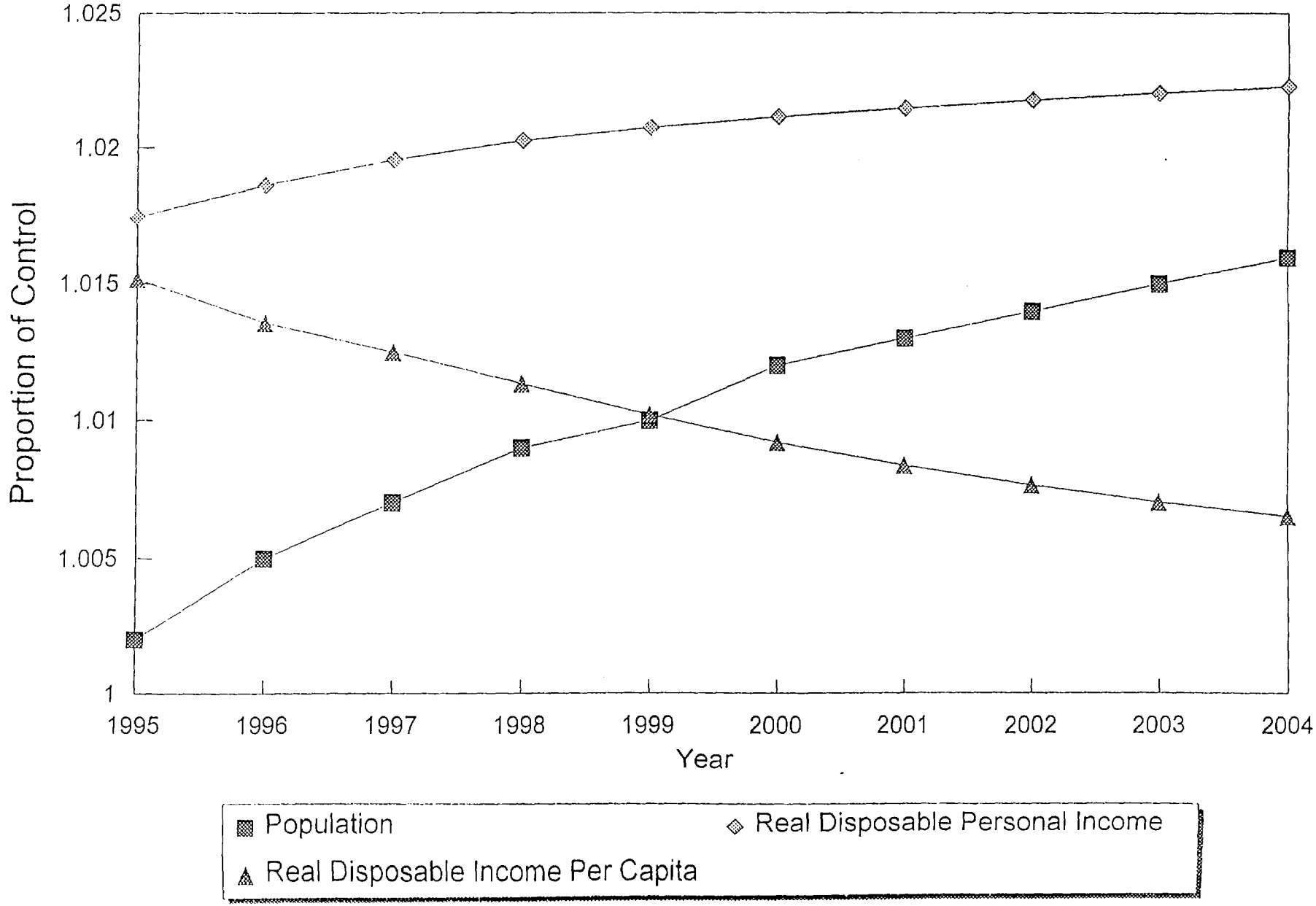
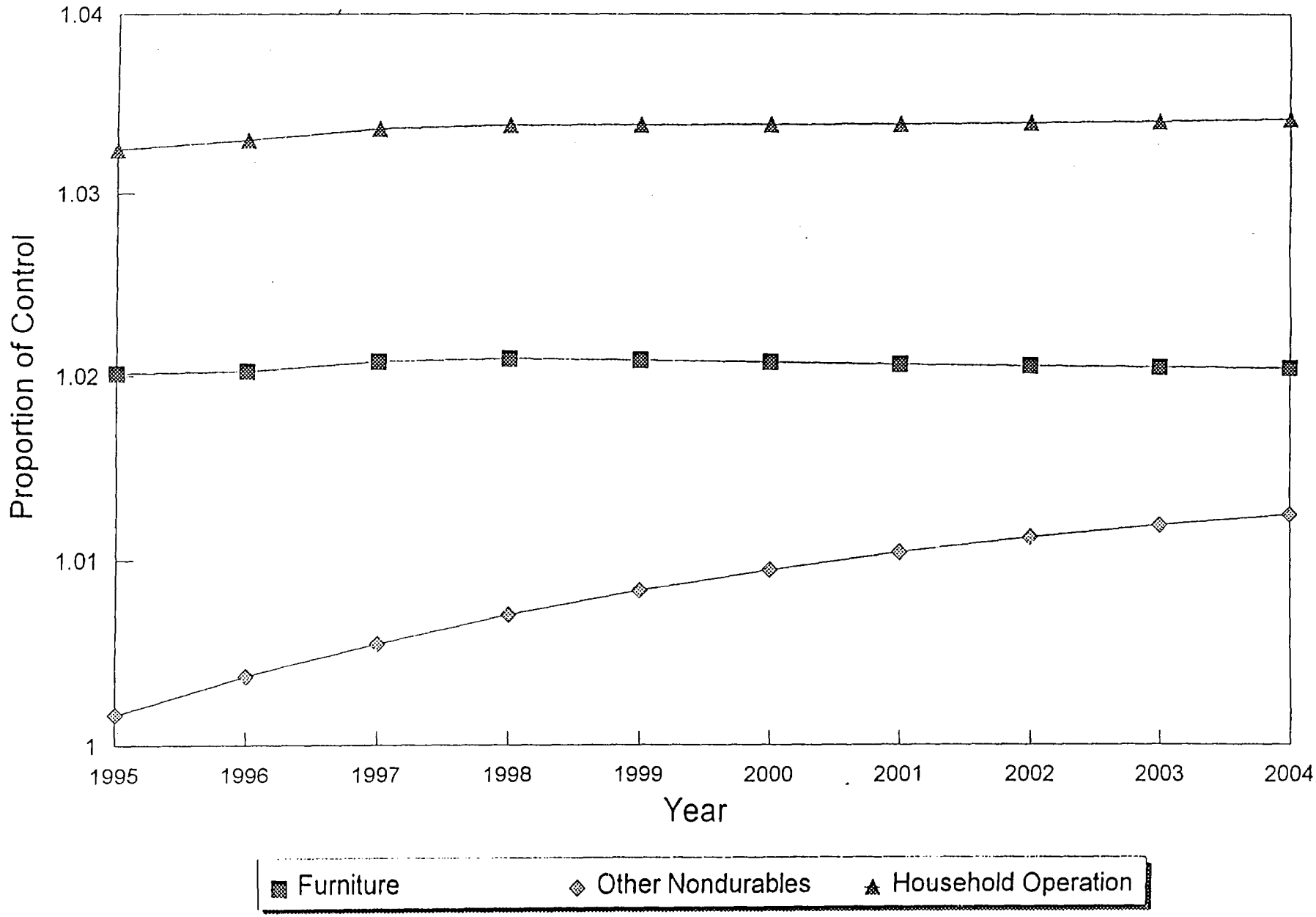


Figure 2: Pennsylvania Investment Tax Credit Effects on Furniture, Household Operations, and Other Nondurable Consumption



# Regional Labor Force Participation Rates

George I Treyz
(President - REMI)

Chris G. Christopher, Jr.
(Economist - REMI)

and

Chengfeng Lou
(Research Assistant - REMI)

March 3, 1996

**Abstract**

We estimate the determinants of regional labor force participation rates with panel and cross-sectional data sets. The panel analysis is gender-specific and offers estimates of regional unexplained time invariant effects and time trends. The cross-section analysis is gender-, age- and race- specific. In order to correct for the well known weakness in cross-section regression we introduce the regional unexplained time invariant effects as explanatory variables in the cross-sectional analysis. We find that there are differences between the genders, age cohorts and races and that the regional unexplained time invariant effects are significant. This implies that modeling at aggregate levels introduces considerable composition bias.

# 1 Introduction

The short-run and long-run dynamics of a regional economy's response to an exogenous shock depends in part on demographic composition. Different demographic groups within a regional economy respond differently to the same shock. Labor market behavior varies by race, gender and age. We find that the answer to the question *Who Benefits from Local Job Growth?* (Bartik, 1992) depends to a significant extent on the local demographic composition.

The purpose of this study is to estimate regional dynamic labor force participation rate equations disaggregated by demographic groups. In national and regional modeling, demographic composition is often overlooked in labor force construction and prediction. Blanchard (1992) and Bartik (1992) investigate the determinants of regional aggregate participation rates without considering racial, gender and age composition. Demographic aggregation is not an issue if the various groups' participation behavior is homogeneous or the demographic composition is static. Since participation rate levels and behavior differ by demographic groups, a significant amount of composition bias in the estimates is introduced in aggregate participation rate estimation. Empirical studies find significant differences in participation rate levels and marginal responses to economic and demographic variables. Bartik (1992) offers a brief summary of the empirical research in this area.

The estimation of regional labor force participation rate equations is problematic since the modeler must choose between panel or cross-section data sets. The Current Population Survey (CPS) contains annual state and MSA

participation rates by race and gender. The 1990 Census contains MSA and state participation rates by race, gender and age. Panel estimation using CPS data allows for the estimation of long-run and short-run dynamics, regional unexplained time invariant effects and time trends. However, the lack of age disaggregation for the race and gender groups introduces composition bias in the estimates. Cross-section estimation for the 1990 Census allows the specification of participation rate equations by race, gender and age.

We specify and estimate dynamic econometric models for the aggregate panels and disaggregate cross-sections. We introduce the panel's regional unexplained time invariant estimates as explanatory variables in the cross-section regressions. Wherever appropriate, the panel and cross-section variables are adjusted in order to diminish compositional bias. The panel estimations provide time trend information and a point of comparison with the cross-sectional results.

In the following section we discuss the theoretical background of participation rate behavior. Section 3 describes the regional and temporal changes in population and participation rates. Section 4 and Section 5 present the econometric issues, variables and results for the panel and cross-sectional analysis respectively. Finally, Section 6 offers a conclusion and suggestions for further research.

# 2 Theoretical Background

We divide the determinants of participation rates into the following categories: regional labor market conditions, competing non-market activities,

demographics, education and dynamics. This categorization is not necessarily disjoint. [1]

The regional labor market conditions include factors that contribute to the definition of the local economic structure, such as job search cost, employment opportunity, real wages and household income. Job search cost is reduced by increases in employment opportunity. (Cogan, 1981). This should increase participation rates. A persistent reduction in employment opportunity will increase the discouraged worker effect.

Wages can have a negative or positive impact on participation rates. For a household, the income effect of an increase in the wage of the highest wage earner may induce the lower wage earner to exit the labor force. On the other hand, a higher real wage can lead to a household member entering the labor force in order to substitute away from non-market activities towards employment. [2]

Competing non-market activities include having children, raising children and attending school. When parents exit the labor force to have or raise children the participation rate of the region decreases. Attending school competes with labor force participation for certain age groups.

The major components of human capital formation are years of schooling and work experience. The level of human capital formation is positively related to the level of education and work experience. A higher level of human

[1] We depart from the theoretical microeconomics categorization since we observe participation rates at the regional level and not as a binary choice of individuals.

[2] The term *non-market activity* can be household production or leisure, see Becker (1991).

capital formation increases the actual and expected return to employment and, therefore, participation rates. (Grounau, 1986, Cogan, 1981). A caveat is that past employment and higher educational attainment are associated with greater lifetime asset accumulation, which can result in a positive income effect. This may lead to an early retirement from the labor force thereby reducing participation rates.

Certain demographic characteristics vary between regions. Ethnic composition and family structure vary between regions and should be accounted for.

Dynamics are important in modeling short-term and persistent responses to key economic conditions. Participation rates are related to present and past values of employment opportunity or past participation rates (Blanchard, 1992 and Bartik, 1992). This is partly based on the premise that past employment increases the likelihood of future involvement in the labor force (Eckstein and Wolpin, 1989 and Gay and Wascher, 1989).

# 3 Population and Participation Rates

In this section, we present and describe the participation rates by race, gender, age and region. We divide race into three categories: white, black and other. The choice of using the category, other, which does not include whites or blacks, is dictated by the availability of adequate data allowing further disaggregation.

The US racial composition by gender for 1967 and 1993 is displayed in Table 1. The percentage of blacks and others increased for both genders.

The racial composition for two age cohorts, 16 to 17 years of age, and over-65 for 1967 and 1993 is presented in Table 2. The percentage of blacks and others increased for both age cohorts while the percentage of whites declined. The changing racial composition in the population is due to different sources of immigration into the US and differences in birth and survival rates by racial groups. The large differences between the younger cohorts indicate that further dramatic changes in the population distribution will occur in the future. In addition to shifts in racial composition, substantial changes have occurred in the regional distribution of the population between 1967 and 1993.

Male and female participation rates for various age cohorts are plotted for the years 1967 through 1993 in Figure 1. It is apparent that female participation rates have increased, and male rates declined over time. The general decline in male participation rates is more dramatic for the older age cohorts. The female participation rate trends are not uniform and exhibit a larger variation than those of their male counterparts for most of the age cohorts. The changing social structure and attitude of female enrollment in higher education, family life and employment is evident in the observation that until the mid-1980s, the female participation rate of the $20 - 24$ age cohort is higher than the $35 - 39$ age cohort. Younger male and female age cohorts exhibit a significant cyclical component around a trend, indicating a relatively greater labor supply sensitivity to the business cycle.

The participation rates by race and gender are plotted across age cohorts for 1967 and 1993 in Figure 2. The hump shape (inverted U) is not surprising

and indicates the importance of age disaggregation. White males exhibit a shift showing the largest participation rate decreases in the 55 – 70 age groups. Black and other males exhibit a different temporal shift across age cohorts. The younger black male participation rate decline is larger relative to their white counterparts. Barring a couple of exceptions, the participation rates of white males are higher than those of non-white males in 1993. The plot for white female participation rates in 1967 across age cohorts shows (approximately) a bimodal (double hump) shape. Black female, other female and 1993 white female participation rates form a single hump shape. The largest increases in female participation rates are for the 20 to 54 years of age whites. With a few exceptions, black female participation rates are higher than white female rates for 1967. The reverse is observed for 1993. The older age cohorts of black and other females exhibit a decline in participation rates.

Hispanic ethnicity crosses racial lines, Figure 3 illustrates that there are considerable differences in participation rates between white female non-Hispanics and white female Hispanics across age cohorts in 1990.

It is apparent from this brief description that participation rate trends and patterns vary by race, gender and age. The aggregate participation rates for the US by age cohort are a reflection of the white rate since whites constitute over 84 percent of the population. Since regions exhibit different racial compositions, the disaggregation of participation rates by race, gender and age is important in understanding regional labor force behavior.

In order to remove the influence of race and age, an adjustment is applied to the regional CPS male and female participation rates. The adjustment

variable is generated by weighting the US race and age participation rate by the regional population of each demographic group. This value is divided by the US race and age participation rate weighted by the US population of each demographic group. [3]

This adjustment variable is divided into the gender specific CPS participation rate. The 1993 adjusted and unadjusted participation rates are plotted in Figure 4. Note that the states exhibit large variation by either measure. Those states that are below the 45-degree line have an above average representation in cohorts with high US participation rates. For example, Alaska's (AK) female unadjusted rate is 67 percent and adjusted rate is 58 percent. This indicates a high representation in those race and age cohorts that have high participation rates at the national level. On the other hand, Florida's (FL) female unadjusted rate is 54 percent and adjusted rate is 58, reflecting a high representation of age and race cohorts with low participation rates.

Using gender-specific regional adjusted participation rates, we examine the changes for two points in time. The 1977 and 1993 adjusted state male and female participation rates are plotted in Figure 5. The regional female participation rates are above the 45 degree line indicating a universal increase. The difference in the increase varies widely between states. For example, Nebraska (NE) and Iowa (IA) show dramatic increases while California (CA) does not. The adjusted state male participation rates generally declined, with

[3] $A_i = (\sum_{j=1}^{n} P_{u,j} N_{i,j} / \sum_{j}^{n} N_{i,j}) / (\sum_{j=1}^{n} P_{u,j} N_{u,j} / \sum_{j=1}^{n} N_{u,j})$, where $A_i$ is the compositional adjustment for region $i$, $P_{u,j}$ is the participation rate for the $j$ age-race cohort in the US and $N_{i,j}$ is the population in the $j$ race and age cohort in region $i$.

a few exceptions. Both figures display the existence of regional variation and changes. It is evident that the US labor force and participation rates necessitate a bottom-up approach from a modeling perspective.

# 4 Panel Estimation

We estimate gender-specific adjusted labor force participation rate equations in a panel framework. The cross-sectional observations are the 50 states and Washington DC. The time span is 1977 – 1993. A restriction on the slope coefficients across regions is imposed and we allow for regional unexplained time invariant effects and time dummies. The regional unexplained time invariant effects will serve as explanatory variables in the cross-section participation rate equations, which follow in Section 5.

## 4.1 Econometric Model

A formal econometric representation of the gender-specific labor force participation rate equation is

$$PR_{it} = \alpha + \gamma RWR_{it} + \beta EA(\lambda)_{it} + Z_i'\nu + \mu_i + \delta_t + \epsilon_{it} \qquad (1)$$

$$i = 1, 2, \ldots, 51$$

$$t = 1977, 1978, \ldots, 1993$$

$$\mu_i \sim iid(0, \sigma_\mu^2)$$

$$\epsilon_{it} \sim iid(0, \sigma_\epsilon^2)$$

$$EA(\lambda)_{it} = EA(\lambda)_{i,t-1} + \lambda(E_{it} - EA(\lambda)_{i,t-1}) \tag{2}$$

$$\lambda \in (0, 1]$$

The time variant variables $PR_{it}$, $RWR_{it}$ and $EA_{it}$ are in logs. The gender-specific labor force participation rate of region $i$ at time $t$ is denoted by $PR_{it}$, $RWR$ is the regional real wage rate, $E$ is the regional employment divided by the population 16 years of age and older, $Z_i$ is a column vector of regional specific time invariant variables, $\mu_i$ is the regional unexplained time invariant effect[4], $\delta_t$ is a time dummy coefficient and $e_{it}$ is a general disturbance term.[5] The time variant variables and most of the time invariant variables are adjusted where appropriate to eliminate race and age bias. This is an important consideration since the regional unexplained time invariant effects need to be purged as much as possible of any inherent composition bias. The appendix discusses in detail the data sources and construction of each variable.

The effect of $RWR$ on labor force participation is theoretically undetermined. A positive sign indicates that the substitution effect dominates the negative income effect. In a model with a traditional family structure, $RWR$ could result in a positive coefficient for male participation rates and negative for females.

[4]This is generally termed a fixed or random effect.

[5]The modeling of spatial autocorrelation is not warranted in $\epsilon_{it}$ and/or $PR_{it}$ since the regional level is states and there are no apparent theoretical reasons to expect significant spill over effects between most states.

the negative income effect. In a model with a traditional family structure, $RWR$ could result in a positive coefficient for male participation rates and a negative coefficient for a females.

The variable $EA(\lambda)$ captures the dynamic effects of general employment conditions on labor force participation rates. The short-run and long-run effects of $EA(\lambda)$ are determined by the numerical value of the speed of adjustment, $\lambda$. A higher value of $\lambda$ indicates that general employment conditions have a more rapid impact on labor force participation rates. There are four compelling reasons for a positive coefficient on $EA(\lambda)$ : (i) the discouraged worker effect is stronger than the extra worker effect, (ii) an increase in employment opportunity reduces the cost of job search, (iii) human-capital formation with respect to past employment for $\lambda \in (0, 1)$ and (iv) favorable individual and community habit formation with respect to employment.

We considered other variables to capture dynamic effects of labor market conditions on labor force participation rates. We rejected the use of the regional unemployment rates due to measurement error, (Bartik, 1991). If labor force measurement error exists then it would appear in the participation rate and the unemployment rate, leading to bias in the estimates. Bartik (1991) suggests using different data sources for the participation rate and the regressors. We use establishment based BEA employment data in constructing our regressor variable, $EA(\lambda)$. A considerable amount of relevant regional educational and demographic data are not available on an annual basis. We use the 1990 values for the time invariant variables.

The time invariant variables are not identical for each gender, although

there is a considerable overlap. Common to both genders is the proportion of the regional population in rural counties $RURAL$, the proportion in a family with two parents and at least one child $CWC$, the proportion of college degree holders $COL$, the proportion of the population that did not complete high school $NOH$, and $HISP$ the percent Hispanic. The natality rate, $NR$, is an explanatory variable in the female participation rate equation. The variables $CWC$, $COL$, $NOH$, $HISP$ and $NR$ are adjusted to remove composition bias.

We assume *a priori* that the time invariant variables are exogenous and asymptotically uncorrelated with $\mu_i$ and $\epsilon_{it}$. Theoretical reasons exist for the asymptotic correlation between the time varying variables and $\epsilon_{it}$.

### 4.1.1 Econometric Methodology

We follow a two step strategy to estimate equation 1. The first step involves estimating $\lambda$ and the coefficients of the time variant variables. The second step involves estimating the coefficients of the time invariant variables.

The parameters to be estimated in the first step are $\gamma, \beta, \delta$ and $\lambda$ . Demeaning (1) by region removes the time invariant variables and coefficients

$$PR_{it} - PR_{i\cdot} = \gamma(RWR_{it} - RWR_{i\cdot}) + \beta(EA(\lambda)_{it} - EA(\lambda)_{i\cdot}) + \delta_t + (\epsilon_{it} - \epsilon_{i\cdot})$$

where the subscript $i\cdot$ denotes the mean value of a variable with respect to region $i$.

The above equation is estimated by Least Squares (LS) and Two-Stage Least Squares (TSLS). The instrumental variable list includes a collection of

regional variables that are considered pre-determined in a larger simultaneous equation model of Treyz (1993), of which this equation is considered a member. These estimates are generally called with-in group or least squares fixed effects estimates. The optimal $\lambda$ is estimated by a grid search with the TSLS procedure that minimizes the sum of the squared residuals. This is done to insure consistent estimates due to the possible asymptotic correlation between $\epsilon_{it}$ and the time variant variables. Notice that the asymptotic correlation of the time variant variables and $\mu_i$ is not an issue since it is removed in the demeaning process. The grid search is conducted on the following sequence: $\{0.1, 0.2, \ldots, 0.9, 1.0\}$. A Hausman test is performed between the LS and TSLS estimates.

The second step involves generating the following variable:

$$d_i = PR_{i\cdot} - \hat{\gamma} RWR_{i\cdot} - \hat{\beta} EA(\hat{\lambda})_{i\cdot}$$

and performing the between group estimation of the time invariant variables

$$d_i = \alpha + Z_i' \nu + \mu_i$$

$$\mu_i \sim iid(0, \sigma_\mu^2) \ .$$

The above procedure provides estimates that are consistent.[6]

[6] Based on a Hausman test for the time variant regression the Random Effects estimator was rejected. In light of this, we decided not to pursue a GLS Random Effects estimator that can simultaneously estimate the time variant and invariant variable coefficients.

### 4.1.2 Results

The estimation from the time variant regressions are displayed in Table 3 and Figure 6. The results from the time invariant regression are displayed in Table 4 and Figure 7.

The time variant regressions exhibit a high adjusted $R^2$. The LS estimates are the preferred estimates for males and TSLS for females. The time varying variable coefficients are significant for each gender. The sign on $RWR$ is positive for males and negative for females, thus implying that an increase in the real wage increases male and decreases female participation rates. The optimal $\lambda$ values indicate that female participation rates exhibit an immediate response to the general employment conditions while male participation rate respond with delay.

The time dummies are graphed separately by gender and verify the trends mentioned in Section 3. An $F - test$ performed for each gender rejects the hypothesis that the time dummies are all equal. The unexplained drop in the participation rate for males is approximately six percentage points while the female participation rates increase approximately seventeen percentage points, approximately one percentage point a year on average.

An $F - test$ performed for each gender rejects the hypothesis that the $d_i$ are all equal. This implies that there are time invariant differences among regions. The question that the time invariant regression answers is to what extent these are unexplained.

The time invariant estimates are presented in Table 4 and Figure 7. The

adjusted $R^2$ for males is 0.093 and the $F$ is 0.927, indicating that the significance of the regression is low. However, $NOH$ is significant at the 15 percent level of significance. The coefficient is negative indicating that not completing high school reduces male participation in the labor force. This fits the view that with less human capital the return from participating in the labor force is lower. The results of the female time invariant regressions are more promising. The variables that are significant at the 15 percent level are $CWC, NOH$ and $HISP$. The sign of $NOH$ is negative and smaller than the male coefficient. The variable $CWC$ exhibits a negative sign on female participation while it is not significant for males. A similar result is found with respect to $HISP$. The results with respect to $CWC$ and $HISP$ imply that traditional family structure behavior is present in the overall population and in particular among Hispanics.

The male and female regional unexplained time invariant effects $\mu_i$ are displayed in Figure 7. There is approximately 3.2 higher variation between the female $\mu_i$ than between the male, implying a larger unexplained regional variation between female participation rates than between male participation rates. The correlation between the male and female $\mu_i$ is 0.62. The $\mu_i$ estimates correspond to so-called first nature properties, (Krugman, 1993).

# 5 Cross-Section Estimation

Since we have adjusted most of the variables in the gender-specific panel estimation with respect to age and race composition and included the percent Hispanic as an explanatory variable, the demographic composition bias in

the $\mu_i$ coefficients is minimal. The cross-section regressions involve a more detailed level of race-age aggregation, due to availability of data that is not available at the regional level across time. However, a fundamental drawback to cross-sectional analysis is that regional unexplained time invariant effects are not identified. This can produce inconsistent estimates.

## 5.1 Econometric Model

The cross-section analysis involves estimating labor force participation rate equations for 20 age cohorts by gender and race categories. This amounts to a total of 120 regression results presented in Table 5 through 11. Each table displays the results for a particular gender-race group for all age cohorts. The age cohort aggregations are displayed along the columns of each table. The explanatory variables are displayed along the rows.

The age groups are single year cohorts for the population that is 16 through 24 years of age since heterogeneous educational and labor market responses exist. The population that is 25 through 74 years of age is represented by ten disjoint and consecutive five year age cohorts, i.e. $25-29, 30-34, \ldots, 70-74$. The population 75 years of age and older is aggregated into one cohort. The aggregation into five year age cohorts reduces the data and regression work needed without introducing severe age aggregation bias.

Since the cross-sectional observations are the 50 states and DC and the *other* race category is approximately 3.7 percent of the US population, further breakdown of this category would introduce a large amount of mea-

surement error due to the small sample size of these populations in many states. Omitting states that do not have a sufficient sample size for certain race representation would in many cases dramatically reduce the degrees of freedom.

For each race, gender and age group the following econometric model is specified:

$$PR_i = \beta_1 + \beta_2 EA(\lambda)_i + \beta_3 RWR_i + \beta_4 RURAL_i + \beta_5 \mu_i + X_i'\gamma + \epsilon_i \quad (3)$$

$$i = 1, 2, \ldots, 51$$

$$\epsilon_i \sim iid(0, \sigma_\epsilon^2)$$

$$\lambda \in (0, 1]$$

The variables $EA(\lambda)$, $RWR$ and $RURAL$ are discussed in the previous section. The variable $\mu_i$ is gender specific and is the regional unexplained time invariant effect estimated from the panel econometric model. To the best of our knowledge this is the first article that attempts to correct the inherent weakness of cross-section regressions in this manner.

The column vector, $X_i$, consists of variables included in a race-, gender- and age cohort-specific regression. The variables represented are race-specific, gender-specific, or race- and gender- and age cohort-specific. Depending on the gender and age cohort, certain variables are included or excluded as regressors, since specific variables are relevant only to certain gender and age cohort groups.

The Couple with Children variable, CWC, is race-specific and is included as an explanatory variable in the 16 though 60 – 64 age cohort regressions for each gender. The percent of a specific race-gender-age cohort-specific group that did not complete high school is denoted by $NOH$ and is an explanatory variable for the 18 years of age and older cohort regressions. $NOH$ is not included as an explanatory for the 16 and 17 years of age cohorts since NOH includes high school attendees. The variable $NOH1824$ is the race- and gender-specific proportion of the 18 through 24 years of age population that did not complete high school. The value of $NOH1824$ is a drop out rate proxy for the 16 and 17 years of age population. A high value of $NOH1824$ is associated with a high dropout rate. A high dropout rate implies an increase in the labor supply for the 16 and 17 years of age population. The variable $COL$ is the proportion of a race-, gender- and age cohort-specific group that completed college. $COL$ is an explanatory variable for the 25 and older age cohorts and $COL2534$ is used for the 18 through 24 age cohorts.

The variable $HISP$ is the race-specific proportion that is ethnically Hispanic. A few states have large Hispanic populations, i.e. California, Texas, New York and Florida, and this may influence the participation rate levels.

### 5.1.1 Econometric Methodology

The econometric strategy is similar to that of the panel estimation with respect to $\lambda$. While $EA(\lambda)$ and $RWR$ are endogenous to a large system of equations , Treyz(1993), the contribution of a race-, gender- and age cohort-specific participation rate to the explanation of either endogenous variable

is sufficiently small. Therefore, $OLS$ estimation of equation 3 is acceptable and $TSLS$ is unnecessary.

### 5.1.2 Results

For brevity, we use the term *significant* in this section to mean statistically *significant at the* 10 *percent level* (two tailed test) unless stated otherwise. None of the significant $\mu_i$ coefficients are negative. The $\mu_i$ coefficients contributed more in explaining white participation rates than those for the black or other races. There appears to be an approximate U shape if the $\mu_i$ coefficients are plotted across age cohorts for a particular race and gender. This is due to the observation noted in Section 3 that participation rates are higher for the middle aged cohorts than for the younger and older age cohorts. The white male $\mu_i$ coefficients tend to be smaller than their female counterparts except for the 19 to 24 years of age cohorts. The $\mu_i$ variable is not significant for the black male age cohorts. For comparative purposes Table 11 displays the results of the white males with $\mu_i$ omitted as an explanatory variable. It is evident that $\mu_i$ contributes significantly to the explanation of white male participation rates and is correlated with some of the regressors. This is typical of omitted variables that are not orthogonal to the other regressors. The regional unexplained time invariant effects are significant in 19, 5 and 8 of the cases for white, black and other females, respectively.

The explanatory variables that fall under the category of competing non-market activities are $NOH1824$, $COL2534$, $CWC$ and $NR$. The $NOH1824$ variable is significant with a negative sign for the 17 years of age cohort among

other males, black females and other females. The $NOH1824$ coefficients that have a $t$-ratio greater than 1 are negative. The prior on this coefficient is positive since a high dropout rate would lead to an increase in the labor force. However, a sociological factor may be present. The variable $COL2534$ is significant for 18 of the 42 cases. It is significantly negative for white males and females.

$CWC$ is significant in 12 of the male and 20 of the female cases. Within the significant male cases, it is positive for whites and others. The coefficient is negative for younger black males. A positive coefficient for males is explained by the need for income to provide for a household. The white female $CWC$ coefficients are significant in 10 of the 20 age cohorts and have a constant negative sign for cohorts of 18 years of age. For blacks and others the significant variables do not have consistent signs.

The natality rate, $NR$, coefficient is negative in all of the significant cases except for the 17 years of age black females. It is significant in 15 of the 39 cases.

The predominance of the significant and negative $RURAL$ coefficients indicates the presence of non-market productive work and the high job search cost in the rural areas.

The employment opportunity coefficient is significantly positive in 95 of the 120 regressions. These long-run elasticities of the participation rate to employment opportunity indicate the percentage change in participation to a percentage change in the general employment to population ratio. They are shown in Figure 8. These figures show that the largest proportional

responses tend to be in the younger and older age cohorts. The youngest and oldest cohorts responses are above unity. The results indicate that black males exhibit a higher response elasticity than white males. The other males are more responsive than whites in the 20 through 45 years of age cohorts.

Clark and Summers (1981) sought to estimate the mean lag of the past employment effect. However, they found that the $\lambda$ value often fell outside the unit interval and set it to 0.10 on *a priori* grounds in most cases. The speed of adjustment, $\lambda$, is a key parameter estimated by this study. The overall response in the long-run as well as the response after the indicated numbers of years are shown in Figure 9. These responses indicate that for females, where the $\lambda$ values are less than one, a difference exists with the estimate value for the panel. However, the $\lambda$'s for the male cohorts average out to be fairly close to the panel regression value. Since the purpose of the panel is to estimate regional unexplained time invariant effects for the cross-section regressions, we performed a sensitivity test by generating $\mu_i$ values for a range of $\lambda$ values in the panel and found that the cross-section estimates are not sensitive to the $\lambda$ used in the panel. The long-run effects of increased employment on participation rates are indirectly related to concepts estimated in other studies. Bartik (1991) estimates that a one percent increase in employment leads to a permanent increase in participation rates of 0.157 percent using his formulation and data and a 0.109 percent change using Blancahrd and Katz (1992) data. The Blanchard and Katz estimate is 0.003 but when modified by Bartik (1991) to include lags it is 0.254. In these cases, the explanatory variable is employment whereas in our case, it is an employment to population

ratio. Since some inward migration will occur if employment increases, our measure of employment to population will increase less than the employment measure of Bartik (1992) or Blanchard and Katz (1992). Our estimates of the elasticity of response of participation rates to employment opportunity change are apparently somewhat higher on average than Clark and Summers (1982). For females 25 to 55 years of age, our range is 0.21 to 0.47, 0.24 to 0.92, and 0.40 to 0.60 for whites, blacks and others, respectively. Our male counterparts are 0.08 to 0.19, 0.21 to 0.48 and 0.18 to 0.37.

The speed of response for Blanchard and Katz peaked in the third year while Bartik's estimated response remained constant after the first year. Again focusing on the 25 to 54 years of age groups, the percent effect in the first year for whites, blacks and others respectively are from 30 to 40 percent, 10 to 40 percent, 60 to 100 percent for females and 30 to 40 percent, 10 to 70 percent, 40 to 70 percent for males.

# 6 Conclusion

There is significant variation in participation rate behavior by race, gender, age and region. However, almost all age cohort participation rates exhibit a positive response to increased employment opportunity. White males tend to increase participation rates in response to an increase in the real wage, while white women decrease their participation rate. Increased education tends to increase participation rates, except for the ages where education competes with work. Family structure, natality and rural versus urban location influence participation rates. Given the variation in response to changes in the

determinants of labor force participation and the shifting demographic composition of the US and states it would appear that any analysis of the effects of policies would have to be analyzed within the framework of a model that included heterogeneous demographic groupings.

# 7 Data Appendix

In order to save space we present two tables that describe our adjustments and data sources for the variables in the text. The sources are Current Population Survey (CPS), Employment and Earnings (EAE), Vital Statistics (Vital) and Census. Table $A1$ describes the age cohort detail code. Table $A2$ presents the adjustment and variables used in the panel and cross-section. Table $A3$ presents the data sources for the variables in Table $A2$ and the text. Since, the cross-section adjustments are similar to the panel adjustments we present only the panel adjustments. The detail code $M$ is male, $F$ female, $W$ white, $B$ black, $O$ other. In Tables $A2$ and $A3$ we suppress the time subscripts for clarity. The subscript $i$ pertains to region and $u$ indicates the aggregate US.

The instrument list is available from the authors.

## 7.1 Table A1

| Code | Age Cohort Groups |
|---|---|
| $12AC$ | $16-19, 20-24, \ldots, 70+$ |
| $7AC$ | $16-19, 20-24, 25-34, \ldots, 65+$ |
| $9AC$ | $16-17, 18-24, 25-34, \ldots, 70+$ |
| $1FAC$ | $15-49$ |
| $7FAC$ | $15-19, 20-24, \ldots, 45-49$ |
| $20AC$ | $16, 17, \ldots, 24, 25-29, \ldots, 75+$ |

## 7.2 Table A2

| Variable | Formulation | Time Span | Detail |
|---|---|---|---|
| $PR_i$ | $P_i/PRADJ_i$ | 1977 – 1993 | $M/F$ |
| $PRADJ_i$ | $\frac{(\sum_j P_{u,j}N_{i,j}/\sum_j N_{i,j})}{(\sum_j P_{u,j}N_{u,j}/\sum_j N_{u,j})}$ | 1977 – 1993 | $M/F$ |
| $E_i$ | $EN_i/EADJ_i$ | 1967 – 1993 | |
| $EN_i$ | $(1+RA_i/YLP_i)(EB_i/POP16_i)$ | 1967 – 1993 | |
| $EADJ_i$ | $\frac{(\sum_j EN_{u,j}N_{i,j}/\sum_j N_{i,j})}{(\sum_j EN_{u,j}N_{u,j}/\sum_j N_{u,j})}$ | 1967 – 1993 | |
| $RWR_i$ | $WR_i/RWRADJ_i$ | 1977 – 1993 | |
| $WR_i$ | $WS_i(RYD_i/YP_i)$ | 1977 – 1993 | |
| $WS_i$ | $YLP_i/EB_i$ | 1977 – 1993 | |
| $RYD_i$ | $YLD_i/PCE_i$ | 1977 – 1993 | |
| $RWRADJ_i$ | $\frac{(\sum_j W_{u,j}ER_{u,j}N_{i,j}/\sum_j ER_{u,j}N_{i,j})}{(\sum_j W_{u,j}ER_{u,j}N_{u,j}/\sum_j ER_{u,j}N_{u,j})}$ | 1977 – 1993 | |
| $NR_i$ | $NNR_i/NRADJ_i$ | 1990 | $F$ |
| $NRADJ_i$ | $\frac{(\sum_j NNR_{u,j}NRP_{i,j}/\sum_j NRP_{i,j})}{(\sum_j NNR_{u,j}NRP_{u,j}/\sum_j NRP_{u,j})}$ | 1990 | $F$ |
| $NOH_i$ | $NOHH_i/NOHADJ_{i,j}$ | 1990 | $M/F$ |
| $NOHADJ_i$ | $\frac{(\sum_j NNOH_{u,j}POP_{i,j}/\sum_j POP_{i,j})}{(\sum_j NNOH_{u,j}POP_{u,j}/\sum_j POP_{u,j})}$ | 1990 | $M/F$ |
| $COL_i$ | $COLL_i/COLADJ_i$ | 1990 | $M/F$ |
| $COLADJ_i$ | $\frac{(\sum_j COLL_{u,j}POP_{i,j}/\sum_j POP_{i,j})}{(\sum_j COLL_{u,j}POP_{u,j}/\sum_j POP_{u,j})}$ | 1990 | $M/F$ |
| $CWC_i$ | $CCWC_i/CWCADJ_i$ | 1990 | $M/F$ |
| $CWCADJ_i$ | $\frac{(\sum_j CCWC_{u,j}HHLD_{i,j}/\sum_j HHLD_{i,j})}{(\sum_j CCWC_{u,j}HHLD_{u,j}/\sum_j HHLD_{u,j})}$ | 1990 | |

## 7.3 Table A3

| Variable | Description | Source | Time Span | Detail |
|---|---|---|---|---|
| | | Panel | | |
| $P_i$ | participation rate | CPS | 1977-1993 | $M/F$ |
| $P_{u,j}$ | participation rate | EAE | 1977-1993 | $M/F, W/B/O, 1$ |
| $N_{u,j}, N_{i,j}$ | population | Census | 1977-1993 | $M/F, W/B/O, 1$ |
| $RA_i$ | residence adjustment | BEA | 1967-1993 | |
| $YLP_i$ | labor & proprietors income | BEA | 1967-1993 | |
| $EB_i$ | employment | BEA | 1967-1993 | |
| $POP16$ | population 16 and over | Census | 1967-1993 | |
| $EN_{u,j}$ | civilian employment rate | EAE | 1967-1993 | $M/F, W/B/O, 1$ |
| $YP_i$ | income | BEA | 1977-1993 | |
| $W_{u,j}$ | median weekly earnings | CPS | 1977-1993 | $M/F, W/B, 7AC$ |
| $ER_{u,j}$ | employment rate | EAE | 1977-1993 | $M/F, W/B, 7AC$ |
| $PCE_i$ | see Treyz (1993) | | | |
| $YLD_i$ | see Treyz (1993) | | | |
| $NNR_i$ | natality rate | Vital | 1990 | $F, W/B/O, 1FA$ |
| $NNR_{u,j}$ | natality rate | Vital | 1990 | $F, W/B/O, 7FAC$ |
| $NRP_{i,j}, NRP_{u,j}$ | female population | Census | 1990 | $F, W/B/O, 7FAC$ |
| $NOHH_i$ | % no high school diploma | Census | 1990 | $M/F$ |
| $NOHH_{u,j}$ | % no high school diploma | Census | 1990 | $M/F, W/B/O, 9A$ |
| $COLL_i$ | % with college degree | Census | 1990 | $M/F$ |
| $COLL_{u,j}$ | % with college degree | Census | 1990 | $M/F, W/B/O, 9A$ |
| $POP_{i,j}$ | population | Census | 1990 | $M/F, W/B/O, 9A$ |
| $CCWC_i$ | % couples with children | Census | 1990 | |
| $CCWC_{u,j}$ | % couples with children | Census | 1990 | $W/B/O$ |
| $HHLD_{i,j}$ | households | Census | 1990 | $W/B/O$ |
| $HISP_i$ | % Hispanic | Census | 1990 | $W/B/O$ |
| $RURAL_i$ | % rural | Census | 1990 | |
| | | Cross-Section | | |
| $PR_i$ | participation rate | Census | 1990 | $M/F, W/B/O, 20$ |
| $NR_i$ | natality rate | Vital | 1990 | $F, W/B/O, 7FAC$ |
| $NOH_i$ | % no high school diploma | Census | 1990 | $M/F, W/B/O, 9A$ |
| $COL_i$ | % with college degree | Census | 1990 | $M/F, W/B/O, 9A$ |
| $HISP_i$ | % Hispanic | Census | 1990 | $W/B/O$ |
| $CWC_i$ | % couples with children | Census | 1990 | $W/B/O$ |

## References


Bartik, Timothy J. (1992) "Who Benefits from Local Job Growth: Migrants or the Original Residents?" *Regional Studies*, 27, 4.

Bartik, Timothy J. (1991) *Who Benefits from State and Local Development Policies?*, W. E. Upjoin Institute for Employment Research, Kalamazoo, MI.

Becker, Gary S. (1994) *A Treatise on the Family*, Harvard University Press.

Blanchard, Olivier J. and Lawarence F. Katz (1992) "Regional Evolutions", *Brookings Paper on Economic Activity*, 1.

Clark, Kim B. and L. H. Summers (1982) "Labor force Participation: Timing and Persistence", in L. H. Summers, ed., *Understanding Unemployment*, 1990.

Cogan, John F. (1981) "Fixed Costs and Labor Supply", *Econometrica*, 49, 4.

Eckstein, Zvi and Kenneth I. Wolpin (1989) "Dynamic Labor Force Participation of Married Women and Endogenous Work Experience", *Review of Economic Studies*, 56, 375-390.

Gay, Robert S. and William L. Wascher (1989) "Persistence Effects in Labor Force Participation", *Eastern Economic Journal*, October-December 1989, 15, 4, 177-187.

Grounau, Beuben (1986) "Home Production - A Survey", in Orley Ashenfelter and Richard Layard, ed., *Handbook of Labor Economics Vol.* I, Elsevier Science Publishers B.V.

Krugman, Paul (1993) "First Nature, Second Nature and Metropolitan Location", Journal of Regional Science, 33,2 .

Treyz, George I. (1993) *Regional Economic Modeling: A Systematic Approach to Economic Forecasting and Policy Analysis*, Kluwer.

Table 1. U.S. Population Distribution by Race and Gender

| | 1967 | | 1993 | |
|---|---|---|---|---|
| | Male | Female | Male | Female |
| %White | 89.6 | 89.2 | 85.4 | 84.1 |
| %Black | 9.2 | 9.7 | 10.9 | 12.1 |
| %Other | 1.2 | 1.1 | 3.7 | 3.8 |

Table 2. U.S. Population Distribution by Race for Two Age Cohorts

| | 1967 | | 1993 | |
|---|---|---|---|---|
| Age | 16-17 | 65 + | 16-17 | 65 + |
| %White | 86.4 | 92.0 | 79.7 | 89.3 |
| %Black | 12.4 | 7.4 | 16.0 | 8.6 |
| %Other | 1.2 | 0.6 | 4.3 | 2.1 |

Table 3. Time Variant Panel Estimates

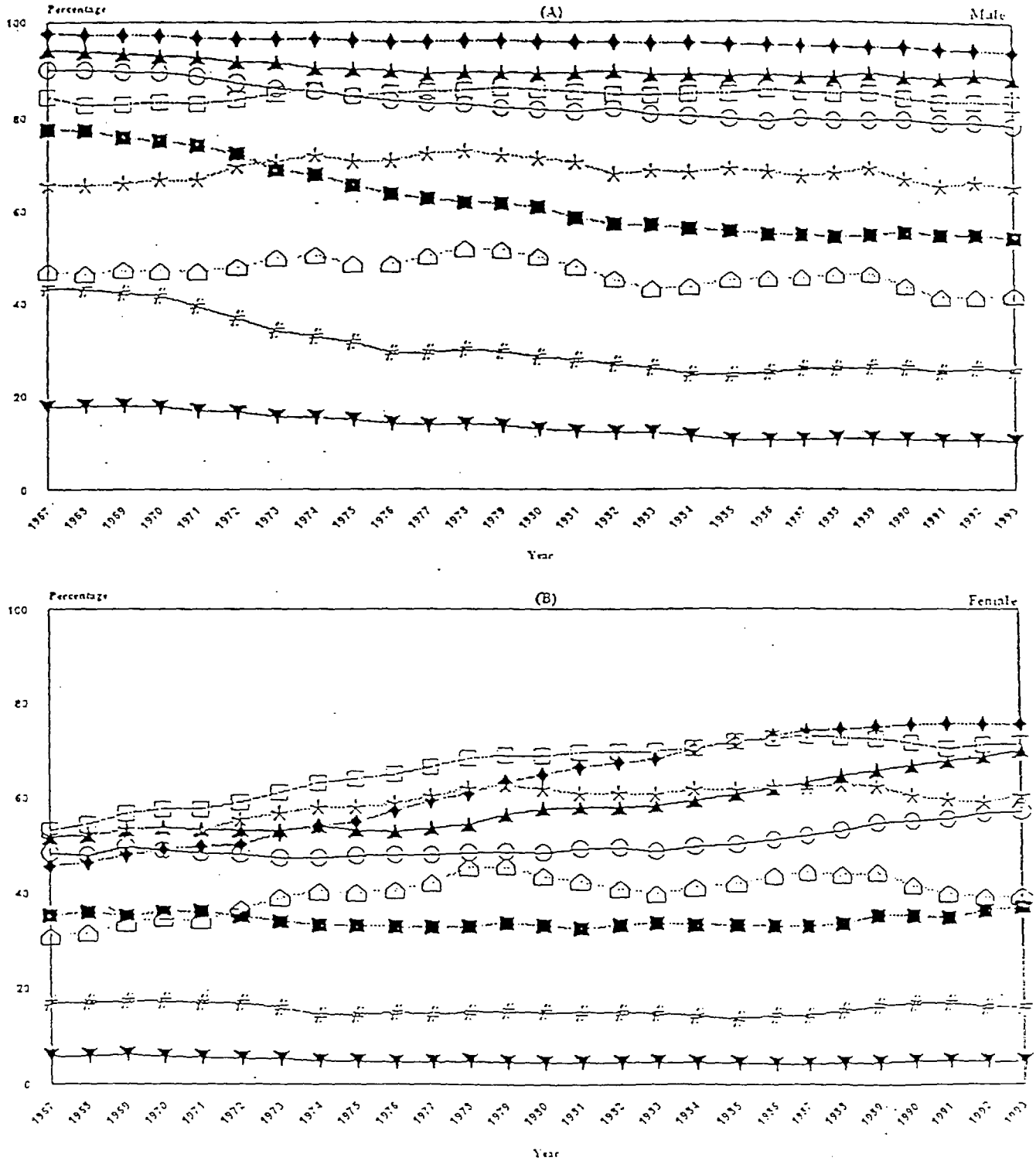
| | Male | | Female | |
|---|---|---|---|---|
| | LS* | TSLS | LS | TSLS* |
| Intercept | 3.309 | 3.430 | 3.494 | 3.736 |
| | (32.1) | (25.0) | (21.2) | (16.2) |
| RWR | 0.052 | 0.044 | -0.116 | -0.278 |
| | (2.99) | (1.56) | (-3.72) | (-5.87) |
| EA(λ) | 0.208 | 0.185 | 0.200 | 0.253 |
| | (9.65) | (5.84) | (6.42) | (5.29) |
| Adjusted $R^2$ | 0.85 | 0.84 | 0.94 | 0.94 |
| Optimal λ | 0.4 | | 1.0 | |

* Prefered estimates

Table 4. Time-Invariant Panel Estimates

| | Male | Female |
|---|---|---|
| Intercept | 0.05787 | 0.30504 |
| | (1.02) | (2.00) |
| NR | | -0.00068 |
| | | (-0.49) |
| NOH | -0.00136 | -0.00553 |
| | (-1.55) | (-3.01) |
| COL | -0.00114 | 0.00227 |
| | (-1.20) | (0.98) |
| Rural | -0.00015 | -0.00044 |
| | (-0.45) | (-0.75) |
| CWC | 0.00029 | -0.00448 |
| | (0.23) | (-1.66) |
| HISP | -0.00043 | -0.00196 |
| | (-0.82) | (-1.77) |
| F-test | 0.93 | 8.63 |
| Adjusted $R^2$ | 0.09 | 0.48 |

Figure 1. U.S. Labor Force Participation Rate
(1967-1993)




16-17 18-19 20-24 35-39 50-54 55-59 60-64 65-69 70 + over

Figure 2. U.S. Labor Force Participation Rate by Race and Gender (1967 and 1993)

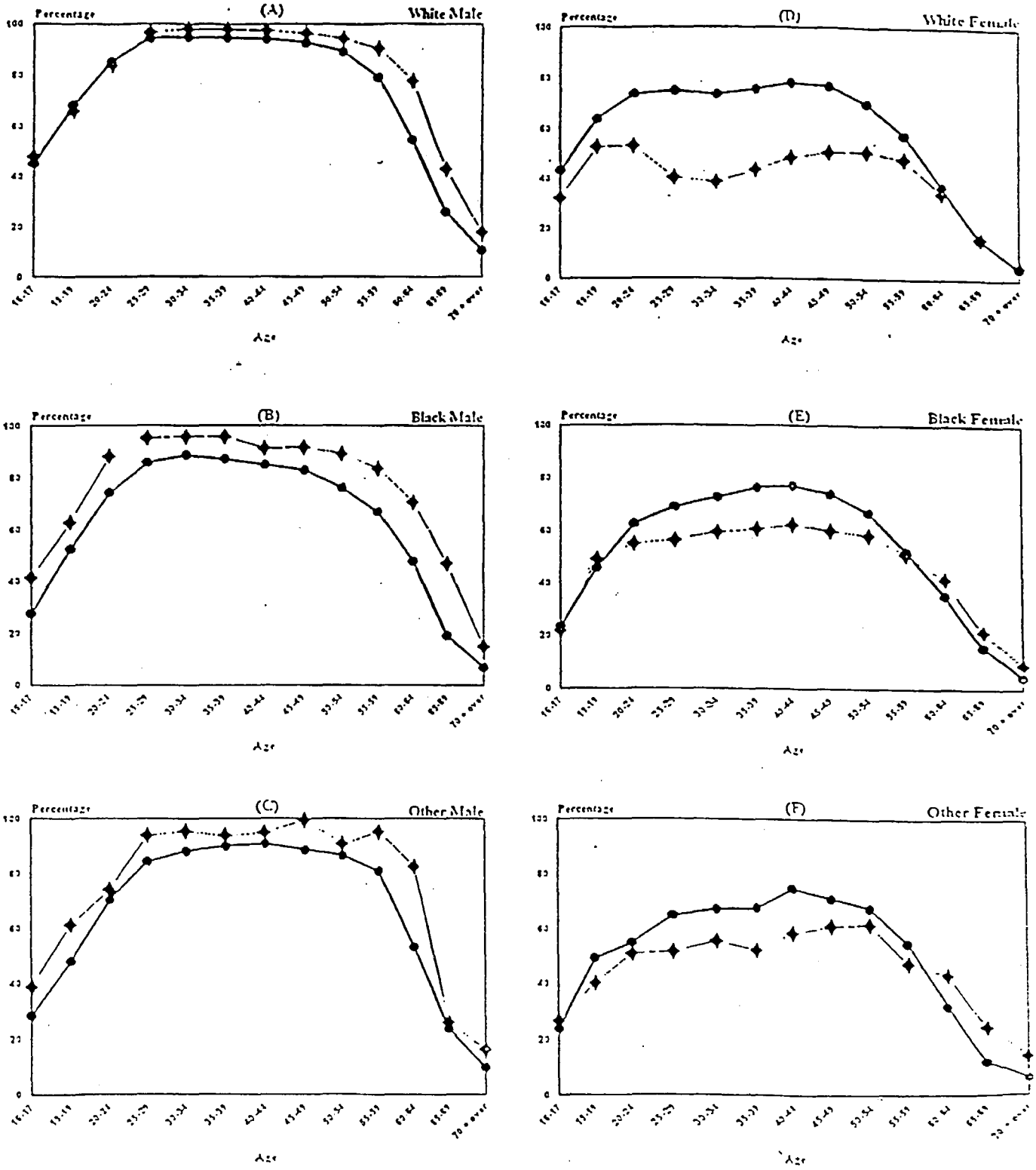


Figure 3. U.S. Labor Force Participation Rate by Gender and Ethnic Breakdown of Whites (1990)

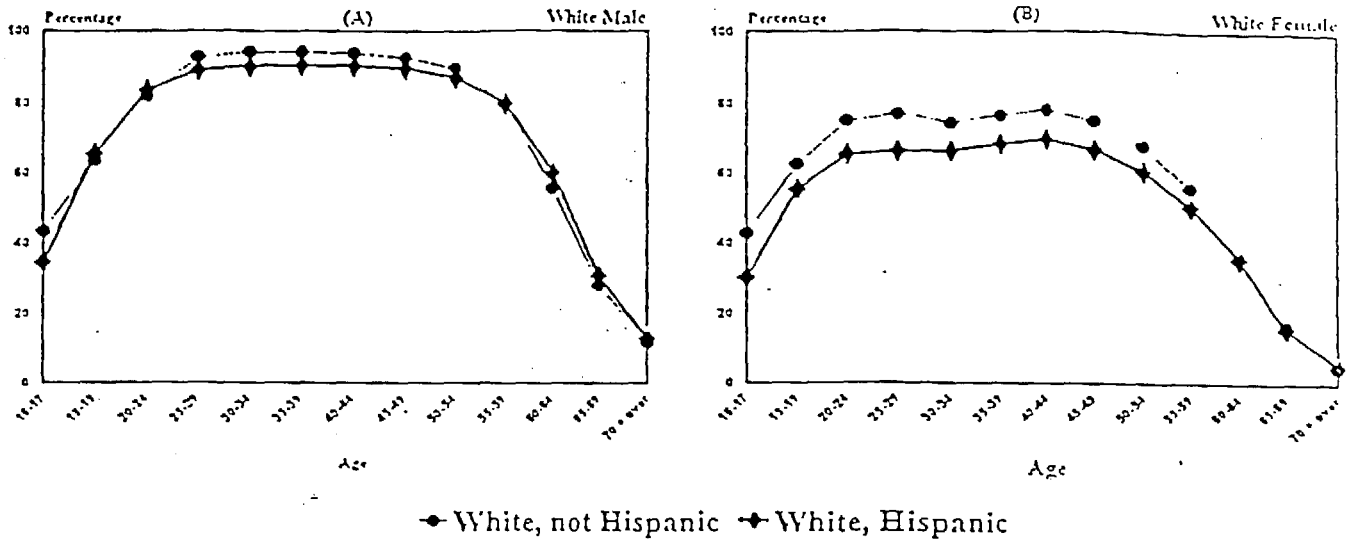


Figure 4a. Participation Rate vs. Participation Rate Adjusted for Age and Race
(Rates for Population over Age 16 for 1993)

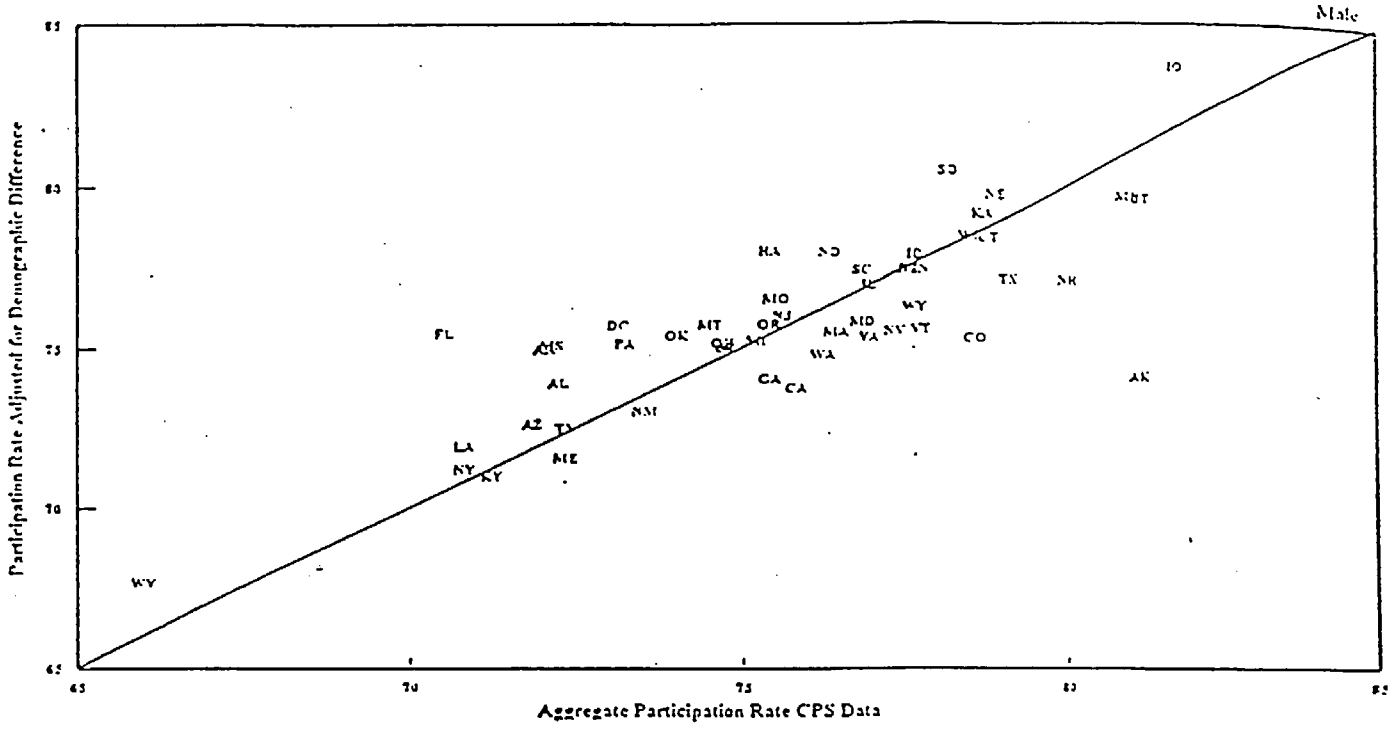



Figure 4b. Participation Rate vs. Participation Rate Adjusted for Age and Race
(Rates for Population over Age 16 for 1993)

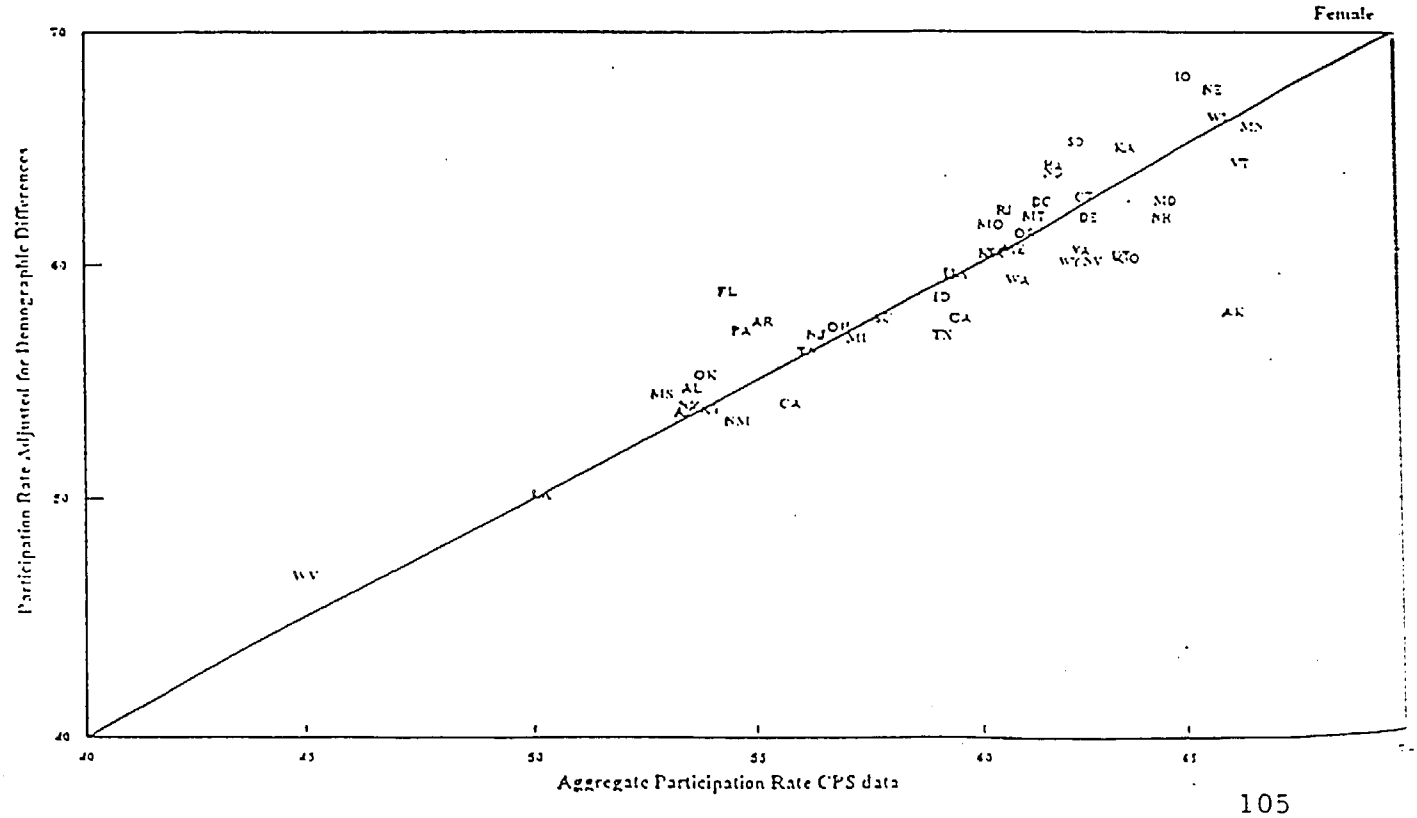


Figure 5a. Adjusted State Participation Rate
(1977 and 1993)

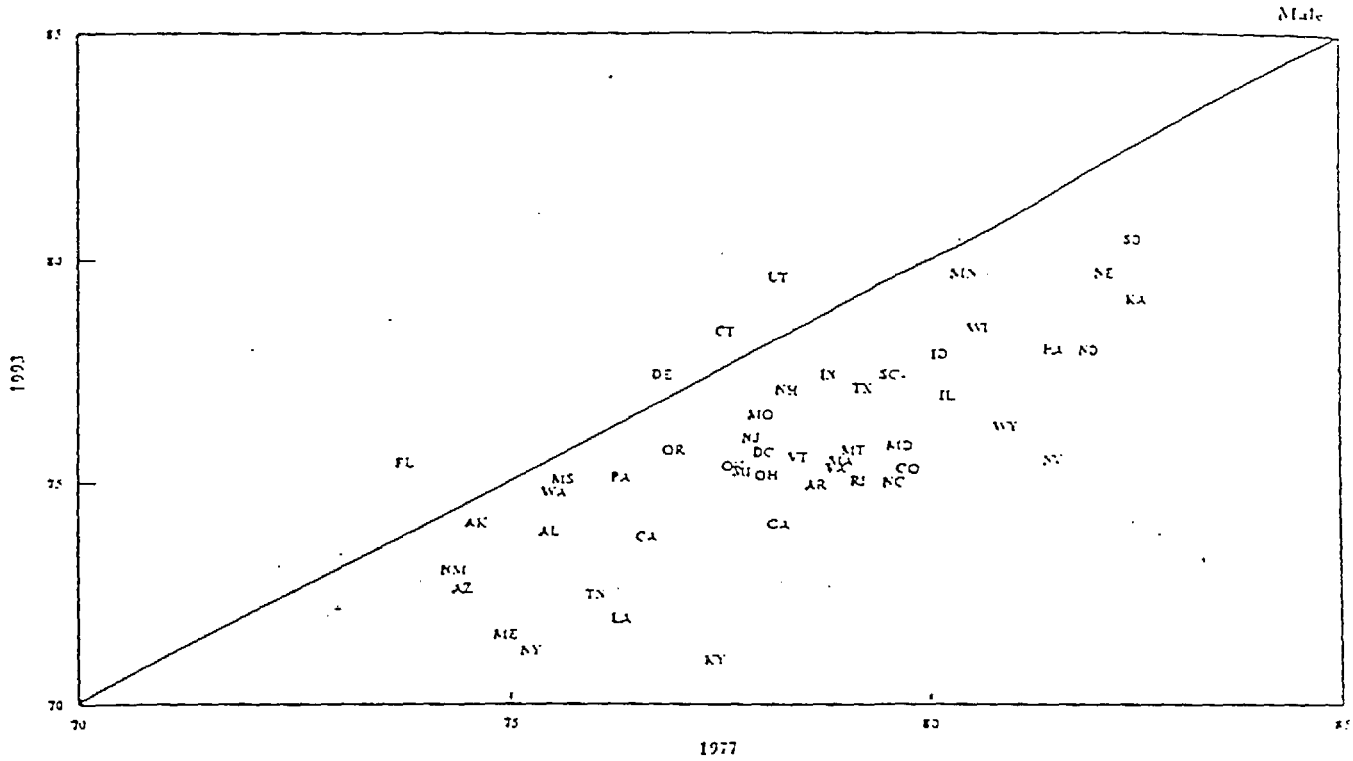



Figure 5b. Adjusted State Participation Rate
(1977 and 1993)

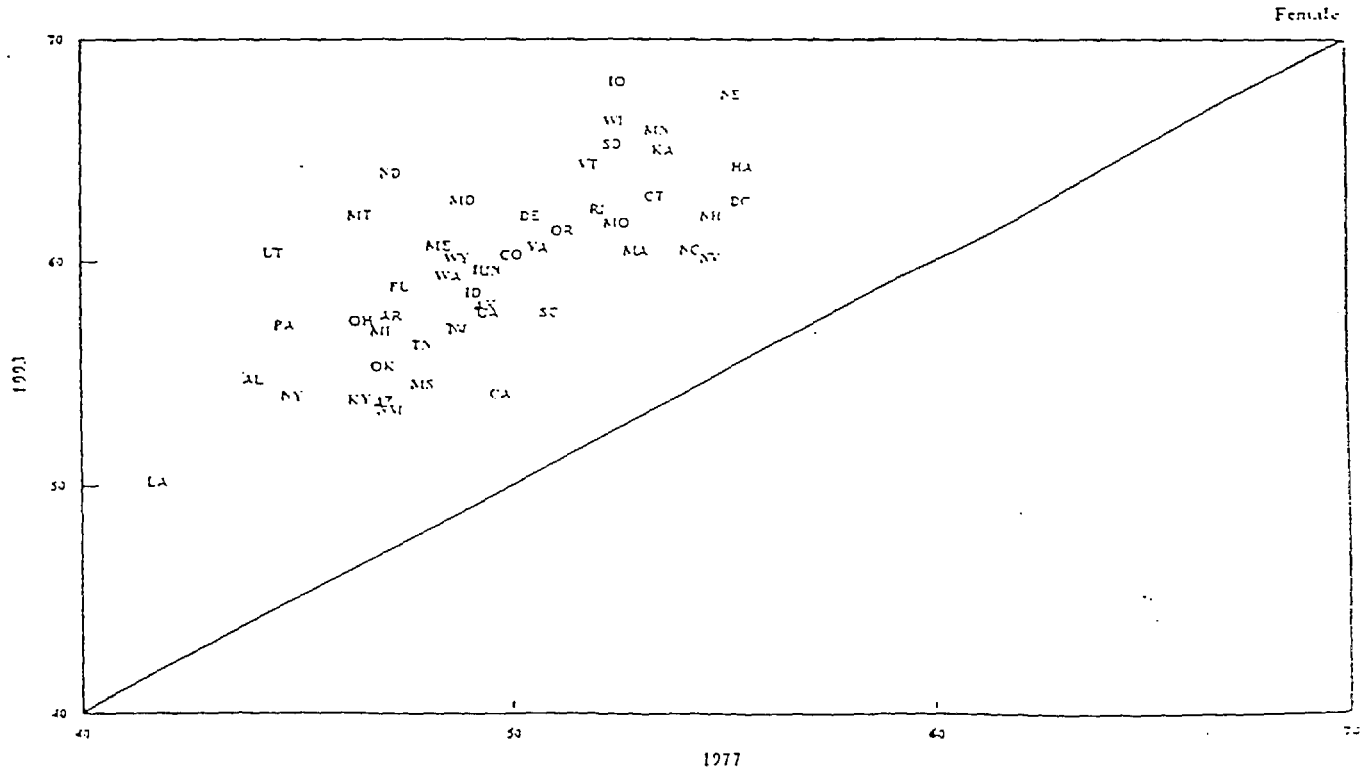
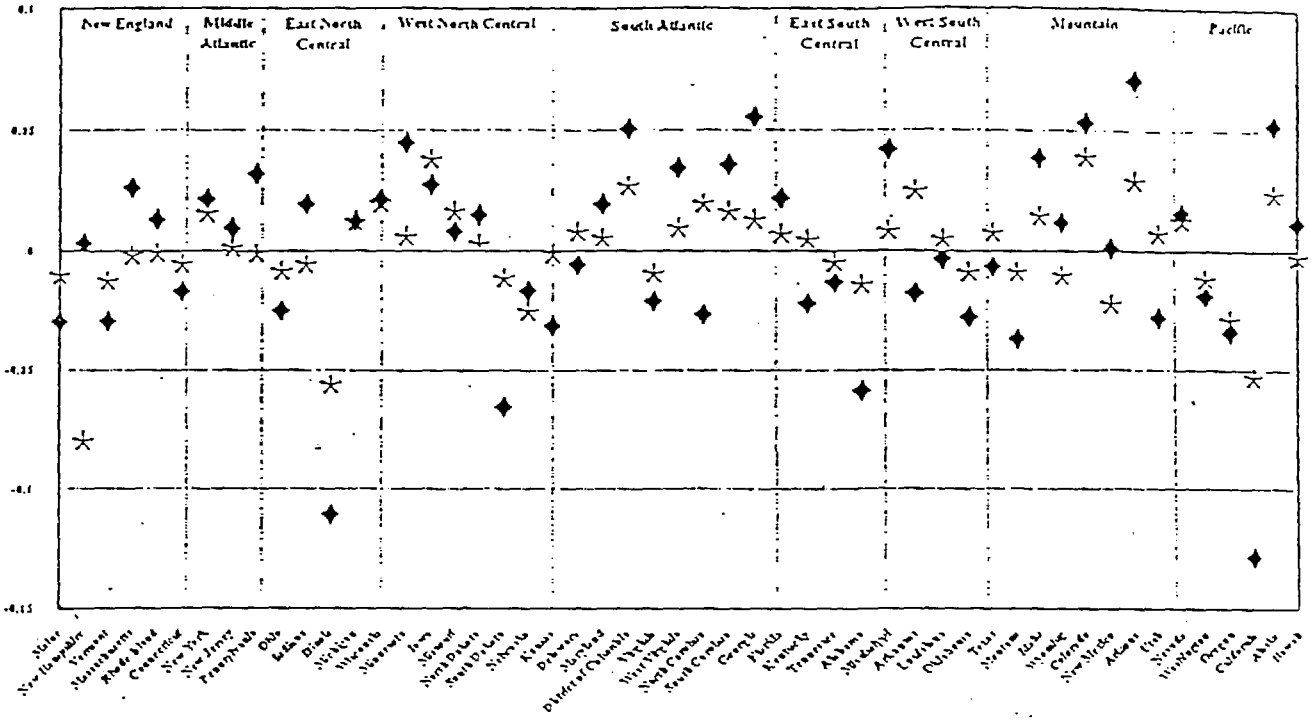


Figure 6. Period Effects in Labor Force Participation Estimation Using Panel Data

Figure 7. State Specific Effects from the Panel Regression




◆ Female ✳ Male

Figure 8. Long Term Percentage Change in Participation Rate for 1% Change in Employment Opportunity (EA)

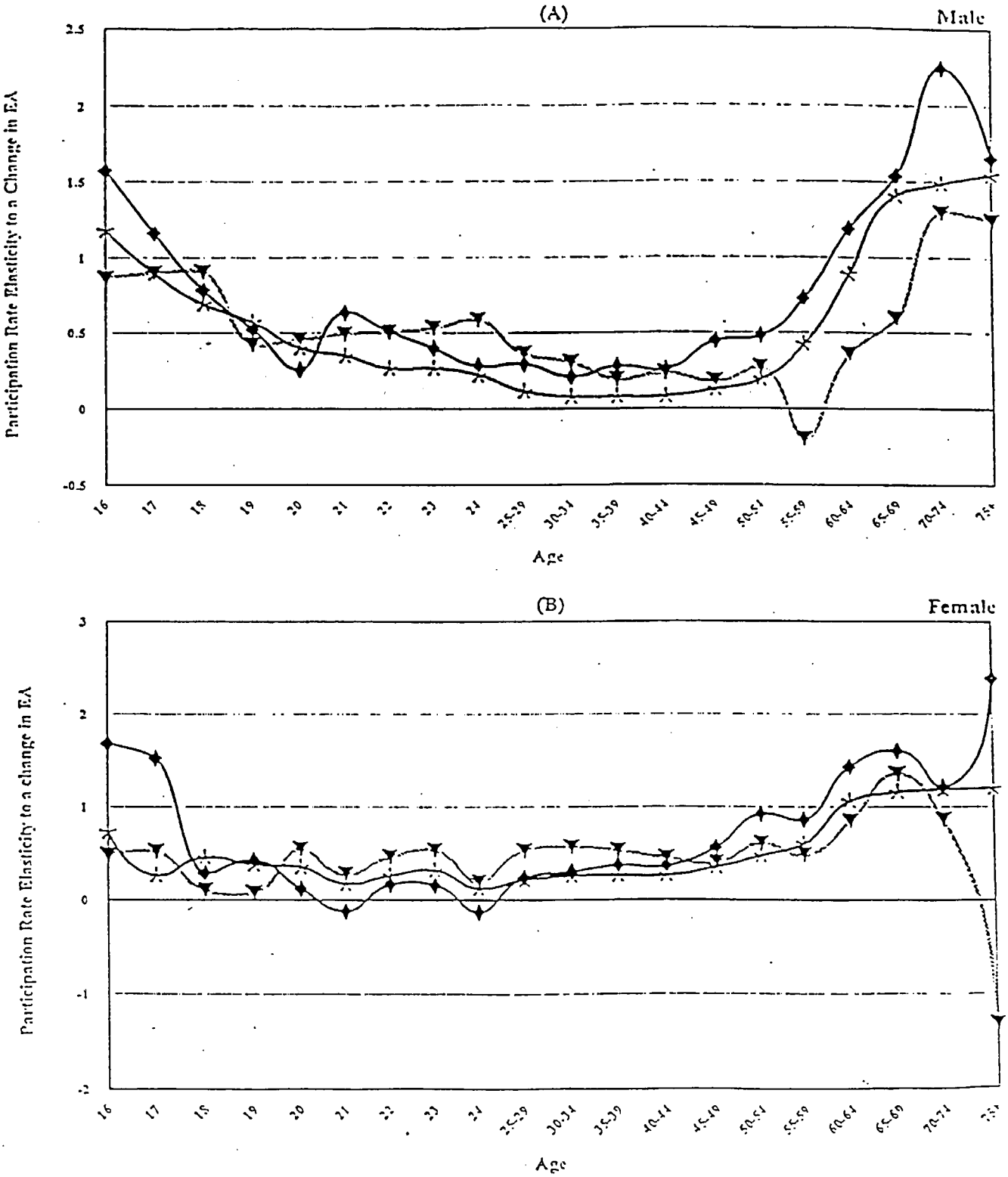


Figure 9. Percentage Change in Participation Rate for 1% Change in Employment Opportunity (EA)

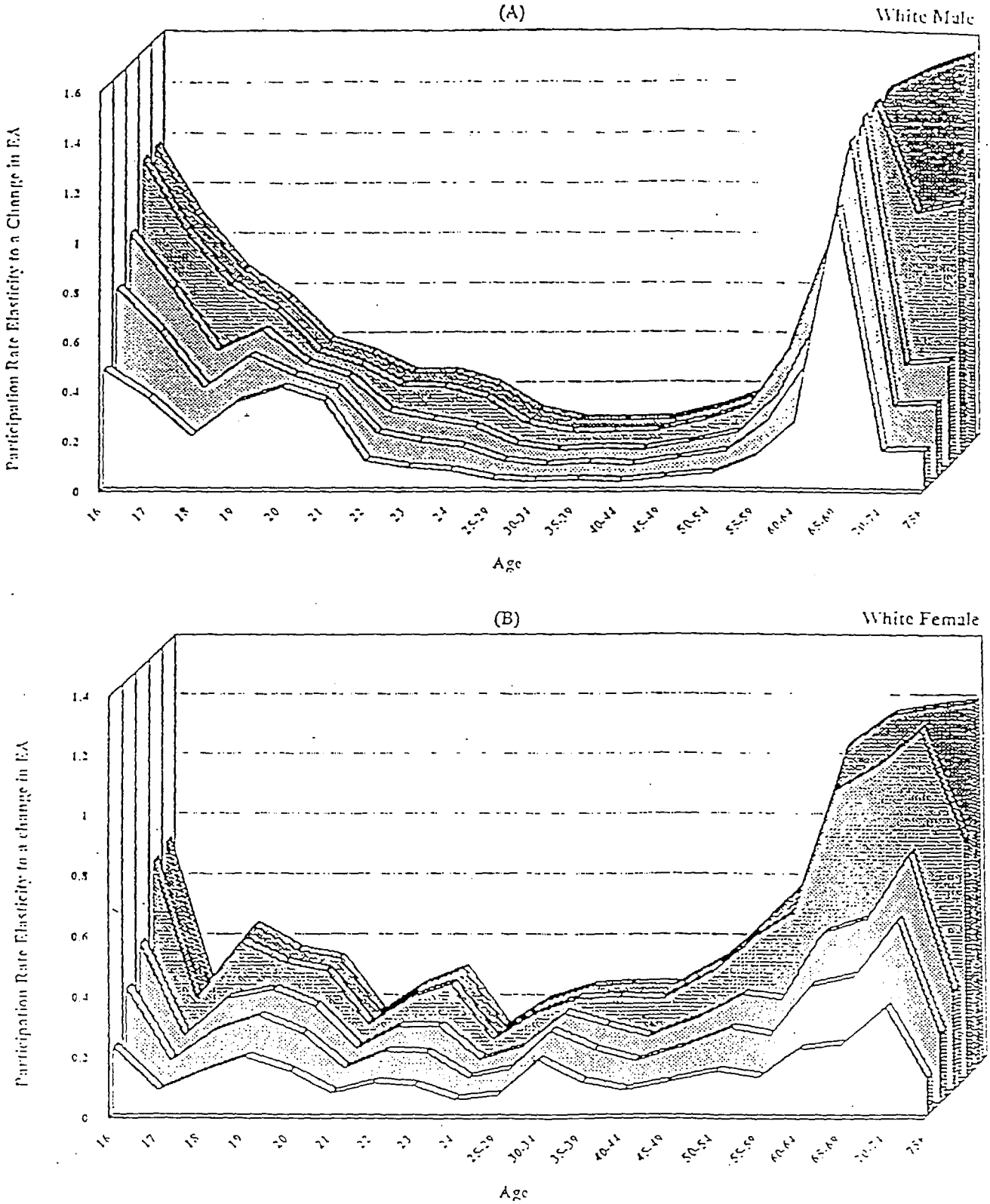


Table 5. Labor Force Participation Equation (3) Estimates Using Cross-Section Data (1990) -- White Male

| Age Cohort | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | -0.37 | 0.87 | 1.20 | 1.65 | 2.35 | 2.58 | 3.11 | 3.23 | 3.36 | 3.73 | 3.94 | 3.90 | 3.90 | 3.58 | 3.15 | 1.67 | -1.93 | -5.38 | -5.92 | -7.62 |
| | (-0.18) | (0.65) | (1.37) | (2.34) | (4.12) | (4.65) | (7.35) | (9.49) | (10.70) | (22.50) | (30.17) | (28.55) | (29.18) | (21.97) | (13.69) | (4.31) | (-2.81) | (-4.46) | (-4.75) | (-4.18) |
| FE | 3.48 | 2.80 | 1.41 | -0.28 | -0.18 | -0.02 | 0.04 | 0.01 | 0.05 | 0.40 | 0.33 | 0.30 | 0.38 | 0.43 | 0.63 | 1.27 | 2.39 | 3.84 | 3.95 | 2.26 |
| | (3.35) | (4.09) | (3.38) | (-0.87) | (-0.69) | (-0.07) | (0.18) | (0.09) | (0.36) | (5.36) | (5.69) | (4.88) | (6.26) | (5.49) | (5.73) | (6.73) | (6.95) | (6.37) | (5.83) | (2.23) |
| NOH 1824 | -0.77 | 0.14 | | | | | | | | | | | | | | | | | | |
| | (-1.13) | (0.32) | | | | | | | | | | | | | | | | | | |
| COL 2534 | | | -1.24 | -0.91 | -0.70 | -0.36 | -0.26 | -0.31 | -0.30 | | | | | | | | | | | |
| | | | (-5.87) | (-5.55) | (-5.31) | (-2.85) | (-2.62) | (-3.73) | (-4.01) | | | | | | | | | | | |
| NOH | | | 0.29 | 0.62 | 0.58 | 0.52 | 0.39 | 0.03 | 0.04 | 0.02 | -0.04 | -0.04 | -0.10 | -0.13 | -0.23 | -0.14 | -0.05 | 0.14 | 0.08 | 0.17 |
| | | | (0.95) | (2.55) | (2.95) | (2.71) | (2.62) | (0.25) | (0.36) | (0.41) | (-0.97) | (-1.03) | (-2.41) | (-3.11) | (-3.94) | (-1.57) | (-0.31) | (0.61) | (0.29) | (0.42) |
| FCOL | | | | | | | | | | -0.09 | -0.00 | 0.02 | -0.00 | -0.03 | -0.08 | -0.13 | -0.14 | 0.71 | 0.68 | 1.03 |
| | | | | | | | | | | (-2.53) | (-0.11) | (0.51) | (-0.04) | (-1.04) | (-1.78) | (-1.61) | (-0.79) | (2.38) | (2.02) | (1.84) |
| RWR | -0.37 | -0.33 | 0.09 | 0.14 | 0.14 | 0.12 | 0.08 | 0.07 | 0.10 | 0.12 | 0.09 | 0.10 | 0.09 | 0.14 | 0.19 | 0.33 | 0.75 | 0.85 | 0.77 | 0.99 |
| | (-1.03) | (-1.40) | (0.52) | (0.98) | (1.24) | (1.05) | (0.98) | (1.05) | (1.49) | (3.54) | (3.35) | (3.82) | (3.72) | (4.75) | (4.50) | (4.66) | (5.87) | (3.89) | (3.15) | (2.79) |
| RURAL | -0.23 | -0.25 | -0.38 | -0.31 | -0.24 | -0.10 | -0.14 | -0.03 | -0.02 | 0.01 | 0.00 | 0.01 | -0.02 | -0.03 | -0.01 | -0.04 | -0.01 | 0.15 | 0.28 | 0.25 |
| | (-1.17) | (-1.94) | (-4.75) | (-4.96) | (-4.79) | (-2.17) | (-3.57) | (-0.84) | (-0.87) | (0.55) | (0.31) | (0.53) | (-1.48) | (-1.68) | (-0.62) | (-1.00) | (-0.17) | (1.19) | (1.97) | (1.26) |
| EA | 1.18 | 0.90 | 0.69 | 0.57 | 0.40 | 0.35 | 0.27 | 0.27 | 0.22 | 0.12 | 0.08 | 0.08 | 0.08 | 0.13 | 0.19 | 0.42 | 0.89 | 1.41 | 1.49 | 1.55 |
| | (3.77) | (4.33) | (5.37) | (5.42) | (4.65) | (4.15) | (4.26) | (5.29) | (4.81) | (4.61) | (4.01) | (3.72) | (3.94) | (4.93) | (5.23) | (6.92) | (8.21) | (7.44) | (7.72) | (5.57) |
| CWC | 0.78 | 0.72 | 0.10 | -0.39 | -0.23 | -0.21 | -0.21 | -0.15 | -0.06 | -0.04 | 0.06 | 0.04 | 0.10 | 0.08 | 0.10 | 0.03 | 0.01 | | | |
| | (1.15) | (1.61) | (0.31) | (-1.51) | (-1.09) | (-1.01) | (-1.34) | (-1.18) | (-0.49) | (-0.61) | (1.37) | (0.87) | (2.01) | (1.31) | (1.14) | (0.23) | (0.04) | | | |
| HISP | -0.52 | -0.53 | -0.33 | -0.13 | -0.13 | -0.09 | -0.21 | -0.07 | -0.07 | -0.07 | -0.04 | -0.04 | -0.03 | -0.02 | -0.03 | -0.08 | 0.10 | 0.30 | 0.26 | 0.76 |
| | (-1.17) | (-1.80) | (-1.79) | (-0.88) | (-1.12) | (-0.84) | (-2.44) | (-0.92) | (-1.08) | (-2.32) | (-1.55) | (-1.44) | (-1.20) | (-0.69) | (-0.53) | (-0.97) | (0.69) | (1.14) | (0.89) | (1.81) |
| R-squared | 0.61 | 0.65 | 0.80 | 0.75 | 0.74 | 0.52 | 0.54 | 0.58 | 0.60 | 0.63 | 0.70 | 0.69 | 0.77 | 0.80 | 0.80 | 0.81 | 0.83 | 0.80 | 0.74 | 0.61 |
| Optimal-lambda | 0.4 | 0.4 | 0.3 | 0.6 | 1.0 | 1.0 | 0.4 | 0.3 | 0.3 | 0.3 | 0.3 | 0.4 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.8 | 0.1 | 0.1 |

Note: The bold print indicates significance for one tailed testt the 5 percent level (t=1.681) or two tailed testt the 10 percent level of significance.

Table 6. Labor Force Participation Equation (3) Estimates Using Cross-Section Data (1990) -- Black Male

| Age Cohort | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | 0.81 | 2.20 | 4.09 | 3.08 | 3.71 | 1.79 | -0.42 | 1.95 | 1.59 | 2.58 | 2.32 | 2.66 | 3.02 | 2.32 | 2.15 | 0.65 | -2.59 | -4.09 | -10.56 | -5.33 |
| | (0.25) | (0.87) | (2.10) | (2.25) | (2.30) | (1.96) | (-0.32) | (1.75) | (2.09) | (4.30) | (5.42) | (5.96) | (6.64) | (3.95) | (3.09) | (0.70) | (-1.87) | (-1.80) | (-2.84) | (-1.55) |
| FE | -0.12 | 1.84 | 0.24 | -1.65 | -0.99 | 0.94 | -1.19 | 0.61 | -0.68 | -0.58 | -0.50 | 0.12 | -0.46 | -0.53 | -0.78 | -0.43 | -0.76 | -2.79 | 2.45 | 1.92 |
| | (-0.04) | (0.85) | (0.45) | (-1.57) | (-0.78) | (1.24) | (-1.09) | (0.65) | (-1.03) | (-1.05) | (-1.28) | (0.30) | (-1.07) | (-1.15) | (-1.42) | (-0.55) | (-0.60) | (-1.53) | (0.73) | (0.69) |
| NOH 1824 | -0.65 | -0.58 | | | | | | | | | | | | | | | | | | |
| | (-0.86) | (-0.98) | | | | | | | | | | | | | | | | | | |
| COL 2534 | | | -0.23 | -0.37 | 0.49 | -0.17 | 0.22 | 0.53 | 0.24 | | | | | | | | | | | |
| | | | (-0.22) | (-0.52) | (0.57) | (-0.37) | (0.32) | (0.94) | (0.62) | | | | | | | | | | | |
| NOH | | | -0.31 | -0.49 | 0.55 | 0.16 | 0.10 | 0.45 | -0.19 | 0.44 | 0.15 | 0.19 | 0.20 | -0.14 | -0.19 | -0.16 | 0.01 | 0.34 | 0.67 | -0.41 |
| | | | (-0.50) | (-1.23) | (1.21) | (0.34) | (0.25) | (1.26) | (-0.78) | (2.56) | (1.29) | (1.95) | (1.92) | (-1.63) | (-1.88) | (-1.17) | (0.05) | (0.85) | (0.95) | (-0.51) |
| FCOL | | | | | | | | | | 0.61 | 0.45 | 0.25 | 0.06 | -0.36 | -0.41 | -0.47 | -0.38 | -2.17 | 0.04 | 1.57 |
| | | | | | | | | | | (2.14) | (2.06) | (1.28) | (0.27) | (-1.60) | (-1.50) | (-1.15) | (-0.47) | (-1.35) | (0.01) | (0.68) |
| RWR | -1.16 | -1.11 | -1.05 | -0.31 | -0.30 | -0.08 | 0.78 | 0.09 | 0.47 | 0.09 | 0.31 | 0.11 | 0.07 | 0.10 | 0.10 | 0.18 | 0.46 | 0.21 | 0.98 | 0.13 |
| | (-1.54) | (-1.89) | (-2.32) | (-0.98) | (-0.80) | (-0.40) | (2.56) | (0.36) | (2.74) | (0.67) | (3.23) | (1.11) | (0.64) | (0.75) | (0.62) | (0.82) | (1.40) | (0.50) | (1.45) | (0.22) |
| RURAL | -0.02 | -0.23 | -0.24 | -0.33 | -0.30 | 0.17 | 0.29 | 0.17 | 0.11 | 0.10 | 0.18 | 0.10 | -0.07 | -0.05 | 0.00 | -0.01 | -0.05 | -0.66 | 0.40 | 0.16 |
| | (-0.06) | (-0.79) | (-1.00) | (-2.10) | (-1.62) | (1.56) | (1.89) | (1.27) | (1.21) | (1.34) | (3.18) | (1.65) | (-1.04) | (-0.68) | (0.02) | (-0.07) | (-0.27) | (-2.24) | (0.85) | (0.39) |
| EA | 1.57 | 1.16 | 0.79 | 0.52 | 0.26 | 0.63 | 0.51 | 0.40 | 0.28 | 0.30 | 0.21 | 0.28 | 0.26 | 0.45 | 0.48 | 0.73 | 1.19 | 1.54 | 2.24 | 1.66 |
| | (2.70) | (2.61) | (2.04) | (2.12) | (0.91) | (3.58) | (1.89) | (1.85) | (1.82) | (2.41) | (2.54) | (3.17) | (2.95) | (4.44) | (4.19) | (4.37) | (4.54) | (4.19) | (3.68) | (3.24) |
| CWC | -1.02 | -0.96 | -1.99 | -0.52 | 0.19 | -0.71 | 0.47 | 0.56 | 0.87 | 0.20 | 0.54 | 0.41 | 0.20 | 0.10 | -0.12 | 0.44 | 0.11 | | | |
| | (-2.58) | (-0.80) | (-2.02) | (-0.77) | (0.23) | (-1.62) | (0.67) | (1.04) | (2.20) | (0.84) | (3.03) | (2.37) | (1.04) | (0.41) | (-0.42) | (0.98) | (0.14) | | | |
| HISP | -0.81 | -0.54 | 0.11 | -0.64 | -0.81 | 0.48 | -0.11 | -0.22 | -0.45 | 0.12 | 0.19 | 0.37 | -0.04 | 0.08 | 0.72 | 1.06 | 2.68 | 5.41 | 7.19 | 5.39 |
| | (-0.44) | (-0.38) | (0.09) | (-0.77) | (-0.81) | (0.90) | (-0.15) | (-0.34) | (-1.03) | (0.33) | (0.74) | (1.32) | (-0.13) | (0.30) | (2.35) | (2.56) | (3.24) | (3.79) | (3.38) | (2.50) |
| R-squared | 0.47 | 0.56 | 0.46 | 0.43 | 0.23 | 0.50 | 0.39 | 0.31 | 0.54 | 0.37 | 0.69 | 0.62 | 0.48 | 0.56 | 0.55 | 0.61 | 0.70 | 0.71 | 0.61 | 0.64 |
| Optimal-lambda | 0.4 | 0.5 | 0.4 | 0.5 | 0.4 | 0.4 | 0.1 | 0.4 | 0.1 | 0.4 | 0.5 | 0.9 | 0.4 | 0.1 | 0.5 | 0.1 | 0.1 | 0.3 | 0.6 | 0.3 |

Note: The bold print indicates significance for one tailed testt the 5 percent level (t=1.684) or two tailed testt the 10 percent level of significance.

Table 7. Labor Force Participation Equation (3) Estimates Using Cross-Section Data (1990) -- Other Male

| Age Cohort | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | 1.73 | 1.02 | 0.56 | 3.40 | 2.51 | 1.99 | 2.67 | 2.71 | 1.87 | 2.90 | 3.68 | 3.90 | 3.48 | 3.47 | -3.48 | -4.80 | -0.04 | 0.74 | -3.09 | -1.81 |
| | (0.52) | (0.35) | (0.26) | (1.94) | (1.52) | (1.54) | (1.96) | (2.38) | (1.90) | (4.09) | (5.89) | (7.53) | (6.89) | (5.93) | (5.69) | (5.52) | (-0.02) | (0.28) | (-0.98) | (-0.27) |
| FE | 1.48 | -0.34 | 0.25 | -0.57 | -0.98 | -1.25 | -1.55 | -0.94 | -0.14 | -0.93 | -1.01 | -0.88 | -0.54 | -0.72 | -0.34 | 0.74 | -0.71 | -0.61 | 1.04 | 3.23 |
| | (0.71) | (-0.17) | (0.16) | (-0.47) | (-0.87) | (-1.41) | (-1.60) | (-1.11) | (-0.21) | (-1.92) | (-2.36) | (-2.57) | (-1.62) | (-1.84) | (-0.80) | (1.33) | (-0.80) | (-0.31) | (0.43) | (0.80) |
| NOH 1824 | -1.01 | -1.05 | | | | | | | | | | | | | | | | | | |
| | (-1.55) | (-1.85) | | | | | | | | | | | | | | | | | | |
| COL 2534 | | | 0.09 | -0.18 | -0.34 | -0.18 | 0.09 | 0.08 | -0.13 | | | | | | | | | | | |
| | | | (0.23) | (-0.55) | (-1.21) | (-0.80) | (0.36) | (0.36) | (-0.75) | | | | | | | | | | | |
| NOH | | | 0.48 | -0.18 | 0.53 | 0.47 | 0.69 | 0.40 | 0.39 | 0.30 | 0.31 | 0.03 | 0.10 | -0.18 | -0.20 | -0.93 | -0.85 | -1.09 | -3.86 | -3.84 |
| | | | (0.88) | (-0.36) | (1.33) | (1.51) | (2.09) | (1.25) | (1.57) | (1.72) | (1.98) | (0.24) | (0.82) | (-1.21) | (-1.15) | (-4.13) | (-3.02) | (-2.26) | (-5.83) | (-2.20) |
| FCOL | | | | | | | | | | 0.08 | 0.30 | 0.30 | 0.39 | 0.29 | 0.35 | -0.13 | 0.51 | 1.23 | -0.83 | -4.85 |
| | | | | | | | | | | (0.65) | (2.86) | (3.84) | (4.99) | (2.60) | (2.65) | (-0.58) | (1.21) | (1.56) | (-0.59) | (-1.61) |
| RWR | -0.75 | -0.57 | -0.29 | -0.47 | -0.24 | -0.09 | -0.33 | -0.33 | -0.12 | -0.11 | -0.28 | -0.18 | -0.11 | 0.01 | -0.15 | 0.20 | 0.91 | 0.12 | 0.63 | 0.47 |
| | (-1.15) | (-1.01) | (-0.63) | (-1.17) | (-0.69) | (-0.33) | (-1.15) | (-1.27) | (-0.59) | (-0.71) | (-2.02) | (-1.61) | (-1.06) | (0.05) | (-0.92) | (0.90) | (2.52) | (0.20) | (0.99) | (0.34) |
| RURAL | -0.21 | 0.27 | 0.08 | -0.24 | -0.20 | -0.08 | -0.04 | -0.08 | -0.18 | -0.10 | 0.01 | -0.00 | 0.03 | -0.02 | 0.02 | -0.19 | 0.06 | -0.04 | 1.08 | -0.09 |
| | (-0.39) | (0.58) | (0.28) | (-0.97) | (-1.00) | (-0.49) | (-0.22) | (-0.44) | (-1.42) | (-1.09) | (0.14) | (-0.07) | (0.44) | (-0.22) | (0.18) | (-1.69) | (0.28) | (-0.09) | (1.71) | (-0.09) |
| EA | 0.87 | 0.90 | 0.90 | 0.43 | 0.46 | 0.49 | 0.51 | 0.53 | 0.59 | 0.37 | 0.31 | 0.20 | 0.24 | 0.18 | 0.27 | -0.20 | 0.36 | 0.59 | 1.30 | 1.25 |
| | (1.53) | (1.81) | (2.61) | (1.49) | (1.79) | (2.47) | (2.41) | (2.89) | (3.84) | (3.06) | (2.93) | (2.29) | (2.72) | (1.76) | (2.49) | (-1.31) | (1.51) | (1.43) | (2.47) | (1.18) |
| CWC | 0.60 | 1.13 | 0.02 | 0.73 | 0.93 | 0.73 | 0.19 | 0.28 | 0.68 | 0.39 | 0.15 | 0.23 | 0.32 | 0.25 | 0.33 | 0.54 | 0.17 | | | |
| | (0.63) | (1.37) | (0.03) | (1.33) | (2.01) | (2.02) | (0.49) | (0.78) | (2.40) | (1.97) | (0.85) | (1.54) | (2.16) | (1.43) | (1.58) | (2.23) | (0.38) | | | |
| HISP | 0.46 | 0.66 | 0.24 | 0.32 | -0.01 | 0.10 | 0.19 | 0.22 | 0.06 | 0.11 | 0.15 | 0.18 | 0.15 | 0.23 | 0.19 | 0.26 | 0.49 | 0.54 | 1.59 | 0.55 |
| | (1.27) | (2.10) | (1.12) | (1.70) | (-0.05) | (0.76) | (1.42) | (1.70) | (0.55) | (1.52) | (2.27) | (3.18) | (2.65) | (3.28) | (2.43) | (3.03) | (3.42) | (1.91) | (3.91) | (0.86) |
| R-squared | 0.30 | 0.34 | 0.33 | 0.43 | 0.56 | 0.60 | 0.56 | 0.53 | 0.68 | 0.51 | 0.45 | 0.62 | 0.66 | 0.71 | 0.67 | 0.77 | 0.80 | 0.66 | 0.59 | 0.38 |
| Optimal-lambda | 0.4 | 0.5 | 0.9 | 0.1 | 0.9 | 0.7 | 0.9 | 0.1 | 0.6 | 0.6 | 0.4 | 0.4 | 1.0 | 1.0 | 0.1 | 1.0 | 0.1 | 0.1 | 0.1 | 1.0 |

Note: The bold print indicates significance for one tailed testt the 5 percent level (t=1.684) or two tailed testt the 10 percent level of significance.

Table 8. Labor Force Participation Equation (3) Estimates Using Cross-Section Data (1990) -- White Female

| Age Cohort | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | 5.80 | 6.67 | 3.10 | 3.01 | 3.29 | 4.97 | 3.62 | 3.26 | 4.63 | 3.72 | 3.74 | 3.70 | 3.56 | 3.13 | 2.30 | 1.16 | -1.91 | -4.16 | -4.78 | -6.48 |
| | (2.25) | (3.25) | (3.62) | (3.89) | (5.64) | (7.25) | (6.93) | (5.59) | (8.15) | (6.89) | (11.00) | (14.26) | (14.14) | (8.25) | (4.32) | (1.67) | (-1.65) | (-2.74) | (-2.59) | (-3.04) |
| FE | 2.91 | 1.95 | 0.84 | 0.49 | 0.67 | 0.72 | 0.74 | 0.66 | 0.87 | 0.71 | 0.68 | 0.56 | 0.50 | 0.53 | 0.66 | 0.80 | 0.91 | 1.28 | 1.24 | 0.55 |
| | (4.31) | (3.62) | (4.01) | (2.63) | (4.87) | (4.44) | (5.93) | (4.70) | (6.45) | (5.52) | (8.91) | (8.80) | (8.00) | (5.50) | (4.93) | (4.34) | (2.91) | (3.11) | (2.51) | (0.86) |
| NR | -0.69 | -0.83 | -0.27 | -0.26 | 0.01 | -0.07 | -0.14 | -0.16 | -0.19 | 0.05 | -0.04 | -0.12 | -0.20 | | | | | | | |
| | (-1.41) | (-2.12) | (-1.77) | (-1.95) | (0.26) | (-1.47) | (-3.96) | (-4.02) | (-4.93) | (0.83) | (-1.02) | (-1.82) | (-0.72) | | | | | | | |
| NOH 1824 | -0.60 | 1.05 | | | | | | | | | | | | | | | | | | |
| | (-0.33) | (0.73) | | | | | | | | | | | | | | | | | | |
| COL 2534 | | | -1.06 | -0.81 | -0.61 | -0.48 | -0.37 | -0.44 | -0.23 | | | | | | | | | | | |
| | | | (-5.66) | (-4.91) | (-4.73) | (-3.15) | (-3.24) | (-3.33) | (-1.81) | | | | | | | | | | | |
| NOH | | | -0.36 | 0.39 | -0.69 | -0.83 | -0.55 | -0.45 | -0.25 | -0.39 | -0.45 | -0.42 | -0.64 | -0.47 | -0.49 | -0.52 | -0.36 | -0.15 | 0.06 | -0.72 |
| | | | (-0.63) | (0.75) | (-2.63) | (-2.69) | (-2.32) | (-1.71) | (-0.99) | (-1.75) | (-3.32) | (-4.78) | (-7.48) | (-5.19) | (-3.83) | (-3.81) | (-1.72) | (-0.60) | (0.20) | (-1.71) |
| FCOL | | | | | | | | | | 0.18 | 0.11 | 0.14 | 0.05 | 0.04 | -0.00 | 0.04 | 0.24 | 1.30 | 1.65 | 1.62 |
| | | | | | | | | | | (1.42) | (1.82) | (2.48) | (0.74) | (0.53) | (-0.02) | (0.23) | (0.61) | (2.52) | (2.67) | (1.29) |
| RWR | -1.79 | -1.36 | -0.11 | -0.01 | -0.04 | -0.35 | 0.01 | 0.06 | -0.15 | -0.06 | -0.12 | -0.11 | -0.05 | -0.03 | 0.05 | 0.20 | 0.43 | 0.69 | 0.56 | 0.87 |
| | (-4.24) | (-4.07) | (-0.67) | (-0.09) | (-0.36) | (-2.75) | (0.07) | (0.57) | (-1.45) | (-0.64) | (-1.99) | (-2.36) | (-1.03) | (-0.51) | (0.52) | (1.55) | (2.00) | (2.46) | (1.66) | (2.21) |
| RURAL | -0.69 | -0.60 | -0.35 | -0.27 | -0.15 | -0.11 | -0.12 | -0.04 | -0.10 | 0.02 | 0.08 | 0.07 | 0.04 | -0.01 | -0.04 | 0.01 | 0.01 | -0.08 | 0.07 | 0.27 |
| | (-2.90) | (-3.15) | (-4.63) | (-4.16) | (-3.03) | (-1.93) | (-2.74) | (-0.87) | (-2.16) | (0.34) | (2.52) | (2.51) | (1.72) | (-0.22) | (-0.77) | (0.12) | (0.11) | (-0.54) | (0.39) | (1.28) |
| EA | 0.73 | 0.26 | 0.45 | 0.37 | 0.35 | 0.17 | 0.26 | 0.31 | 0.12 | 0.21 | 0.26 | 0.27 | 0.26 | 0.35 | 0.47 | 0.61 | 1.06 | 1.16 | 1.19 | 1.20 |
| | (1.82) | (0.82) | (3.49) | (3.16) | (3.90) | (1.64) | (3.24) | (3.53) | (1.41) | (2.61) | (4.75) | (6.43) | (6.55) | (5.89) | (5.69) | (5.63) | (5.91) | (5.08) | (4.27) | (3.60) |
| CWC | 1.55 | 1.58 | -0.36 | -0.30 | -0.49 | -0.28 | -0.26 | -0.22 | -0.11 | -0.66 | -0.58 | -0.54 | -0.46 | -0.45 | -0.52 | -0.60 | -0.68 | | | |
| | (1.89) | (2.42) | (-1.12) | (-1.06) | (-2.15) | (-1.01) | (-1.30) | (-0.95) | (-0.50) | (-3.05) | (-4.62) | (-5.85) | (-5.14) | (-3.49) | (-2.88) | (-2.43) | (-1.61) | | | |
| HISP | -0.47 | -0.45 | -0.12 | -0.05 | -0.01 | 0.03 | -0.01 | 0.00 | -0.08 | -0.23 | -0.05 | -0.08 | -0.02 | -0.09 | -0.15 | -0.27 | -0.24 | -0.22 | 0.14 | 0.91 |
| | (-0.85) | (-1.02) | (-0.67) | (-0.30) | (-0.12) | (0.19) | (-0.12) | (0.04) | (-0.74) | (-2.47) | (-0.82) | (-1.62) | (-0.35) | (-1.27) | (-1.45) | (-1.96) | (-1.01) | (-0.68) | (0.36) | (1.85) |
| R-squared | 0.69 | 0.61 | 0.78 | 0.70 | 0.76 | 0.70 | 0.83 | 0.77 | 0.81 | 0.83 | 0.90 | 0.93 | 0.93 | 0.90 | 0.87 | 0.85 | 0.78 | 0.76 | 0.66 | 0.62 |
| Optimal-lambda | 0.3 | 0.3 | 0.3 | 0.5 | 0.4 | 0.4 | 0.4 | 0.3 | 0.4 | 0.3 | 0.7 | 0.4 | 0.3 | 0.3 | 0.3 | 0.2 | 0.2 | 0.2 | 0.3 | 0.1 |

Note: The bold print indicates significance for one tailed testt the 5 percent level (t=1.684) or two tailed testt the 10 percent level of significance.

Table 9. Labor Force Participation Equation (3) Estimates Using Cross-Section Data (1990) -- Black Female

| Age Cohort | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | -0.96 | -1.64 | 6.01 | 6.47 | 5.57 | 5.87 | 3.41 | 3.87 | 3.95 | 3.21 | 2.90 | 2.22 | 2.67 | 1.42 | -0.44 | 0.03 | -3.29 | -7.13 | -2.75 | -13.01 |
| | (-0.20) | (-1.30) | (2.48) | (3.99) | (3.14) | (3.19) | (1.87) | (2.49) | (2.40) | (3.29) | (3.81) | (3.18) | (3.72) | (2.16) | (-0.53) | (0.02) | (-2.21) | (-1.82) | (-0.87) | (-2.52) |
| FE | 0.91 | 0.51 | 1.28 | 1.27 | 0.75 | 1.07 | 0.10 | 0.70 | 0.08 | 0.43 | 0.34 | -0.07 | -0.02 | -0.21 | -0.41 | -0.00 | -0.15 | -0.95 | -0.30 | 1.24 |
| | (0.68) | (0.53) | (1.96) | (2.87) | (1.81) | (2.43) | (0.25) | (1.93) | (0.21) | (1.28) | (1.33) | (-0.27) | (-0.07) | (-0.99) | (-1.46) | (-0.01) | (-0.29) | (-0.79) | (-0.30) | (0.71) |
| NR | 0.12 | 0.64 | -0.26 | -0.38 | -0.22 | -0.23 | -0.08 | -0.27 | -0.08 | -0.24 | -0.29 | -0.41 | -0.57 | | | | | | | |
| | (0.26) | (1.82) | (-1.04) | (-2.22) | (-1.55) | (-1.62) | (-0.57) | (-2.19) | (-0.59) | (-2.17) | (-2.67) | (-2.47) | (-1.17) | | | | | | | |
| NOH 18-24 | -1.69 | -2.23 | | | | | | | | | | | | | | | | | | |
| | (-1.03) | (-1.77) | | | | | | | | | | | | | | | | | | |
| COL 25-34 | | | -0.47 | -2.08 | -1.14 | -0.64 | -0.56 | -0.32 | 0.79 | | | | | | | | | | | |
| | | | (-0.46) | (-2.86) | (-1.51) | (-0.83) | (-0.73) | (-0.49) | (1.14) | | | | | | | | | | | |
| NOH | | | -0.06 | -0.36 | -0.67 | -0.57 | -0.90 | -0.40 | -0.81 | 0.42 | 0.21 | 0.11 | 0.00 | -0.29 | -0.36 | -0.01 | -0.15 | -0.22 | -0.59 | -0.29 |
| | | | (-0.07) | (-0.65) | (-1.34) | (-1.11) | (-1.76) | (-0.92) | (-1.75) | (1.50) | (0.99) | (0.68) | (0.02) | (-2.70) | (-2.77) | (-0.05) | (-0.67) | (-0.43) | (-1.31) | (-0.31) |
| FCOL | | | | | | | | | | 0.72 | 0.97 | -0.13 | 0.32 | -0.32 | -1.05 | 0.50 | 0.30 | -2.29 | -0.26 | -3.81 |
| | | | | | | | | | | (1.12) | (1.97) | (-0.32) | (0.81) | (-0.92) | (-2.25) | (0.99) | (0.36) | (-0.78) | (-0.11) | (-1.04) |
| RWR | -0.74 | 0.44 | -0.90 | -1.05 | -0.41 | -0.20 | 0.14 | 0.05 | 0.29 | 0.04 | 0.05 | 0.24 | 0.06 | 0.22 | 0.35 | 0.15 | 0.32 | 1.15 | 0.09 | 1.49 |
| | (-0.65) | (0.52) | (-1.59) | (-2.65) | (-1.09) | (-0.53) | (0.36) | (0.15) | (0.82) | (0.26) | (0.40) | (1.94) | (0.36) | (1.51) | (1.99) | (0.50) | (0.98) | (1.69) | (0.16) | (1.67) |
| RURAL | 0.18 | 0.55 | -0.34 | -0.61 | -0.45 | -0.36 | -0.19 | -0.01 | -0.08 | 0.02 | 0.02 | -0.06 | -0.08 | -0.08 | -0.04 | -0.18 | 0.26 | 0.10 | -0.02 | 0.50 |
| | (0.37) | (1.48) | (-1.28) | (-3.51) | (-2.78) | (-2.17) | (-1.15) | (-0.04) | (-0.55) | (0.21) | (0.25) | (-0.74) | (-0.92) | (-0.91) | (-0.35) | (-1.07) | (1.36) | (0.19) | (-0.04) | (0.83) |
| EA | 1.68 | 1.53 | 0.29 | 0.42 | 0.12 | -0.12 | 0.16 | 0.16 | -0.13 | 0.24 | 0.30 | 0.37 | 0.36 | 0.56 | 0.92 | 0.86 | 1.43 | 1.61 | 1.22 | 2.39 |
| | (2.76) | (3.37) | (0.85) | (1.90) | (0.50) | (-0.47) | (0.64) | (0.75) | (-0.58) | (1.22) | (1.98) | (2.76) | (3.16) | (5.21) | (6.43) | (3.77) | (5.59) | (2.47) | (2.43) | (2.88) |
| CWC | -3.52 | -0.36 | -1.75 | -1.30 | 0.25 | 0.67 | 1.11 | 0.77 | 1.57 | 0.17 | 0.11 | 0.00 | -0.08 | 0.16 | -0.19 | -0.48 | -0.96 | | | |
| | (-1.64) | (-0.23) | (-1.68) | (-1.72) | (0.42) | (1.04) | (1.80) | (1.46) | (2.73) | (0.55) | (0.44) | (0.01) | (-0.25) | (0.53) | (-0.48) | (-0.70) | (-1.16) | | | |
| HISP | 2.24 | 4.64 | -0.45 | -0.73 | -1.60 | -1.08 | -0.85 | -0.66 | -1.50 | -0.05 | 0.21 | 0.07 | -0.07 | 0.42 | 0.64 | 0.71 | 2.95 | 1.34 | 1.63 | 5.20 |
| | (0.79) | (2.18) | (-0.31) | (-0.74) | (-1.82) | (-1.19) | (-0.94) | (-0.86) | (-1.84) | (-0.10) | (0.48) | (0.17) | (-0.18) | (1.44) | (1.70) | (1.12) | (4.02) | (0.74) | (0.89) | (1.91) |
| R-squared | 0.51 | 0.59 | 0.38 | 0.72 | 0.53 | 0.48 | 0.50 | 0.59 | 0.62 | 0.32 | 0.48 | 0.43 | 0.54 | 0.72 | 0.74 | 0.65 | 0.73 | 0.28 | 0.39 | 0.49 |
| Optimal-lambda | 0.4 | 0.6 | 0.4 | 0.4 | 0.4 | 0.1 | 0.6 | 0.4 | 0.1 | 0.1 | 0.1 | 0.1 | 0.4 | 0.1 | 0.1 | 0.1 | 0.2 | 1.0 | 0.1 | 0.6 |

Note: The bold print indicates significance for one tailed testt the 5 percent level (t=1.684) or two tailed testt the 10 percent level of significance.

Table 10. Labor Force Participation Equation (3) Estimates Using Cross-Section Data (1990) -- Other Female

| Age Cohort | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | 5.00 | 5.36 | 6.31 | 4.99 | 3.68 | 3.25 | 1.57 | 3.21 | 3.62 | 1.64 | 1.70 | 2.28 | 2.31 | 3.05 | 1.81 | 2.40 | -0.44 | -3.68 | 0.61 | 5.98 |
| | (1.69) | (1.99) | (3.20) | (3.64) | (2.02) | (2.35) | (1.06) | (2.07) | (2.75) | (1.50) | (1.67) | (2.33) | (2.94) | (4.24) | (1.31) | (2.51) | (-0.24) | (-0.93) | (0.14) | (0.75) |
| FE | 3.37 | 4.06 | 1.53 | 1.42 | 1.59 | 0.80 | -0.98 | 0.01 | 0.19 | 0.15 | 0.13 | 0.24 | 0.26 | 0.54 | 0.33 | 0.80 | 1.20 | 0.97 | 2.04 | 0.64 |
| | (3.43) | (4.82) | (2.35) | (3.29) | (2.56) | (1.48) | (-1.60) | (0.02) | (0.35) | (0.39) | (0.39) | (0.76) | (1.02) | (2.03) | (0.70) | (2.21) | (2.00) | (0.75) | (1.43) | (0.26) |
| NR | -0.25 | -0.11 | -0.20 | -0.15 | -0.20 | -0.05 | 0.00 | -0.09 | -0.01 | 0.01 | -0.13 | -0.15 | -0.11 | | | | | | | |
| | (-1.54) | (-0.73) | (-1.30) | (-1.40) | (-2.50) | (-0.70) | (0.03) | (-1.27) | (-0.17) | (0.24) | (-1.83) | (-1.33) | (-0.51) | | | | | | | |
| NOH 1824 | -0.56 | -1.76 | | | | | | | | | | | | | | | | | | |
| | (-0.89) | (-3.06) | | | | | | | | | | | | | | | | | | |
| COL 2534 | | | -0.79 | -1.33 | -0.99 | -0.59 | -0.14 | -0.24 | 0.00 | | | | | | | | | | | |
| | | | (-1.69) | (-4.35) | (-2.06) | (-1.41) | (-0.30) | (-0.55) | (0.01) | | | | | | | | | | | |
| NOH | | | -0.88 | -1.52 | -0.42 | -1.08 | 0.31 | 0.00 | -0.51 | -0.33 | -0.17 | -0.06 | -0.34 | -0.72 | -0.76 | -1.01 | -0.46 | -1.72 | -1.10 | -0.21 |
| | | | (-1.89) | (-5.01) | (-0.75) | (-2.20) | (0.59) | (0.01) | (-1.37) | (-1.14) | (-0.60) | (-0.24) | (-1.69) | (-4.06) | (-2.40) | (-5.54) | (-1.67) | (-2.14) | (-1.27) | (-0.13) |
| FCOL | | | | | | | | | | 0.04 | -0.06 | 0.09 | 0.16 | 0.01 | -0.01 | 0.37 | 1.51 | -0.43 | 1.25 | 5.43 |
| | | | | | | | | | | (0.23) | (-0.34) | (0.51) | (1.16) | (0.06) | (-0.03) | (1.13) | (2.37) | (-0.20) | (0.50) | (1.21) |
| RWR | -0.90 | -0.99 | -0.72 | -0.15 | -0.42 | 0.03 | 0.05 | -0.45 | -0.12 | 0.10 | 0.13 | -0.08 | 0.07 | -0.04 | -0.01 | 0.08 | 0.34 | 0.56 | -0.55 | 0.35 |
| | (-1.62) | (-1.96) | (-1.88) | (-0.57) | (-1.26) | (0.12) | (0.17) | (-1.59) | (-0.44) | (0.45) | (0.66) | (-0.41) | (0.49) | (-0.26) | (-0.05) | (0.40) | (1.02) | (0.68) | (-0.62) | (0.23) |
| RURAL | -0.59 | -0.77 | -0.75 | -0.77 | -0.65 | -0.51 | -0.15 | -0.39 | -0.40 | -0.20 | -0.25 | -0.19 | -0.17 | -0.19 | -0.27 | -0.35 | -0.42 | 0.13 | -0.32 | -0.60 |
| | (-1.39) | (-2.13) | (-2.80) | (-4.26) | (-2.43) | (-2.18) | (-0.57) | (-1.59) | (-1.79) | (-1.42) | (-1.94) | (-1.57) | (-1.61) | (-1.71) | (-1.42) | (-2.31) | (-1.70) | (0.21) | (-0.47) | (-0.49) |
| EA | 0.48 | 0.52 | 0.10 | 0.07 | 0.55 | 0.27 | 0.46 | 0.53 | 0.19 | 0.53 | 0.57 | 0.53 | 0.44 | 0.40 | 0.60 | 0.48 | 0.81 | 1.35 | 0.56 | -1.32 |
| | (1.09) | (1.28) | (0.35) | (0.37) | (1.94) | (1.24) | (1.93) | (2.22) | (0.92) | (2.97) | (3.50) | (3.42) | (3.55) | (3.26) | (2.67) | (2.88) | (2.75) | (2.20) | (1.27) | (-1.06) |
| CWC | -2.00 | -0.91 | 0.00 | 0.66 | 0.21 | 0.66 | 1.15 | 0.55 | 0.91 | 0.16 | -0.24 | 0.07 | -0.18 | -0.26 | 0.23 | -0.67 | -1.02 | | | |
| | (-2.57) | (-1.27) | (0.01) | (1.82) | (0.46) | (1.64) | (2.66) | (1.42) | (2.39) | (0.55) | (-0.92) | (0.26) | (-0.85) | (-1.31) | (0.67) | (-2.55) | (-2.02) | | | |
| HISP | -0.05 | 0.01 | -0.33 | -0.32 | -0.34 | -0.09 | -0.04 | -0.06 | -0.11 | 0.08 | 0.01 | 0.00 | 0.07 | 0.14 | 0.13 | 0.21 | -0.09 | 0.26 | 0.25 | 0.25 |
| | (-0.16) | (0.04) | (-1.58) | (-2.18) | (-1.75) | (-0.51) | (-0.20) | (-0.33) | (-0.68) | (0.70) | (0.05) | (0.03) | (0.81) | (1.57) | (0.86) | (1.88) | (-0.57) | (0.75) | (0.66) | (0.36) |
| R-squared | 0.49 | 0.61 | 0.47 | 0.72 | 0.52 | 0.40 | 0.46 | 0.55 | 0.49 | 0.41 | 0.48 | 0.45 | 0.56 | 0.65 | 0.51 | 0.78 | 0.68 | 0.45 | 0.39 | 0.20 |
| Optimal-lamluda | 0.6 | 0.8 | 1.0 | 1.0 | 0.8 | 0.1 | 0.1 | 0.9 | 0.1 | 0.6 | 1.0 | 1.0 | 0.9 | 0.1 | 0.6 | 0.1 | 0.7 | 0.6 | 0.4 | 1.0 |

Note: The bold print indicates significance for one tailed testt the 5 percent level (t=1.684) or two tailed testt the 10 percent level of significance.

Table 11. Labor Force Participation Equation (without State-Specific Effects) Estimates Using Cross-Section Data -- White Male

| Age Cohort | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | 0.95 | 1.94 | 1.68 | 1.54 | 2.28 | 2.57 | 3.12 | 3.24 | 3.37 | 3.88 | 4.07 | 3.96 | 3.98 | 3.53 | 3.12 | 1.47 | -2.43 | -6.19 | -7.43 | -8.71 |
| | (0.43) | (1.27) | (1.76) | (2.23) | (4.09) | (4.77) | (7.57) | (9.75) | (11.02) | (18.51) | (23.99) | (23.54) | (21.79) | (16.49) | (10.26) | (2.63) | (-2.43) | (-3.76) | (-4.35) | (-4.24) |
| NOH 1824 | -1.42 | -0.39 | | | | | | | | | | | | | | | | | | |
| | (-1.98) | (-0.78) | | | | | | | | | | | | | | | | | | |
| COL 2534 | | | -1.26 | -0.91 | -0.69 | -0.36 | -0.26 | -0.31 | -0.30 | | | | | | | | | | | |
| | | | (-5.39) | (-5.54) | (-5.33) | (-2.88) | (-2.66) | (-3.78) | (-4.07) | | | | | | | | | | | |
| NOH | | | 0.01 | 0.68 | 0.62 | 0.52 | 0.38 | 0.03 | 0.03 | -0.02 | -0.07 | -0.07 | -0.13 | -0.15 | -0.27 | -0.14 | 0.00 | 0.39 | 0.34 | 0.29 |
| | | | (0.03) | (2.91) | (3.29) | (2.87) | (2.70) | (0.27) | (0.27) | (-0.29) | (-1.37) | (-1.31) | (-2.33) | (-2.76) | (-3.42) | (-1.15) | (0.00) | (1.21) | (1.03) | (0.72) |
| FCOL | | | | | | | | | | -0.07 | 0.01 | 0.01 | -0.00 | -0.08 | -0.14 | -0.27 | -0.47 | 0.47 | 0.26 | 0.74 |
| | | | | | | | | | | (-1.66) | (0.25) | (0.33) | (-0.11) | (-1.82) | (-2.37) | (-2.34) | (-1.85) | (1.15) | (0.63) | (1.25) |
| RWR | -0.68 | -0.58 | -0.02 | 0.16 | 0.15 | 0.12 | 0.08 | 0.07 | 0.09 | 0.07 | 0.05 | 0.08 | 0.07 | 0.13 | 0.17 | 0.30 | 0.71 | 0.75 | 0.77 | 1.02 |
| | (-1.79) | (-2.21) | (-0.10) | (1.17) | (1.41) | (1.10) | (0.97) | (1.06) | (1.46) | (1.82) | (1.56) | (2.43) | (1.92) | (3.36) | (3.05) | (2.96) | (3.82) | (2.51) | (2.48) | (2.73) |
| RURAL | -0.21 | -0.23 | -0.37 | -0.31 | -0.24 | -0.10 | -0.13 | -0.03 | -0.02 | 0.01 | 0.01 | 0.01 | -0.02 | -0.03 | -0.01 | -0.05 | -0.05 | 0.03 | 0.17 | 0.20 |
| | (-0.95) | (-1.54) | (-4.16) | (-5.01) | (-4.84) | (-2.19) | (-3.61) | (-0.85) | (-0.87) | (0.53) | (0.34) | (0.53) | (-0.97) | (-1.28) | (-0.44) | (-0.90) | (-0.49) | (0.15) | (0.97) | (0.99) |
| EA | 1.13 | 0.85 | 0.68 | 0.58 | 0.41 | 0.35 | 0.27 | 0.27 | 0.22 | 0.11 | 0.08 | 0.09 | 0.09 | 0.16 | 0.23 | 0.52 | 1.08 | 1.65 | 1.81 | 1.75 |
| | (3.25) | (3.54) | (4.71) | (5.49) | (4.73) | (4.21) | (4.30) | (5.36) | (4.85) | (3.45) | (2.90) | (3.14) | (3.00) | (4.60) | (4.74) | (5.93) | (6.97) | (6.53) | (6.91) | (5.59) |
| CWC | 0.56 | 0.54 | -0.01 | -0.37 | -0.22 | -0.21 | -0.22 | -0.15 | -0.06 | -0.06 | 0.05 | 0.01 | 0.05 | -0.00 | -0.02 | -0.21 | -0.47 | 0.43 | 0.48 | 0.85 |
| | (0.74) | (1.04) | (-0.03) | (-1.44) | (-1.04) | (-1.02) | (-1.39) | (-1.21) | (-0.54) | (-0.78) | (0.72) | (0.12) | (0.81) | (-0.03) | (-0.22) | (-1.07) | (-1.32) | (1.18) | (1.28) | (1.92) |
| HISP | -0.44 | -0.47 | -0.29 | -0.13 | -0.13 | -0.09 | -0.21 | -0.07 | -0.07 | -0.08 | -0.04 | -0.04 | -0.04 | -0.02 | -0.02 | -0.06 | 0.18 | | | |
| | (-0.89) | (-1.36) | (-1.43) | (-0.94) | (-1.17) | (-0.85) | (-2.46) | (-0.93) | (-1.07) | (-1.90) | (-1.27) | (-1.28) | (-1.01) | (-0.38) | (-0.30) | (-0.51) | (0.82) | | | |
| R-squared | 0.51 | 0.51 | 0.75 | 0.75 | 0.74 | 0.52 | 0.54 | 0.58 | 0.60 | 0.38 | 0.46 | 0.52 | 0.56 | 0.66 | 0.64 | 0.61 | 0.64 | 0.61 | 0.59 | 0.56 |
| Optimal-lambda | 0.4 | 0.4 | 0.3 | 0.6 | 1.0 | 1.0 | 0.4 | 0.3 | 0.3 | 0.3 | 0.3 | 0.4 | 0.3 | 0.4 | 0.3 | 0.4 | 0.4 | 0.7 | 0.7 | 1.0 |

Note: The bold print indicates significance for one tailed testt the 5 percent level (t=1.684) or two talled testt the 10 percent level of significance.

# CHAPTER XII

# DEMOGRAPHIC/MIGRATION MODULE

## A. INTRODUCTION

The REMI Demographic/Migration module divides regional population changes into two distinct but interactive processes. First, a region's population may change as the net result of natural causes such as births and deaths within the population. Second, a region may gain or lose net population because of inter-regional movement of the U.S. population (herein referred to as "net migrants"). Moreover, the migrants become part of the receiving region's population to which the natural population processes apply.

**The first section of the chapter will describe the calculation of the natural changes in population. The components of net migrants and how they are predicted comprise the second section. A final section is added to explain the appended population tables.**

NOTE: The tables have been omitted from this Supplemental Reference Materials Guide. For information omitted, please refer to the Tables in your Policy Insight software.

## B. NATURAL POPULATION CHANGES

Natural population changes are predicted for one-year age cohorts for both males and females by race (white, black, and other). To obtain the natural population changes, a cohort algorithm is employed. The cohort algorithm consists of applying fertility and survival rates to the appropriate cohorts and adding the births while subtracting the deaths.

The cohort algorithm begins in 1970 with the 1970 Bureau of Census Survey providing the benchmark data for each region. Starting with 1970 and continuing through history, the census population data are normalized to that from the Bureau of Economic Analysis for consistency with the economic model. Furthermore, beginning in 1970, military personnel and their dependents are subtracted from the population and predicted separately. The cohort algorithm is then applied to the remaining cohorts which we now refer to as the "civilian population."

Natality rates for five-year age cohorts by race for each state are drawn from the 1990 Census data. Single-year rates are calculated based on variations of 1990 U.S. single-year natality rates within corresponding five-year cohort rates as reported by Vital Statistics. The single-year rates are calculated forward and backward from 1990 based on single-year historical U.S. trends in Vital Statistics and projected U.S. trends from Current Population Reports P25-1130 (Census). County-based models use their state-specific natality rates.

Survival rates for single-year age cohorts by race for each state are obtained from 1980 Vital Statistics data. For most states, this means using the survival rates of the state for whites and blacks and the U.S. rates for others. Further, we normalize the rates to the known number of deaths in the area by race in 1980. These single-year rates are calculated forward and backward from 1980 based on the projected trends in the U.S. Census reports. Once again, county-based models use the state-specific survival rates.

Starting with 1971, predicted civilian births and deaths by cohort and race are adjusted to obtain

total birth and death rates by race calculated from Vital Statistics data. The adjustment to births and deaths in the last year of history is carried forward into the forecast period. In addition to predicting the civilian population, the number of military personnel and their dependents are also calculated. BEA Local Area Personal Income data provides regional military employment estimates. The number of military personnel is adjusted to account for the fact that the BEA concept for military employment includes the reserves and national guard. Consequently, we alter the military concept by multiplying by an active to total state ratio to truly capture active military personnel. The proportion of military personnel that is male and female, and the number of male and female dependents, were obtained from the Census. The age distribution by race of the military personnel and dependents was estimated using single-year and five-year cohort data also from the Census, based on 1995.

## C. MIGRANTS

The components of net migrants are: international migrants; retired migrants; special populations including college students and prisoners; change in military personnel and their dependents; and economic migrants. The motivation for each group's migration decision and method of calculation are presented in the following paragraphs.

*International migration* is assumed exogenous to the region and is based on each region's share of U.S. international immigrants by race. Net international immigration, by race, for the U.S. is obtained from the Census Current Population Reports for recent history (beginning in 1981). The Statistical Abstract provides an early history (1971-1980). Racial detail is obtained by assigning a racial category to each country of origin available in Immigration and Naturalization Service (INS) data, and then reconciling to the total from the Statistical Abstract data. The INS provides information on immigrants admitted by selected country of birth and state of intended residence. We used this series for the years 1971, 1982, 1983, 1986, 1987, 1988, 1990, and 1994 and assigned a racial category to each country of origin. We

*migrants* are defined as persons under 65 years of age who move in response to differential changes in inter-regional economic opportunities. The exact equation and explanation is given in Chapter 3, Section 1. The two variables for the equation that the population model produces are the dependent variable, net economic migrants, and the natural labor force for the independent variable REO.

*Net economic migrants* are calculated as a residual between observed civilian population under 65 and the predicted population by the cohort algorithm and other components of migration. An additional adjustment is required though, due to more complete coverage of the population by the 1980 census relative to the 1970 census. The Bureau of Census distributes the 1970 undercount over the intercensal years with the largest intercensal adjustment occurring in 1971 and gradually declining to zero in 1980.

Total undercount adjustments over all years by state are obtained directly from the Bureau of Census. The economic migrant series is adjusted by first breaking the undercount for each state into those over and those under 65 according to "Coverage of the National Population in the 1980 Census by Age, Gender, and Race: Preliminary Estimates by Demographic Analysis." The proportion of the under-65 undercount by state is allocated across the years by taking the proportion of predicted positive net economic migrants in each year summed across all states to the total over all intercensal years. The total residually predicted undercount is within ten percent of the middle census projected undercount for the U.S. The ten percent discrepancy is distributed across states in proportion to their population. For regions within states, undercount and residual adjustments are based on the region's share of the corresponding population. Therefore, the resultant sum of net economic migrants, summed across all states, equals zero.

A region's *labor force* is calculated as the sum of individual age/gender/race cohorts multiplied by the corresponding labor force participation rates. The participation rates are calculated by age, gender, and race endogenously. The participation rate equations were derived econometrically and are largely based on wage rates, employment opportunity, demographic characteristics, state differences, and the U.S. participation rate. The paper, "Regional Labor Force Participation Rates", which describes this work in

detail, is available from REMI. The participation rates are "actual" rates which try to reflect the economic conditions of an area. For example, in general when economic conditions improve, people will shift back to the labor force to capture employment. The participation rates follow the trends in history from the U.S. Bureau of Labor Statistics (BLS) and the projected U.S. trends also from BLS. As a final note, the labor force cohorts are adjusted to reflect the undercount adjustments applied to the residual economic migrant series.

With the above two variables produced by the population model, equation 2-19 in Chapter 2 produces predicted net economic migrants. The migrants are distributed according to the state-specific migrant distributions given by 1975-1980 migration reports by the Census. Moreover, the cohort algorithm is also applied to economic migrants in the current year, consistent with that for the other migration components.

The final step is to multiplicatively adjust the civilian population. For 1970-1992, total predicted single-year cohort levels are multiplicatively adjusted to obtain corresponding, normalized five-year cohort levels by race reported in "Intercensal Estimates of the Population of Counties by Age, Gender and Race: 1970-1980," and similar data based upon the 1990 Census, and their most current five-year cohort estimates by county. For all periods of history, the cohorts are multiplicatively adjusted to obtain the normalized total population from BEA to be consistent with the economic model

# DATA SOURCES AND ESTIMATION/CALIBRATION PROCEDURES

## A. PRIMARY HISTORICAL DATA

The basic national, state and county data source for the REMI EDFS Model is the Bureau of Economic Analysis (BEA) employment, wage, and personal income series. It is an internally consistent data set covering the years from 1969 to the present (usually updated in the fall/winter for states, spring/summer for counties). The BEA data is available for states at the two-digit level (53+ industries), and available for counties at the one-digit level (14+ industries).

The second major source of data available is the Bureau of Labor Statistics (BLS) ES-202 annual average employment and total annual wages. This data is collected on a monthly basis in conjunction with the unemployment insurance program, and forms the basic data upon which the wage and salary employment components of the BEA total employment data is based. It is available at the two-digit Standard Industrial Classifications (SIC) level for all counties and states. Since this data series is based on unemployment insurance reports, it does not include private households, membership organizations, railroads, or the military.

The final source of employment and wage data available is County Business Patterns (CBP) data, which is collected in conjunction with the Social Security program in March of each year. This data is available at the four- and five-digit levels, and while it has many suppressions due to confidentiality requirements, its advantage is that when the data is suppressed, ranges for the firms are supplied. For example, for a given suppressed industry, there may be two establishments in an area with between 50,000-99,999 employees, and one establishment with 100,000 or more. This provides some basis from which to make a rough estimate of employees in that industry in the absence of any other information.

### i. BEA DATA

REMI receives four data series from the Bureau of Economic Analysis (BEA): Total Employment, Wage and Salary Employment, Wage and Salary Disbursements, and Personal Income. Total Employment is considered any activity that is compensated in the monetary economy, including people who work for wages and salaries or are self-employed. Wage and Salary Employment is the same as Total Employment but does not contain those people who are self employed (proprietors). Wage and Salary Disbursements are defined as the monetary remuneration of employees and includes the compensation of corporate officers in the form of commissions, tips, or bonuses as well as the receipts-in-kind that represent income to the recipient.

The Personal Income data series is comprised of many types of income. It includes the following: wage and salary disbursements, other labor income, proprietors' income, rental income, personal dividend income, personal interest income, transfer payments, and personal contributions for social insurance. Other labor income consists of employer contributions to private pensions, private welfare funds and private workers' compensations funds. The proprietors' income is that income or income-in-kind which is received by sole proprietorships, partnerships, or tax-exempt cooperatives. Rental income is made up of the income received from the rental of real property, and the royalties received from patents, copyrights, and rights to natural resources. Also included in rental income is an imputed net rental income of owner-occupants of non-farm dwellings. Dividends are payments by for-profit corporations to non-corporate stockholders residing in the United States in the form of cash or other assets. The interest component of Personal Income is determined in part by the income received from such items as savings accounts. The BEA also calculates an "imputed" interest which reflects the value of financial services. Transfer payments are income payments to persons for which a service is not provided. Personal contributions for social insurance are the contributions made by individuals under the various federal social insurance programs. They are calculated into personal income as a deduction.

The components of personal income are measured in two manners: by place-of-residence or by place-of-work. Place-of-residence refers to the location where the income is received. The components of personal income which are calculated in this manner include transfer payments, personal dividend income, personal interest income, rent income and proprietors' income. The remaining components of personal income; wage and salary disbursements, other labor income and personal contributions to social insurance are measured by place-of-work.

In order to calculate the flow of dollars within a multi-region model, REMI has in the past used the net Residence Adjustment figure from the Bureau of Economic Analysis (BEA). The net Residence Adjustment value is the product of subtracting the inward flow of dollars (income of those people who live in one region and work in another) from the outward flow of dollars (those people who work in the region and live in another). This methodology could not provide specific information such as how many dollars moved from region to region and to which region the dollars moved.

REMI has been able to improve its multi-regional models with the addition of Journey to Work (JTW) data from the Regional Economic Measurement Division. This data shows the movement of dollars from a specific county to all other counties which take monies away from it as well as the amount of dollars brought into the county from all other counties. This data set is internally consistent in that dollars moving from county A to county B equals those dollars moving to B from A. Using a simple RAS procedure to normalize the data so that the new movement of dollars equal those that the BEA indicates in its net Residence Adjustment value, REMI is able to obtain a matrix of intraregional flows. With this matrix, REMI can better allocate the flow of intraregional dollars.

Each of the data series contains two natural classifications; regional and industrial. Regional information is provided for the United States, for each individual state, and for the eight Major Regions. Industries are identified using SIC (Standard Industrial Classification) codes which are divided at the one-digit, two-digit and three-digit levels. For the regions, BEA begins by adding up county data to get state level observations for each data series. State observations are then summed to obtain a Major Region data series. Combining the Major Regions gives a complete set of observations for the United States.

**Table 1:** Major Regions and their state components.

| | |
|---|---|
| New England | Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont |
| Mideast | Delaware, District of Columbia, Maryland, New Jersey, New York, and Pennsylvania |
| Great Lakes | Illinois, Indiana, Michigan, Ohio and Wisconsin |
| Plains | Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota |
| Southeast | Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia |
| Southwest | Arizona, New Mexico, Oklahoma, and Texas |
| Rocky Mountains | Colorado, Idaho, Montana, Utah, and Wyoming |
| Far West | Alaska, California, Hawaii, Nevada, Oregon, and Washington |

**Table 2:** Industry Detail

- Farm
- Nonfarm
  - Private
    - Agricultural Services, Forestry, Fisheries, and Other
      - Agricultural Services
      - Forestry, Fisheries, and Other
        - Forestry
        - Fisheries
        - Other
    - Mining
      - Coal Mining
      - Oil and Gas Extraction
      - Metal Mining
      - Nonmetallic Minerals, Except Fuels
    - Construction
      - General Building Contractors
      - Heavy Construction Contractors
      - Special Trade Contractors
    - Manufacturing
      - Nondurable goods
        - Food and Kindred Products
        - Textile Mill Products
        - Apparel and Other Textile Products
        - Paper and Allied Products
        - Printing and Publishing
        - Chemicals and Allied Products
        - Petroleum and Coal Products
        - Tobacco Products
        - Rubber and Miscellaneous Plastics Products
        - Leather and Leather Products
      - Durable Goods
        - Lumber and Wood Products
        - Furniture and Fixtures
        - Primary Metal Industries
        - Fabricated Metal Products
        - Machinery and Computer Equipment
        - Electronic Equipment, Except Computer Equipment
        - Transportation Equipment Excl. Motor Vehicles
        - Motor Vehicles and Equipment
        - Ordnance
        - Stone, Clay, and Glass Products
        - Instruments and Related Products
        - Miscellaneous Manufacturing Industries
    - Transportation and Public Utilities
      - Railroad Transportation
      - Trucking and Warehousing
      - Water Transportation
      - Other Transportation
        - Local and Interurban Passenger Transit
        - Transportation by Air

- Pipelines, Except Natural Gas
- Transportation Services
- Communications
- Electric, Gas, and Sanitary Services

Wholesale Trade

Retail Trade
- Building Materials and Garden Equipment
- General Merchandise Stores
- Food Stores
- Automotive Dealers and Service Stations
- Apparel and Accessory Stores
- Home Furniture and Furnishings Stores
- Eating and Drinking Places
- Miscellaneous Retail

Finance, Insurance, and Real Estate
- Depository and Nondepository Credit Institutions
- Other Finance, Insurance, and Real Estate
  - Security & Commodity Brokers and Services
  - Insurance Carriers
  - Insurance Agents, Brokers, and Services
  - Real Estate
  - Combined Real Estate, Insurance, Etc.
  - Holding and Other Investment Companies

Services
- Hotels and Other Lodging Places
- Personal Services
- Private Households
- Business Services
- Auto Repair, Services, and Parking
- Miscellaneous Repair Services
- Amusement and Recreation Services
- Motion Pictures
- Health Services
- Legal Services
- Educational Services
- Social Services
- Museums, Botanical, Zoological Gardens
- Membership Organizations
- Engineering & Management Services
- Miscellaneous Services

Government and Government Enterprises
- Federal, Civilian
- Military
- State and Local

In some instances throughout the data series, the BEA has suppressed a value for a particular observation. This could be due to the fact that a small estimate was constructed or that the BEA wishes to protect the confidentiality of a firm within that industry. Suppressions are made in the data after the aggregation process is completed at the U.S. level. This allows for an internally consistent data set even though certain data have been withheld.

### a. Estimation of Data Suppressions in Major Regions and States

The current solving methodology is to minimize a constrained quadratic loss function. In order to begin this process, we obtain estimates and variances from regressions which will be used in our loss function. The loss function looks like:

$$L(\ ) = \sum_{i=1}^{n} \frac{(x_i^f - x_i^e)^2}{\sigma_i} \qquad \text{Equation (1)}$$

where $x_i^f$ is the actual value of the suppression and $x_i^e$ is the estimated value. This function would be minimized when $x_i^f = x_i^e$ for all i. It is good to have smaller variances as this means you are closer to the actual value. This function can also be expressed as a matrix function in the following manner:

$$\min (x^f - x^e)' \, Q^{-1} \, (x^f - x^e) \qquad \text{Equation (2)}$$

This hierarchal scheme of regions and industries is where REMI obtains the necessary constraints. Within the system provided by the BEA, there exists three regional and industrial constraints. The first one is inherent in the SIC classification. Within each one-digit industry, there exist two-digit industries, and within that two-digit, industries are broken down further into three-digit. The remaining constraints exist through the regional configuration of the data series. First, states are grouped into Major Regions. Hence, the sum of a particular industry over all the states in a Major Region will give the value for the Major Region. Secondly, summing all the Major Regions will equal the United States value for that industry. Let A, B, and D equal our three types of constraints and let C equal the constraint value.

$$\Omega = \begin{bmatrix} A \\ \cdots \\ B \\ \cdots \\ D \end{bmatrix} \qquad \text{Equation (3)}$$

$$C = \begin{bmatrix} C_A \\ \cdots \\ C_B \\ \cdots \\ C_D \end{bmatrix} \qquad \text{Equation (4)}$$

Finally, combine the loss function described above with the known constraint structure within each of the data series and obtain the following normalized Lagrangian equation:

$$L = (x^f - x^\varepsilon)' Q^{-1} (x^f - x^\varepsilon) + \lambda [\Omega x^f - C] \qquad \text{Equation (5)}$$

Calculate the first derivatives.

The system begins with Total Employment. Initial estimates and variances are calculated using regressions, share and shift measures, and interpolation. Data is used for the regressions in the following order: Wage and Salary Employment, Wage and Salary Disbursements, and Gross State Product data. The share and shift measures are used when a regression is not possible. A ratio of the region to the United States is calculated and then is applied to the available U.S. values (as suppressions do not exist at the U.S. level). If there exists a string of suppressions either in the beginning or ending years of an industry, interpolation is used. The rate of change between the last two data points is calculated and then used to interpolate the missing points.

Once there exists estimates and variances for all data points, each year is processed separately. Within each year, it is possible to break up the suppressions into their respective states. If there exists a Major Region suppression, all states associated with that region must be processed at the same time in order to maintain the regional system. If there exists two Major Regions with suppressions in the same group of industries, it is necessary to process all states that make up both of the Major Regions. If there is an overwhelming amount of suppressions as a result of this, it is possible to break out some the industries. Each one digit industry is considered its own system. Therefore, it is possible to solve each system of industries independently within a region.

Suppressions in the Personal Income series are solved next in the REMI methodology. The procedure is basically the same as Total Employment. The regressions are run using only the final Total Employment data set (all suppressions filled in). If a share and shift measure is used, the ratio is first calculated using the Major Region. However, if a suppression exists at the Major Region level, then the ratio is determined using the United States values. Interpolation is not used in calculating estimates for Personal Income. The data series also differ slightly in their final stage. The goal during Total Employment is to obtain all positive values during the final optimization. However, in the Personal Income data series, it is possible to have negative values as losses are calculated as negative income. therefore, the protocol is to accept a negative income value if the initial estimated value is a negative.

Wage and Salary Employment suppressions are solved next. The regression for this data series are run also using the final Total Employment data set. The share and shift measure follows the same procedure as the Personal Income. An additional consideration is added in the processing of this data. If there exists a zero in the Total Employment data set in a particular region, in a particular industry, and there exists a suppression in the same locale in the Wage and Salary Employment data set, then the fitted (or final) value for employment must be zero. Redundancies (systems that contain the fitted value) are solved and the system moves along as with Total Employment.

Finally, Wage and Salary Disbursements suppressions are solved in generally the same manner. The final Wage and Salary Employment data series (all suppressions filled in) is used in the regressions for this data series. The share and shift measures follow the same procedure as Personal Income. In this series, zeros are also placed in the Disbursements data set when there exists a zero in the Employment data set. Redundancies are solved and the process continues.

Once all the suppressions have been estimated, the constraints are checked and resumed to the United States levels.

**b. SIC Code Changes**

The BEA one- and two-digit industries are defined within the 1987 Standard Industrial Classifications (SIC) code system for 1988-present. Because REMI also had 1988 data in 1972 SIC codes, we were able to calculate state-specific employment ratios of the 1987 SIC to 1972 SIC for 1988. These ratios were applied to the BEA 1969-1987 data series for both employment and income to convert the entire series to 1987 SIC codes.

In the BEA data, the industry Ordnance (under Durable Goods) was defined as a unique two-digit industry within the 1967 SIC codes for 1969-1974. After 1974 when the 1972 SIC codes are used, Ordnance is dropped and reallocated. The employment in Ordnance is allocated back to fabricated metal and rest of transportation equipment for 1969-1974 based on an algorithm that uses: (a) the U.S. fabricated metal and transportation equipment shares of Ordnance from the BEA Table (*Survey of Current Business*, April 1974), and (b) the 1974-1975 employment data for the three industries. Wages are allocated by applying the Ordnance wage rate to the additional employment in fabricated metal and transportation equipment.

Social Services (under Services) was new under the 1972 standard industrial classification, therefore estimates prior to 1975 do not exist. REMI subtracts 23% from education and distributes it to non-profit organizations for the 1969-1974 social services series.

**c. Industry Combinations**

Some of the detailed BEA industries are combined into the REMI EDFS sectors:

| EDFS | BEA |
|---|---|
| Other transportation and transportation services | Water transportation;<br>Pipelines, except natural gas;<br>Transportation services |
| Brokers, credit, and other investment | Security and commodity brokers and services;<br>Holding and other investment companies |
| Insurance | Insurance carriers;<br>Insurance agents, brokers, and services |
| Real Estate | Real estate;<br>Combined real estate, ins., etc. |
| Rest of Retail | Building materials and garden equipment;<br>General merchandise stores;<br>Food stores;<br>Automotive dealers and service stations;<br>Apparel and accessories stores;<br>Home furniture and furnishings stores;<br>Miscellaneous retail stores |
| Personal & Repair Services | Personal services;<br>Miscellaneous repair services |

| | |
|---|---|
| Legal & Misc. Services | Legal services;<br>Engineering and management services;<br>Miscellaneous services |
| Non-Profit Member Organizations | Social services;<br>Membership organizations;<br>Museums, botanical, zoological gardens |

### ii. BLS ES-202 Data

The BLS ES-202 data begins in 1975. The first step in processing this data is to convert data for the years 1975-1987 (available only in 1977 SIC codes) to 1987 SIC code form. There is no overlap year of data available to us as there was for BEA data. The only available coefficients showing the relationship between 1987 SIC codes and 1977 SIC codes are at the national level. These coefficients were used to make preliminary adjustments to the 1975-1987 data. From 1988 forward data is received in 1987 SIC codes and does not require any conversion.

Next the two-digit employment data are aggregated to the 49 private non-farm industries in the REMI EDFS model. This is done at both the state and county levels. With the employment data, the comparable wage and salary disbursements are used to calculate the 49 private non-farm sector wage rates for each county. The next step in processing this data is to take the ratio of the BEA state employment data in relation to the ES-202 state employment data for all of these industries, and then to multiply this ratio by the ES-202 county employment data. Through this process an adjustment is made from a wage and salary employment base to a total employment base (includes proprietors) which is used by the BEA. Underlying this process is the assumption that the ratio of total employment, that is, employment including the self-employed, to the wage and salary employment as reported by ES-202, is the same in the state as it would be in each of the counties.

The ES-202 adjusted data will be used in combination with similarly adjusted CBP data to make initial estimates of employment for the county database that will actually be used in the EDFS-53 model. The ES-202 data series is not used in county EDFS-14 models, because 2-digit data is not needed. Likewise, ES-202 data is not needed for state models as the 2-digit data is available from BEA. Therefore, only EDFS-53 county models contain the ES-202 data.

### iii. CBP DATA

A single (most recent) year of CBP data that is defined within the 1987 SIC codes is used in the estimation/calibration system. This data is suppressed where the confidentiality of a firm would be violated, but these suppressions can be estimated using information as to the number of firms in employment size classes by industry.

The data is organized within the 1987 SIC code system as: (1) the state total; (2) within the total the one-digit industries; (3) within each one-digit, the two-digit industries; (4) within each two-digit, the three-digit industries; and (5) within each three-digit, the four-digit industries. Each one-digit industry has an administrative/auxiliary sector that is treated as a two-digit industry. Further, the employment and wage values for any level sum up to their higher-level value. For example, the one-digit SIC 50 (wholesale trade) equals the sum of two-digit SIC 5000 (wholesale trade & durable goods) and SIC 5100 (wholesale trade & non-durable goods). Likewise, the three-digit SIC categories 5010 (motor vehicles and motor vehicle parts and supplies), 5020 (furniture and home furnishings), 5030 (lumber and other construction materials), 5040 (professional and commercial equipment and supplies), 5050 (metals and minerals, except

petroleum), 5060 (electrical goods), 5070 (hardware, and plumbing and heating equipment and supplies), 5080 (machinery, equipment, and supplies), and 5090 (miscellaneous durable goods) all add up to the two-digit SIC 5000. Finally, the four-digit SIC categories 5012 (automobiles and other motor vehicles), 5013 (motor vehicle supplies and new parts), 5014 (tires and tubes), and 5015 (motor vehicle parts, used), all sum up to the three-digit SIC 5010.

In both state and county systems the current year of CBP data is used to generate Regional Purchase Coefficients (RPC) estimates for the detailed 526 industries. In the county system, CBP data is used to fill in suppressions in BEA/BLS employment and wage data.

**a. Estimation of Data Suppressions in Counties**

Initial estimates for all CBP employment suppressions are calculated by multiplying the number of firms in each of nine (9) size classes by that class' midpoint value, and summing the resulting values. Wages are estimated by using the wage rate from the next higher-level SIC industry. The initial estimates are made in a top-down order to ensure that the higher-level SIC industry's wage rate is available. At the same time, the separate sums of the data and estimates for each set of lower-level industries within a higher level are calculated. The initial estimates within a higher-level industry are adjusted by the amount that distributes higher-level industry available for allocation to suppressions relative to the sum of the estimates.

Basic CBP data is provided for over 1,000 industries. The employment data, including estimated data aggregated to this level, is not directly comparable either conceptually or in the way it is measured to the BEA data, making it necessary to adjust the data in a manner similar to the adjustment used for the BLS 202 data. This is done by taking the ratio of the state BEA employment data to the total CBP employment for all of the counties making up the state. This ratio is then applied to the county CBP data. There are now two sets of data for employment, both of which are estimates of the final data that will be used for completing the two-digit count historical series. The next step is to merge these two data sets as follows:

A. When both ES-202 employment and wage rates are available, these are used (with wage rates being converted back to wage bills).

B. When ES-202 employment is available, but the ES-202 wages are suppressed, the employment is used as is, and the wages are calculated by multiplying the employment by the state wage rates calculated from state BEA data (after it has been processed by the REMI model).

C. When ES-202 employment is not available, CBP employment is used (where possible) and the wages are then calculated by multiplying this employment by the BEA-based state wage rates.

D. When neither ES-202 employment or CBP employment is available remaining suppressions in employment and wages are filled in by an estimating program, EST202.

ES-202 estimates the suppressions for wage bill and total employment data series one county (ie., file) at a time. The FORTRAN code which performs the estimations is divided into 5 program routines. The first routine creates a 1-to-1 correspondence between the wage and employment series with respect to zero-valued observations. The second and third routines calculate estimates of suppressed observations using different algorithms, including two forms of interpolation, depending on the available data. Suppressed employment is determined in the second routine and suppressed wages in the third. The most used type of interpolation takes the form of:

$$E(T)^{i}_{Cnty} = E(t_0)^{i}_{Cnty} + (E(t_1)^{i}_{Cnty} - E(t_0)^{i}_{Cnty}) x \left( \frac{E(T)^{i}_{State} - E(t_0)^{i}_{State}}{E(t_1)^{i}_{State} - E(t_0)^{i}_{State}} \right)$$

where $E$ represents employment, $i$ the industry, $t_0$ & $t_1$ the year endpoints of suppressed employment series, $T$ the year of estimated employment, and *Cnty* & *State* the county and state of the series being estimated. The year $T$ takes values $t_0 < T < t_1$ in the formula.

The remaining two routines of code are auxiliary to the second and third sections and are used to account for potential math errors, to set logical conditions, and to prevent calculation of a wage bill with an implicit wage rate outside a fixed tolerance interval. The tolerance interval is defined to be

$$\left(\frac{1}{3}\right) x WR_S < WR_C < 3x WR_S$$

where $WR_C$ is the county specific implicit wage rate and $WR_S$ is the corresponding state wage rate.

The above procedures create one file for each county that contains the merged data. The final processing step (a program called SMUTH) checks those industries in the years 1975-1987 known to be affected by the change from 1977 SIC codes to 1987 SIC codes. First the industries are grouped according to SIC code changes. For instance, Miscellaneous Business Services, Motion Pictures and Miscellaneous Professional Services form a group because employees moved between them. We take the sum of the number of employees in the group in 1987 and divide by the sum of the number of employees in the group in 1988. This ratio is compared to the ratio of 1987 employees to 1988 employees in each of the three industries. The absolute value of the difference between the ratio of the group and the individual ratios is calculated. If the absolute value is more than .03 then the group ratio is applied to the 1988 individual values to give an adjusted 1987 value. In order to maintain the integrity of the industry totals by year a normalizing process is performed. The total number of employees in a related group of affected industries is calculated from the pre-SMUTH file (OLDSUM) and also from the data after the first adjustment in SMUTH (PRELSUM). The ratio of OLDSUM/PRELSUM is then applied to the individual component industries of PRELSUM. The resulting files are then aggregated according to client-specific region definitions.

The individual county data files created are used as initial estimates for the county employment and wages. These initial estimates are used in the same sort of reconciliation process that was explained earlier, using higher- and lower-order industries in higher- and lower-order regions. In the case of counties, the higher-order region is the state. This process is undertaken in such a way that the employment and wage totals for all of the counties in the state will add up to the state total employment and wages for each of the industries. On the other hand, the BEA major industry data has been developed for all of the counties within the state based again on the same process of reconciliation of the county-major industries with the state-major industries and, thus, the major industry data for each county within the state have to be reconciled with the total of the two-digit industries within the major industry in question. This is done using the same iterative process explained in the first section. The only modification of this procedure is for durables and non-durables, where we are not given the BEA major industry data. Instead, the data is apportioned between durable and non-durables based on initial estimates that we get from the combined CBP/202 data file. This is done prior to the final reconciliation so that all of the two-digit industries add up to the state two-digit industries, and all of the two-digit industries in a county add up to the one-digit total within that county. The end result of this process is employment data by area and wage data by area.

## B. SUPPLEMENTARY HISTORICAL DATA

### i. Fuel Cost Data

State-specific relative fuel costs for three types of fuel (electricity, natural gas, residual fuel) are calculated for the industrial (all manufacturing) and commercial (all non-manufacturing) sectors of the model based on unit cost data obtained from the Energy Information Administration/State Price and Expenditure Report.

### ii. Fuel Weight Data

Purchased fuel weights represent the proportion of an industry's fuel expenditures that are for electricity, natural gas, or residual fuels.

Purchased fuels and electric energy consumed by manufacturing industries are obtained from the 1982 Census of Manufactures-Subject Series; Table 3. These numbers are available for twenty manufacturing industries (SIC codes 20-39) for each state. All suppressed numbers are estimated using state proportions, and states that do not report a specific SIC code are assigned zero for that entire industry.

Purchased fuels and electric energy consumed by mineral industries are obtained from the 1982 Census of Mineral Industries-Subject Series; Table 3. The average state proportions are used for REMI's mining industry 22. Disclosures are treated the same way as above.

Purchased fuels and electric energy consumed by construction industries are obtained from the 1982 Census of Construction Industries-Industry Series; Table 4. This information is available only for the United States, so these numbers have been used for all states for the REMI construction industry 23.

Fuels and electric energy expenditures by transportation and electric utility sectors are obtained from the 1982 State Energy Price and Expenditure Report-Energy Information Administration. U.S. weights are used for all states for the REMI transportation and public utility industries 24-30.

Fuels and electric energy expenditures by service industries are obtained from the 1982 Census of Service Industries-Industry Series; Table 4. Again, U.S. numbers are used for all states for REMI service industries 31-34 and 38-48. Since natural gas is not reported, previous proportions found in the model are used for breaking out natural gas from residual fuels.

Fuels and electric energy expenditures by retail industries are obtained from the 1982 Census of Retail Trade-Industry Series; Table 4. U.S. proportions are used for all states for the REMI retail trade industries 35-36. The natural gas weight is determined using the same method as above.

Fuels and electric energy expenditures by wholesale industries are obtained from the 1982 Census of Wholesale Trade-Industry Series; Table 4. U.S. proportions are used for all states for the REMI wholesale industry 37. The natural gas weight is determined the same way as for the other service industries.

Purchased fuels and electric energy consumed by the agriculture, food, and fisheries service industry are obtained from the 1982 Census of Agriculture-United States Data; Table 7. U.S. proportions are used for this industry, which is the 49th REMI industry.

### iii. Tax Data

To calculate the cost of capital variable, the EDFS model requires both state-specific and national-average corporate profit and property tax rates. In the absence of a consistent and complete data source, the tax rates are estimated as follows.

*State and U.S. corporate profit tax rates* are defined as the amount of tax collections divided by the amount of corporate profits. The tax collections are found in the Government Finances (Revenue) publication and are converted from fiscal year to calendar year. Profits for states are constructed by: 1) converting the industry BEA wage data into output by dividing by the technological ($a_{i,j}$) matrix's labor per dollar output coefficient; 2) calculating profits as the product of the technological matrix's capital per dollar output coefficient times output for the industry in question; 3) smoothing the profit time series; and 4) normalizing to reproduce the reported average state and local corporate profit tax rate. Corporate profits for the U.S. are taken from the *Survey of Current Business.*

*State and U.S. property tax rates* are defined as the amount of tax collections divided by the level of residential and non-residential capital stock. Again, tax collections are taken from the Government Finances (Revenue) publication, and converted from fiscal to calendar year. Non-residential capital stock is calculated by estimating the state's share of national non-residential capital stock based on implicit profits. Residential capital is estimated similarly, but a permanent disposable income concept is used as weight. U.S. investment and capital stock data for residential and non-residential structures are also found in the *Survey of Current Business.*

### iv. Cost of Capital Data

In addition to the tax rates described above, exogenous variables for the cost of capital equation include Moody's AAA bond rates, investment tax credit rates, and the proportion of business capital financed by bonds and loans. The latter is estimated from the Quarterly Financial Report for Manufacturing, while all of the other variables are taken from the *Survey of Current Business.*

### v. Gross State Product Data

For a state, gross state product originating (GSPO) by industry is the contribution of each industry, including government, to GSP. An industry's GSPO, often referred to as its "value added," is equal to its gross output (sales or receipts and other operating income, plus inventory change) minus its intermediate inputs (consumption of goods and services purchased from other industries or imported). GSP measured as the sum of GSPO in all industries is the state counterpart of the Nation's gross domestic product (GDP) by industry from the national income and product accounts (NIPA's).

BEA prepared GSPO estimates in 61-industry detail (which REMI aggregates to its 53 industries). For each industry, estimates of gross product is composed of four components: (1) compensation of employees; (2) proprietor's income with inventory valuation adjustment (IVA) and capital consumption allowances; (3) indirect business tax and non-tax liability (IBT); and (4) other, mainly capital related charges. Most of the compensation and proprietor's income components of GSP are primarily based on BEA'S estimates of earnings by place of work, an aggregate in the state personal income series. The IBT component of GSP reflects liabilities charged to business expense, most of which are sales and property taxes levied by state and local governments. The capital charges component of GSP comprises corporate profits with IVA, corporate capital consumption allowances, business transfer payments, net interest, rental income of persons, and subsidies less current surplus of government enterprises.

The Gross State Product data is made available by the BEA for all states. The 1987-present series is based on 1987 SIC codes, while the 1969-86 series uses 1972/77 SIC codes. Also, the 1977-present

series is available in 1987 dollars, while the 1969-1976 series is in 1982 dollars. REMI has converted this data to 1987 SIC codes and 1987 dollars. To make the model internally consistent, the GSP data is renormalized for all states to match the value-added data from the BLS U.S. input-output table. BEA reports the GSP data in local (state) real dollars (local nominal dollars deflated by U.S. prices). The REMI model requires U.S. real dollars, so we further adjust the data for the local price of value added. Therefore, a relatively large GSP value may be due to higher prices, not higher production levels. This data represents value-added for each industry and is used in the model for the computation of the relative productivity measure and other variables as explained in Chapters 2 and 3. For sub-state regions, the relative productivity measures estimated for the state are used.

The other side of gross state product is the demand side, which in the REMI model is divided into components of consumption, investment and government spending. The thirteen consumption categories in the model are adjusted to be region specific for a base year using cross-section consumer/expenditure survey data. Using this consumer survey data, we are able to estimate the special consumption patterns at the 13-component level for the region relative to the rest of the United States. These consumption patterns are then used to adjust the time-series data estimates based on changes in disposable income within the sub-state areas in question relative to the United States. Residential and non-residential investment is estimated by apportioning the U.S. series by the region's share of Gross Construction Product. State and local government spending is apportioned by state and local government employment in the area.

U.S. Gross Domestic Product data for the 25 final demand categories are updated from the *Survey of Current Business* for history years.

#### vi. Housing Price Data

State-specific median values of owner-occupied housing units are obtained from the Census of Housing for the years 1970, 1980 and 1990. The National Association of Realtors' regional and metropolitan growth rates for median sales price of existing single-family homes are then used to interpolate between these three census benchmark years and to estimate state housing prices after 1990. To determine the national housing price figure, from which selling price for real estate relative to the U.S. is calculated, a weighted average of all the states is determined using real disposable income as the weight. County specific median values of owner-occupied housing units are also obtained from the Census of Housing for 1970, 1980 and 1990. State and metropolitan housing price values are used to interpolate between these three census benchmark years, and to extend the series beyond 1990.

### C. NATIONAL FORECAST DATA

#### i. BLS Forecast Data

The REMI EDFS model's baseline forecast is primarily based on the BLS Employment Outlook: 1997-2008 projections, published in the November 1997 issue of the *Monthly Labor Review*.

Between 1997 and 2008, REMI uses a labor-force-trended forecast. The output measures follow the definitions and conventions used by the BEA in their input-output tables. These input-output measures are based on producer's value and include both primary and secondary products and services. The main data sources for compiling the output times series for manufacturing industries are the Census and the Annual Survey of Manufacturers. Data sources for non-manufacturing industries are more varied. They include the Service Annual Survey, National Income and Products Accounts data on new construction and personal consumption expenditures, IRS Business Income Tax Receipts, Department of Transportation data, and many other sources. The employment data are from the BLS surveys of establishments and of households. They also include unemployment insurance data.

After 2008, the BLS moderate-growth labor force participation rates and the Census Bureau's middle population projections for the U.S. are used to forecast the labor force. An initial estimate of final demand is made, and then adjusted until the resulting industry employment comes in line with the labor force. Once the BLS trended forecast is in place, and then extended to 2035, the U.S. interface procedure is run using the latest short-term national forecast from the University of Michigan's Research Seminar in Quantitative Economics (RSQE). This updates the U.S. forecast with current national business cycles.

The technical coefficients matrix and PCE bridge matrix are estimated and obtained in the following manner. The industry-by-industry national input-output models under industry-based technology are first built for the benchmark years of 1983-1997 and 2008. These benchmark input-output accounts are provided by Bureau of Labor Statistics (BLS) and expressed in terms of 1992 dollars and 1987 SIC codes. For the non-benchmark years between 1997 and 2008, a linear interpolation method is used to estimate the coefficients. The 1983 coefficients are held fixed back to 1969, and the 2008 coefficients are held fixed forward to 2035. Holding the coefficients constant for the years prior to 1983 does not guarantee that the input-output matrix will produce the same industry outputs that BLS reported when given the historical data of final demand.[1] To ensure that total final demand will equal total value-added from input-output models, we use generated industrial output instead of BLS outputs.

The BLS includes as "special industries" noncomparable imports, scrap, and used and secondhand goods. For noncomparable imports and used and secondhand goods, there is no production in the United States, and thus, no domestic commodity or industry output. For scrap, there is domestic production, although that production is not by a "scrap" industry, but by other industries as a part of the production of their output. For REMI purposes, we need to account for these values in our industry-by-industry matrix. For scrap and used and secondhand goods, the great majority of which automobiles, we made the assumption that most of these goods would at some point pass through the wholesale industry, so we simply aggregated them with wholesale. For noncomparable imports, we added the values (which are negative) to the industry that "used" these imported goods (the commodity by industry diagonal in the USE table), and then balanced the table by subtracting them from the commodity by imports column in the demand table.

### ii. Occupation Data

The REMI model system uses a fixed-proportion occupation-by-industry matrix in calculating the occupational demands used in the model's wage equation. The matrix is also used in estimating the parameters in the wage equation (see Treyz, George I. and B.H. Stevens, "The TFS Regional Modeling Methodology," *Regional Studies*, Vol. 19.6, 1985, pp. 547-562).

The occupation matrix used in the REMI model is based on the BLS 1997 and projected 2008 National OES Matrices which contain employment for over 600 detail and summary occupations and over 300 detail and summary industries. The detailed industries are aggregated to 53 or 14 industries for the EDFS-53 model or the EDFS-14 model, respectively. The detailed occupational employment data within each of the industries is aggregated to 94 or 17 occupations for EDFS-53 or EDFS-14 models, respectively. The fixed proportion of occupational employment is calculated by summing the employment across an industry, and then dividing each occupation by the industry total. The rates of occupational change between 1997-2008 are calculated by linear interpolation, are then extended back historically at the same rate of change, and extended forward at one-half the rate of change.

---

[1] For more detail on how REMI creates the U.S. model, see REMI "Building U.S. National and Regional Forecasting and Simulation Models", Economic Systems Research, Vol.5, No.1, 1993.