Socioeconomic Indicators (SEI) Module Documentation

Table of Contents

Income Distribution

Changes in SEI Module

Results displayed by industry quintile show regionally calculated quintiles. Regional quintiles are calculated by ordering all industries by the model's All Regions average annual compensation rate in the last history year and creating five groups with roughly the same number of industries in each.

A new result, Weighted Compensation Rate, shows how compensation rates have changed by quintile, while excluding any effects caused by changes to the quintile's employment mix. For each quintile, the result is calculated as a weighted average of each member industry's average annual compensation rate, where each weight is the industry's share of total employment in the baseline forecast.

Decomposing Policy Effects on Employment, Wages, and Prices by Income Groups

A table entitled "Income Distribution" is generated for the purpose of evaluating the economic impacts of policies on different compensation and income groups. The following documentation discusses the occupation classifications, the economic background, and the operation of this procedure.

Occupational Classification

Median weekly U.S. wage rates for 95 occupations are obtained from the 2019 BLS Employment and Earnings. The wage rates are ranked in ascending order, and then divided into five groups. The range of occupational wage rates are listed below (see Exhibit 2 for a more detailed list):

	Range of Occupational Wage Rates
Group 1 (First 20%)	\$490 - \$618
Group 2 (Second 20%)	\$624 - \$767
Group 3 (Third 20%)	\$799 - \$997
Group 4 (Fourth 20%)	\$1012 - \$1213
Group 5 (Fifth 20%)	\$1228 - \$2027

Personal Expenditure Classification

Average annual expenditures for consumers by quintiles or by ranges of income are obtained from the 2019 BLS Consumer Expenditure Survey. The eight income ranges are as follows:

Group	**Range**
1	<\$15,000
2	\$15,000-\$29,999
3	\$30,000-\$39,999
4	\$40,000-\$49,999
5	\$50,000-\$69,999
6	\$70,000-\$99,999
7	\$100,000-\$149,999
8	\$150,000+

Economic Background

The percentage changes from control forecasts for industrial compensation rates and occupational wage rates, employment, and compensation are reported on the table. Note that the simulation (alternative) forecast must be generated before running the software. For each item, the percentage change is calculated as follows:

Industrial employment:

$$\Delta E_I = \left(\sum_{j \in I} E_j^a \Big/ \sum_{j \in I} E_j^c - 1 \right) * 100$$

j = 1,…,66

I = 1,…,5

where ΔE_I is percentage change of employment for industry group *I*, and E_j^a and E_j^c are employment for industry *j* (in group *I*) from a(alternative) and c(control) forecasts.

Industrial compensation:

$$\Delta WSD_I = \left(\sum_{j \in I} WSD_j^a \Big/ \sum_{j \in I} WSD_j^c - 1 \right) * 100$$

j = 1,…,66

I = 1,…,5

where ΔWSD_I is percentage change of compensation for industry group *I*.

Industrial compensation rate:

$$\Delta w_I = \left[\left(\sum_{j \in I} WSD_j^a \Big/ \sum_{j \in I} E_j^a \right) \Big/ \left(\sum_{j \in I} WSD_j^c \Big/ \sum_{j \in I} E_j^c \right) - 1 \right] * 100$$

j = 1,…,66

I = 1,…,5

where Δw_I is percentage change of compensation rate for industry group *I*.

Occupational employment:

$$\Delta OE_I = \left(\sum_{j \in I} OE_j^a \Big/ \sum_{j \in I} OE_j^c - 1 \right) * 100$$

j = 1,…,95

I = 1,…,5

where ΔOE_I is percentage change of employment for occupation group *I*, OE_j^a and OE_j^c are employment for occupation j (in group *I*) from a(alternative) and c(control) forecasts.

Occupational wage bill:

$$\Delta OWSD_I = \left(\sum_{j \in I} OWSD_j^a \Big/ \sum_{j \in I} OWSD_j^c - 1 \right) * 100$$

j = 1,…,95

I = 1,…,5

where $\Delta OWSD_I$ is percentage change of wage bill for occupation group *I*. The REMI model does not predict the occupational wage bill directly, but the change in occupational wage rate (i.e. $A_{j,t+1} = {ow_{j,t+1}} / {ow_{j,t-1}}$, where $ow_{j,t}$ is wage rate for occupation *j* at year *t*). In order to obtain *OWSD*, we apply

$$ow_{j,T+1} = ow_{j,T}^{us} * \left(A_{j,T+1} + 1\right)$$

and

$$ow_{j,T+k+1} = ow_{j,T+k} * \left(A_{j,T+k+1} + 1\right)$$

for *k*>0

where the subscript *T* denotes the last history year, and the superscript *us* represents the U.S.

Then,

$$OWSD_j = OE_j * ow_j$$

Occupational wage rate:

$$\Delta ow_I = \left[\left(\sum_{j \in I} OWSD_j^a \Big/ \sum_{j \in I} OE_j^a \right) \Bigg/ \left(\sum_{j \in I} OWSD_j^c \Big/ \sum_{j \in I} OE_j^c \right) - 1 \right] * 100$$

j = 1,…,95

I = 1,…,5

where Δow_I is percentage change of wage rate for occupational group *I*.

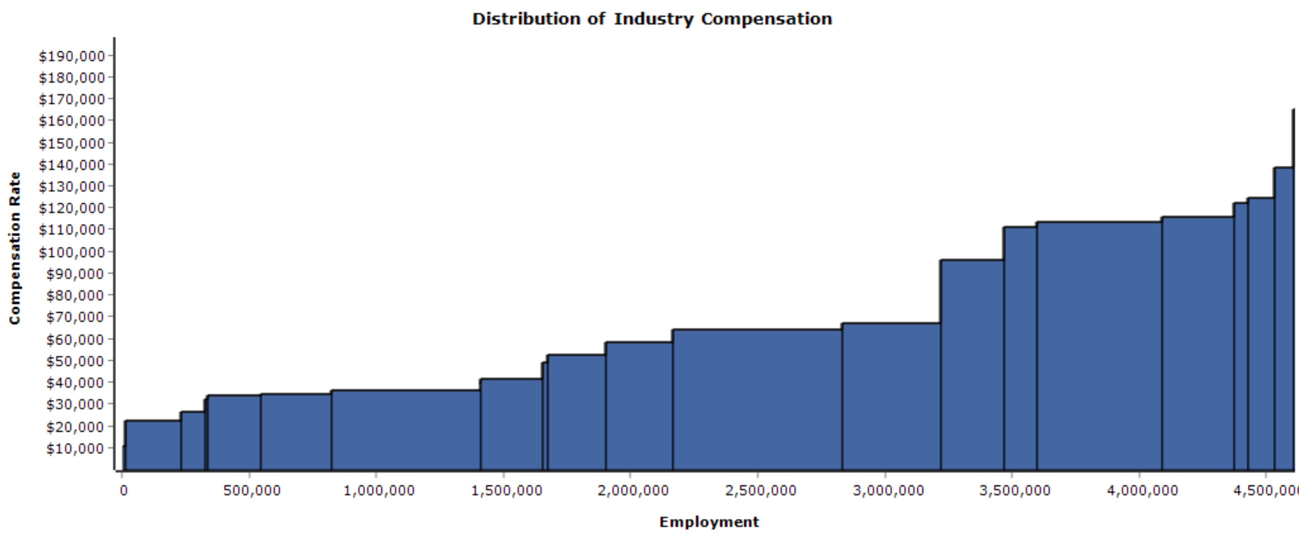
Consumption Price by Income Quintile & Range

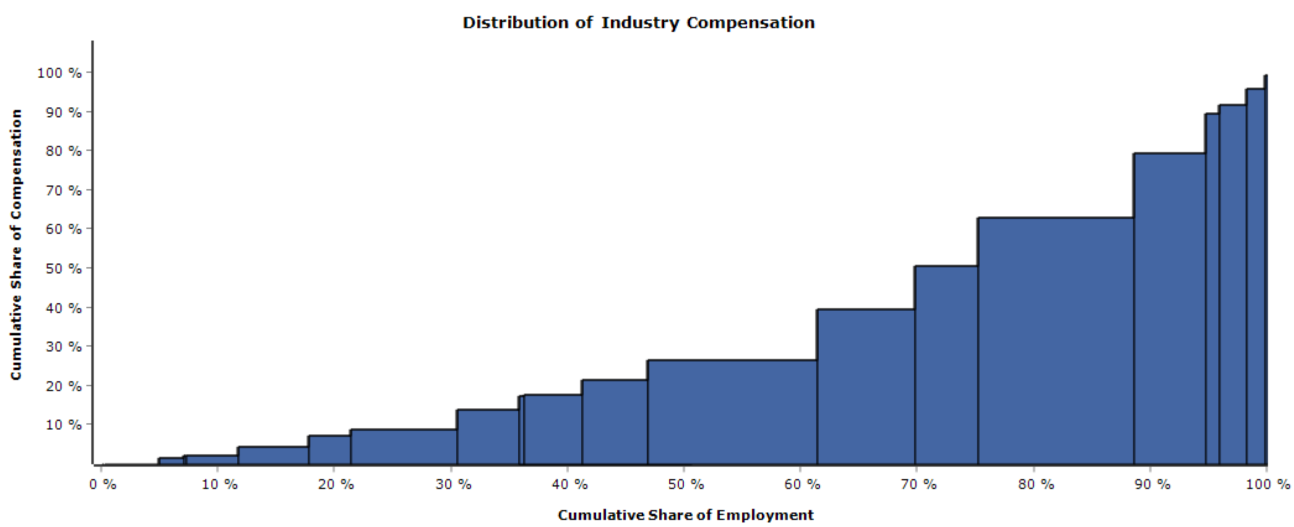
A new result, the Consumption Price Distribution, is calculated as the weighted average of Consumption Commodity Prices for each income group (either quintiles or eight income ranges). The weights for each commodity price are based on each group's national average share of spending devoted to that commodity category. Spending data by income group come from the BLS Consumer Expenditure Survey.

Inequality Coefficient

The SEI module's compensation inequality coefficient result is a measure of inequality in the distribution of industry-average compensation rates across jobs. The coefficient is calculated by applying the Gini methodology to REMI's employment by industry and average annual compensation rate by industry data. Being a type of Gini coefficient, the value of the coefficient is bounded between 0 and 1. Values closer to 0 represent a more equal distribution of compensation, while values closer to 1 represent a more unequal distribution.

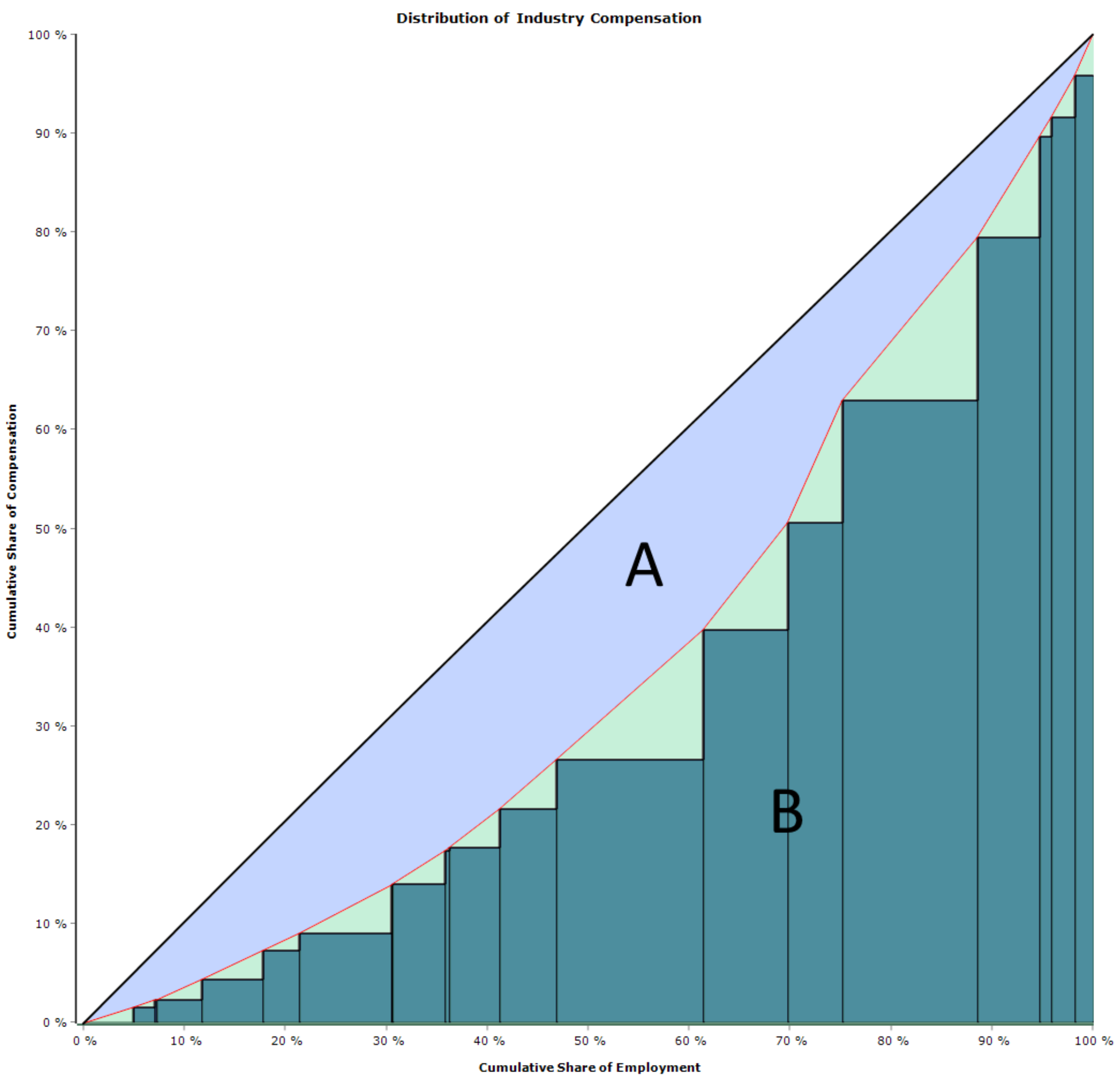
A sample distribution of industry compensation for a 23-sector model, ordered by average compensation rate, is shown in the following figure. Each bar represents an industry, with the bar's width being equal to the amount of employment in the industry and the bar's height being equal to the industry's average annual compensation rate. The area of each bar is equal to the total amount of compensation paid in that industry.





To use a distribution like this to calculate a Gini coefficient, the distribution must first be normalized into cumulative shares. The sample distribution is shown in normalized, cumulative share form in the next figure. Set up this way, a distribution can be analyzed for how much each industry deviates from complete equality. For example, in this instance it can be seen that the bottom 30% of jobs earn only about 10% of all compensation.

Connecting each point on the distribution at this stage yields a line known as a Lorenz Curve. Compared against the Line of Equality — a Lorenz Curve given by a distribution with perfect compensation equality, equal to a straight 45-degree line — we can visualize the degree of inequality in our sample distribution. This is demonstrated in the last figure. Finally, a Gini coefficient can be calculated from this type of visual representation as the ratio $\boldsymbol{G} = \frac{\boldsymbol{A}}{\boldsymbol{A}+\boldsymbol{B}}$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ correspond to areas marked in the below figure. $\boldsymbol{A}$ is the area between the Lorenz Curve and the Line of Equality, while $\boldsymbol{B}$ is the area below the Lorenz Curve.



Employment by Race and Gender

The SEI module's employment by demographics feature was developed as follows.

First, a weighted average of 5-Year (2015-2019) sample with 2019 as the latest year is calculated using ACS as the external data source. This dataset provides national-level "employment by race and gender shares for each occupation" across REMI's 95 occupations. Dividing the national ACS data by "REMI National Labor Force by Race and Gender" data, an "Occupation Weights" dataset is produced to feed into model building. The "Occupation Weights" dataset describes how national employment represents national labor force for each occupation. It captures the occupational effects in employment demographic distributions.

National occupation weights are applied to all regions. Although ACS data has both state- and county-level data, this procedure only uses the national level because there are many missing values in state- and county-level data, making it difficult to estimate missing values and ensure data quality at the same time. Using national level data can help minimize sampling errors. Additionally, there is no apparent historical or theoretical reason that explains regional difference in employment by race and gender shares for each occupation. In the forecast years, we assume occupation weights will not change significantly as there will not be drastic demographic distribution changes for each occupation.

The REMI labor force measurement "by place of residence" is converted to "by place of work," to be consistent with the measurement of Employment. Using county-level "Commuter Income Flows," "Earnings by Place of Residence," and county "Labor Force," an estimation of the labor commuters (unit = individuals) inflows and outflows for a certain county are produced. This procedure assumes that commuters have the same demographic distribution as laborers in their place of residence. The adjusted "Labor Force by Place of Work" will be equal to "Labor Force by Place of Residence," plus commuter inflows and minus commuter outflows.

In history years, the labor force demographic shares of the origin regions are used to distribute commuter inflows. In forecast years, since we do not have the labor force demographic shares forecasts for regions outside of the model region, we estimate these shares based on last history year data, which gives a relatively accurate demographic distribution of commuter inflows. Then, national labor force demographic distribution forecasts are used to adjust "last history year commuter inflows by race and gender."

Employment by occupation, race, and gender shares are calculated using new national "Occupation Weights" and "Labor Force by Place of Work Shares," which give the distribution of employment by race and gender within individual occupations and counties. To get county-level total employment demographic distribution, the calculated shares are applied to REMI baseline "Employment by Occupation", the resulting employment by occupation, race and gender are then summed across all occupations.

Methodology

Step 1: Calculate national shares by race/gender for each occupation, using ACS data (year 2015 - year 2019 weighted average).

For demographic group i, occupation j:

$$National\ Employment_{i,j} = \sum_{k=1}^{51} Employment_{i,j}$$

$$Total\ National\ Employment_{j} = \sum_{k=1}^{51} Total\ Employment_{j}$$

$$National\ Employment\ Shares_{i,j} = \frac{National\ Employment_{i,j}}{Total\ National\ Employment_{j}}$$

where:

State k = 1, 2, ..., 51

Demographic group i = 1,2, ..., 8

Occupation j = 1,2, ..., 95

Step 2: Calculate Occupation weights for each race, gender and occupation.

$$Occupation\ Weights_{i,j} = \frac{National\ Employment\ Shares_{i,j}}{National\ Labor\ Force\ Shares_{i}}$$

where:

Demographic group i = 1,2, ..., 8

Occupation j = 1,2, ..., 95

Step 3: Estimate labor force race shares by place of work.

Step 3.1: Transform commuter data from number of jobs to number of labor.

$$\frac{Commuter\ Income\ from\ county\ w\ to\ county\ r}{Earnings\ by\ place\ of\ residence\ in\ county\ r} = \frac{Estimated\ Commuters\ by\ Labor\ from\ county\ r\ to\ county\ w}{Labor\ Force\ in\ county\ r}$$

$$Estimated\ Commuters\ by\ Labor\ from\ county\ r\ to\ county\ w = \frac{Commuter\ Income\ from\ county\ w\ to\ county\ r}{Earnings\ by\ place\ of\ residence\ in\ county\ r} * Labor\ Force\ in\ county\ r$$

where:

$$County\ r = Place\ of\ Residence$$

$$County\ w = Place\ of\ Work$$

Step 3.2: Calculate commuter outflows by race and gender.

For demographic group i:

$$Commuter\ outflows\ by\ race_i = Commuter\ outflows * Labor\ force\ shares_i\ in\ county\ j$$

where:

Demographic group i = 1,2, ..., 8

$$County\ j = County\ of\ interest$$

Step 3.3: Calculate commuter inflows by race and gender.

For demographic group i, origin county K:

$$Commuter\ inflows\ by\ race_i\ from\ county\ K$$
$$= Commuter\ inflows\ from\ county\ K * \ Labor\ force\ shares_i\ in\ county\ K$$

where:

Demographic group i = 1,2, ..., 8

$$County\ K = Other\ counties$$

$$K \neq j$$

Step 3.4: Calculate labor force by place of work.

For demographic group i, county of interest j, origin county K:

$$Labor\ force\ by\ place\ of\ work\ _{i,j}$$
$$= \ Labor\ force\ by\ place\ of\ residence_{i,j}$$
$$+ \ sum(commuter\ inflows_{i,j}\ from\ county\ K) - \ commuter\ outflows_{i,j}$$

where:

Demographic group i = 1,2, ..., 8

$$K \neq j$$

Step 4: Calculate the employment by occupation, race and gender shares.

For region k, demographic group i, occupation j:

$$Employment\ Shares_{k,i,j} = Labor\ Force\ by\ place\ of\ work\ shares_{k,i} * \ Occupation\ Weights_{i,j}$$

where:

Demographic group i = 1,2, ..., 8

Occupation j = 1,2, ..., 95

Employment by Educational Attainment

The SEI module's employment by educational attainment feature estimates the number of new jobs created that do not require a college education.

This result was calculated using the REMI policy variable "Employment by Occupation" multiplied by the percentage of educational attainment by Occupation, the data for which came from Table 5.3 Educational attainment for workers 25 years and older by detailed occupation, 2019, BLS.

The 790 BLS line item occupations were given by seven education levels. This data was processed so that the occupations were aggregated into three levels, 18 major occupations, 95 occupations, and 481 detailed occupations. To calculate summary occupations, the line item occupations from the BLS Employment Projections program table were matched to the REMI Occupations list occupation codes.

Summary occupations were calculated as follows:

$$\frac{\text{Employment for education level \& summary occupation}}{\text{Employment for summary occupation}} = \frac{\sum \text{Education attainment} * \text{Employment for line item occupations}}{\sum \text{Employment for line item occupations}}$$

To address military occupation, as there is no military in the BLS data, this procedure used educational attainment for the total of all other occupations.

The new results give a table with the REMI 95 occupations by two aggregative education levels, "Bachelor's degree and above" and "No Bachelor's degree" and a table with the REMI 95 occupations by seven education levels: Less than high school diploma; High school diploma or equivalent; Some college, no degree; Associate's degree; Bachelor's degree; Master's degree; Doctoral or professional degree.

Labor Force Participation by Race and Gender

Labor Force Equations

$$LF_t^k = \sum_{i=1}^{n} PR_{i,t}^k * N_{i,t}^k$$

Regional Participation Rates

The national participation rate is calculated using data from BLS. REA and RWR are variables calculated in the model.

$$PR_{i,t}^k = \beta_1^k * \left(REA_t^k\right)^{\beta_2} * \left(RWR_t^k\right)^{\beta_3} * PR_{i,t}^u + PRIPVA_t^k$$

(3-11)

where:

LF_t^k =The labor force in area *k*.

$PR_{i,t}^k$ = The participation rate (i.e. the proportion of the relevant population that is in the labor force) in age cohort *i*, area *k*.

$N_{i,t}^k$ = The number of people in cohort *i,* area *k*.

β_1^k = The fixed effect for area *k*.

β_2, β_3 = The parameters estimated on the basis of pooled or national time series.

$$REA_t^k = \left(\frac{EA_t^k}{EA_t^u}\right)$$

(3-12)

$$EA_t^k = EA_{t-1}^k + \lambda_E\left(EO_t^k - EA_{t-1}^k\right)$$

$$EA_t^u = EA_{t-1}^u + \lambda_E\left(EO_t^u - EA_{t-1}^u\right)$$

EO_t^u = A synthetic labor force based on the local population at fixed national participation rates.

EO_t^k =The residence adjusted employment.

RWR_t^k = The relative real compensation rate.

λ_E = An estimated parameter $0 < \lambda_E < 1$.

The β_2, β_3values by age cohorts, gender, and racial/ethnic groups have been estimated for 160 (20x2x4) age cohorts in the U.S. The β_1^k parameter is a fixed effect for area *k* calibrated to the measured labor force (see Treyz, Christopher, and Lou, 1996).

Parameters

Estimated Labor Force Participation Rates Elasticities (Males) Using Time Fixed Effects Regression, Years 2010-2014

			Male-White Alone-Non-Hispanic				Male-Black Alone-Non-Hispanic			
			RWR (2)		REA (3)		RWR (2)		REA (3)	
Age group available from the CPS (1)	Age group in PI+		PI+	Estimates	PI+	Estimates	PI+	Estimates	PI+	Estimates
00 to 17	1	16-19	-0.12	-0.77	0.84	1.54	-0.91	0.41	1.01	0.46
18 to 24	2	20-21	0.13	0.01	0.38	0.31	-0.19	-0.27	0.45	0.61
	3	22-24	0.08	*0.01*	0.25	*0.31*	0.45	*-0.27*	0.40	*0.61*
25 to 34	4	25-29	0.12	0.01	0.12	0.26	0.09	0.21	0.30	0.62
	5	30-34	0.09	*0.01*	0.08	*0.26*	0.31	*0.21*	0.21	*0.62*
35 to 44	6	35-44	0.10	0.16	0.08	0.25	0.09	0.22	0.27	0.61
45 to 54	7	45-54	0.17	0.03	0.16	0.40	0.10	0.21	0.47	0.43
55 to 59	8	55-59	0.33	0.01	0.42	0.62	0.18	0.28	0.73	0.42
60 to 64	9	60-61	0.75	0.37	0.89	1.22	0.46	0.39	1.19	0.61
	10	62-64	0.75	*0.37*	0.89	*1.22*	0.46	*0.39*	1.19	*0.61*
65 to 74	11	65-69	0.85	0.52	1.41	1.49	0.21	0.33	1.54	0.44
	12	70-74	0.77	*0.52*	1.49	*1.49*	0.98	*0.33*	2.24	*0.44*
75 to 80+	13	75+	0.99	1.32	1.55	1.46	0.13	0.77	1.66	-0.14

(1) CPS = Current Population Survey Table Creator for Adult Civilian Persons
(2) RWR = Relative Wage Rate
(3) REA = EAk / EAu, where EAk is moving average employment for state
EAu is moving average synthetic labor force for nation
(4) A figure in italics is assumed to be the same as the one above it.

Age group available from the CPS (1)	Age group in PI+		Male-Other-Non-Hispanic				Male-Hispanic			
			RWR (2)		REA (3)		RWR (2)		REA (3)	
			PI+	Estimates	PI+	Estimates	PI+ (5)	Estimates	PI+ (6)	Estimates
00 to 17	1	16-19	-0.52	-0.69	0.78	0.79	-0.12	-0.70	0.84	0.17
18 to 24	2	20-21	-0.17	-0.69	0.48	0.22	0.13	-0.69	0.38	-0.69
	3	22-24	-0.26	*-0.69*	0.54	*0.22*	0.08	*0.69*	0.25	*-0.69*
25 to 34	4	25-29	-0.11	0.45	0.37	0.18	0.12	0.06	0.12	0.10
	5	30-34	-0.28	*0.45*	0.31	*0.18*	0.09	*0.06*	0.08	*0.10*
35 to 44	6	35-44	-0.15	0.80	0.22	-0.11	0.10	0.26	0.08	-0.40
45 to 54	7	45-54	-0.07	0.47	0.23	-0.16	0.17	0.01	0.16	-0.10
55 to 59	8	55-59	0.20	0.61	-0.20	0.42	0.33	0.21	0.42	-0.01
60 to 64	9	60-61	0.91	0.05	0.36	0.06	0.75	0.22	0.89	0.22
	10	62-64	0.91	*0.05*	0.36	*0.06*	0.75	*0.22*	0.89	*0.22*
65 to 74	11	65-69	0.12	-0.64	0.59	0.25	0.85	0.37	1.41	0.34
	12	70-74	0.63	*-0.64*	1.30	*0.25*	0.77	*0.37*	1.49	*0.34*
75 to 80+	13	75+	0.47	0.63	1.25	-0.55	0.99	0.94	1.55	-0.29

(5), (6) These coefficients are the same as for the male white non-Hispanic table

Estimated Labor Force Participation Rates Elasticities (Females) Using Time Fixed Effects Regression, Years 2010-2014

			Female-White Alone-Non-Hispanic				Female-Black Alone-Non-Hispanic			
			RWR (2)		REA (3)		RWR (2)		REA (3)	
Age group available from the CPS (1)	Age group in PI+		PI+	Estimates	PI+	Estimates	PI+	Estimates	PI+	Estimates
00 to 17	1	16-19	-0.82	-1.10	0.45	2.11	-0.56	-0.28	0.98	0.97
18 to 24	2	20-21	-0.20	-0.16	0.26	0.56	-0.31	-0.55	0.00	0.23
	3	22-24	-0.03	*-0.16*	0.23	*0.56*	0.16	*-0.55*	0.06	*0.23*
25 to 34	4	25-29	-0.06	0.17	0.21	0.42	0.04	0.03	0.24	0.31
	5	30-34	-0.12	*0.17*	0.26	*0.42*	0.05	*0.03*	0.30	*0.31*
35 to 44	6	35-44	-0.08	0.07	0.27	0.53	0.15	-0.36	0.37	0.03
45 to 54	7	45-54	0.01	0.18	0.41	0.64	0.29	0.11	0.74	0.60
55 to 59	8	55-59	0.20	0.30	0.61	0.82	0.15	0.28	0.86	0.66
60 to 64	9	60-61	0.43	0.54	1.06	1.40	0.32	0.31	1.43	0.24
	10	62-64	0.43	*0.54*	1.06	*1.40*	0.32	*0.31*	1.43	*0.24*
65 to 74	11	65-69	0.69	0.75	1.16	1.81	1.15	0.23	1.61	0.03
	12	70-74	0.56	*0.75*	1.19	*1.81*	0.09	*0.33*	1.22	*0.03*
75 to 80+	13	75+	0.87	0.72	1.20	1.73	1.49	0.37	2.39	0.34

(1) CPS = Current Population Survey Table Creator for Adult Civilian Persons
(2) RWR = Relative Wage Rate
(3) REA = EAk / EAu
EAk is moving average employment for state
EAu is moving average synthetic labor force for nation
(4) A figure in italics is assumed to be the same as the one above it.

Age group available from the CPS (1)		Age group in PI+	Female-Other-Non-Hispanic RWR (2) PI+	Female-Other-Non-Hispanic RWR (2) Estimates	Female-Other-Non-Hispanic REA (3) PI+	Female-Other-Non-Hispanic REA (3) Estimates	Female-Hispanic RWR (2) PI+ (5)	Female-Hispanic RWR (2) Estimates	Female-Hispanic REA (3) PI+ (6)	Female-Hispanic REA (3) Estimates
00 to 17	1	16-19	-0.69	-0.80	0.29	0.73	-0.82	-0.64	0.45	0.00
18 to 24	2	20-21	-0.20	-0.74	0.41	0.48	-0.20	-0.26	0.26	0.84
	3	22-24	-0.17	*-0.74*	0.39	*0.48*	-0.03	*-0.26*	0.23	*0.84*
25 to 34	4	25-29	0.10	-0.11	0.53	0.59	-0.06	0.18	0.21	0.60
	5	30-34	0.13	*-0.11*	0.57	*0.59*	-0.12	*0.18*	0.26	*0.60*
35 to 44	6	35-44	-0.01	0.63	0.49	0.42	-0.08	0.07	0.27	0.82
45 to 54	7	45-54	-0.03	-0.07	0.50	0.59	0.01	-0.24	0.41	0.68
55 to 59	8	55-59	0.08	-0.44	0.48	0.10	0.20	0.18	0.61	0.65
60 to 64	9	60-61	0.34	-0.49	0.84	0.55	0.43	-0.10	1.06	0.08
	10	62-64	0.34	*-0.49*	0.84	*0.55*	0.43	*-0.10*	1.06	*0.08*
65 to 74	11	65-69	0.56	-0.39	1.35	0.92	0.69	-0.25	1.16	0.81
	12	70-74	-0.55	*-0.39*	0.86	*0.92*	0.53	*-0.25*	1.19	*0.81*
75 to 80+	13	75+	0.35	0.57	-1.32	-0.42	0.87	1.16	1.20	-0.07

(5), (6) These coefficients are the same as for the female white non-Hispanic table

Demographic Component of REMI Model

The SEI module's labor force participation by demographics feature uses the following methods to break down labor force participation by race, ethnicity, and gender. This demographic component is used in other parts of the REMI model.

Overview

The demographic component of the REMI model uses a "cohort-component" method to forecast the population for a region. The components of demographic change are calculated every year for each of the age cohorts by sex and race. The population at the end of the year is equal to the population at the beginning of the year (starting population) plus births and net migration, minus deaths. The rate of change for each of the components depends on both observed historical trends in the region and on forecasted national trends. There are also several types of special populations that have different characteristics than the rest of the population and need to be treated differently. They are military, military dependents, prisoners, and college students.

Historical Data

Population

The model contains historical demographic data starting from the year 2001. Some of this data comes from official sources and some of it is estimated.

The Bureau of Economic Analysis (BEA) provides the total population for each county from its personal income and population summary tables. The BEA uses the population estimates from the U.S. Bureau of the Census. It is important that the population estimates are consistent with the personal income estimates, so the total population data is taken from the BEA instead of directly from the Census in case one bureau revises its estimates and the other does not.

The Census provides population estimates annually in 5-year age groups by sex, race, and Hispanic origin. The Census treats race and Hispanic origin as two different concepts in accordance with the guidelines from the Office of Management and Budget (OMB). The Census asked people whether they were Hispanic or Non-Hispanic and asked them to select all of the races that apply to them. Each person has a race and a separate Hispanic origin attribute, so a Hispanic person may be of any race.

There are four races in the REMI model: White, Black, Other, and Hispanic. The category White in the REMI model includes non-Hispanic people who are White *alone*, the Black category includes non-Hispanic people who are Black *alone*, and non-Hispanics of all other races *and combinations of races* are grouped into Other. Hispanic contains all people who are of Hispanic Origin, regardless of their race.

Historical population estimates for single years of age are estimated by taking the starting population in the first history year, applying the components of change by age, and adjusting the ages within each 5-year age group so the total matches the Census estimates.

Components of Change

The Census provides annual estimates of the total number of births, deaths, net domestic migration, and net international migration into each county, which are used to calibrate the county's birth rates, survival rates, and migration rates.

Birth rates can vary greatly by region and are difficult to calculate for each county because of small sample sizes. State birth rates are calculated by race and age group using data from the Center for Disease Control, National

Center for Health Statistics. Regional birth rates are created by adjusting the state rates to fit the total number of births that are estimated by the Census.

The Census publishes its own population projection and the assumptions that are used to generate it, including a natality rate and survival rate forecast. The assumed national survival rates are specific to each age, sex, race, and Hispanic origin. Regional survival rates in the model are estimated by adjusting the national survival rates to fit the total number of deaths estimated in the area.

Net international migration is the net number of people who enter the region from outside the fifty states and District of Columbia. This includes net migration from Puerto Rico and U.S. territories, Armed Forces, permanent migrants, temporary migrants (such as students), refugees, and illegal migrants. Net international migrants in each county are divided up by race according to the data from the state population projections from the Census. Each county in the state has the same racial breakdown of net international migrants as the whole state.

Interregional migration is the difference between the in- and out-migration of an area within a year, and only consist moves where the origin and destination are within the United States. The Census published its annual estimates for Net Domestic Migration at county level. In the REMI model, interregional migration is divided into two categories: retired migrants and economic migrants.

People aged 65 and older who move from one area to another are called retired migrants. They do not respond to economic conditions. Data from the Census 2000 Migration DVD is used to calculate a migration rate by age for each of the counties.

The interregional migrants under the age of 65 are called economic migrants. Economic migrants are calculated as the difference between the net domestic migrants and the net retired migrants. The labor force, relative employment opportunity, relative wage rate, and the commodity access index are used, along with the historical economic migration data, to calculate an amenity term for the area which is used in the migration equation to predict future migration.

Unexplained growth is calculated as the residual of population growth of the region during the year minus all of the other components of change.

Population Forecast

The U.S Census periodically produces national population projections based on assumptions about demographic components of change, including future births, deaths, net international migration, and population. In the latest national population projections, the Census provided deaths data including six categories of race (White Alone, Black Alone, American Indian and Alaska Native Alone, Asian Alone, Native Hawaiian and Other Pacific Islander Alone, and Two or More Races) and two categories for Hispanic origin groups (Hispanic and non-Hispanic). REMI estimated the survivals rates to form national assumptions that matched the race categories in the model. The projected birth rates used by Census in these projections contained race categories that were not compatible with the REMI categories, so the 2012 fertility projections were adjusted to form national assumptions with the appropriate race categories. The annual changes in birth and survival rates from the national assumptions are applied to the last history year regional birth and survival rates to form the forecasted rates. These birth and survival rates are multiplied by the population by sex, race, and age to predict the number of births and deaths.

The net international migration forecast for the nation by race and age is also based on the latest Census assumptions with some adjustments to create a smooth transition between the history and forecast. Each area receives the same percentage of the nation's net international migrants by race as it had in the last history year. The international migrants acquire the birth rates and survival rates of the area that they move into. Births and deaths are calculated for the migrant population by applying the birth and survival rates to half of the migrants, because the migrants arrive during the whole year and will only be in the region for half of the year, on average.

Economic migration is an endogenous calculation in the model. It depends on the economic conditions, the current labor force, and the amenity of the area. Births and deaths are added and subtracted from the economic migrants in the same manner that they are for the international migrants.

Retired migration for the area is calculated using the retired migration rates by age group. If the rate is a positive number, then the net retired migration into the area is based on the size of the 65 and older population in the rest of the nation. Otherwise, the net migration leaving the area is related to the migration rate and size of the over-65 population in the area.

Special Populations

Special populations are also estimated by age, sex, race, and Hispanic origin. The special populations are important because they are pockets of the general population that do not appear to age over time and have other special characteristics.

Active Military

The active military consists of people in the full-time duty in the active service of the Army, Navy, Marine Corps, or Air Force. It includes uniformed personnel on the active list, in training, or in military schools. Total active military population data by base is available from the military and civilian personnel reports by Defense Manpower Data Center (DMDC) of Department of Defense (DoD) until 2009. After then, the DMDC reports only provide military population data at the state level, and we use the American Community Survey (ACS) data to estimates of Armed Forces population by county. National sex and race totals for the active military are from the DoD, Office of Personnel and Readiness publication "Population Representation in the Military Services". Federal Military employment data from BEA differs from active military strength because federal military includes all active military, Coast Guard, and military reserve members who meet regularly for training. Active military personnel are not part of the labor force, and only active military members have military dependents. The ratio of active military to Federal Military employment is calculated in the last history year, and number of active military personnel in the forecast is calculated by applying that ratio to the forecasted Federal Military employment total. Military population by race and ethnicity is estimated based on Population Representation in the Military Services reports from CNA.

Active Military Dependents

Active military dependents are the family members that live with active military personnel and move when the person in the military is reassigned to a different base. They are the spouses, children, and other adult family members that depend on an active member of the military. Dependent totals are available from the DoD publication "Demographics Report". A ratio of dependents to active military member is calculated excluding active military living in group quarters housing. Estimates of Armed Forces members living in group quarters housing is provided by the ACS. This ratio is applied to the portion of active military that is not living in group quarters housing to estimate the number of dependents in each county.

College Students

College population estimates are very important, because people in the United States enrolled in college have exceeded 20 million since 2009, and they mostly fall within a very narrow age range. Students that live in places other than their hometowns during the school year are counted by the Census as residents of their new towns. It is difficult to estimate college population, because not all college students necessarily live in the same county as the college where they are enrolled. Census enrollment data by county, race, and sex, college enrollment data from American Community Survey (ACS), and data from the National Center for Educational Statistics are used to estimate the college population of an area by year. With the development of distance education technologies, the enrollment for online classes has increased dramatically in recent years. Students enrolled in online universities do not have to be physically present in the county where the college is located. Thus, college population is adjusted by online university enrollment using data from IPEDS Data Center. College students are assumed to have labor force participation rates lower than the general college-aged population.

Prisoners

Prisoners are estimated using data from the Census and annual data from the Department of Justice Statistics. The Census data provides a distribution of prisoners by race, sex, and type of facility for each county. The change in prisoners by year is based on state and national level data about local jails and state, federal, and military prisons from the Bureau of Justice Statistics and reports from various states' Department of Corrections. Prisoners are not included in the labor force.

What Makes Special Populations Special?

For the population estimates to be reasonable, it is important to recognize the special populations because they can comprise a very large portion of the population in an area that does not appear to age over time. In a college town, for example, there may be thousands of people between the ages of 18 and 22 years. Ten years later, there will not be an abundance of 28-to-32-year-olds; instead, the same 18-to-22-year-old bulge in the population will still exist. If special populations were simply not allowed to age, this would create problems in large models that have a large total special population but a small net special population. To model this situation, before the population in an area is aged, all of the special populations are returned to their "home areas". The special populations estimated in the area are taken out and the estimated special populations from the area that currently exist in other areas are brought back in. The population is then aged one year. New special populations are added to and taken out of the population in the same age distribution as the previous year. In this way, the special population appears not to age and the total population of the nation is allowed to grow normally.

Special populations are also treated differently in the labor force calculations. Labor force participation rates are only applied to the civilian, non-institutional population.

Labor Force

Historical labor force totals by county are taken from the Bureau of Labor Statistics. Participation rates by race and age are calculated using the relative compensation rate, employment opportunity, demographic characteristics, and national participation rates. They are calibrated in the history so the labor force will be consistent with the data reported by the BLS. The participation rates are multiplied by the civilian non-institutional population to generate the labor force. Forecasted national rates from BLS are used in the participation rate equation to help shape the participation rates in the model forecast.

Unemployment

The REMI Unemployment Rate result is only available in simulation forecasts and as a percent change from the baseline control. Conceptual incompatibilities[1] between the underlying employment and labor force data necessitate us to estimate the unemployment rate in a way that is unlikely to closely match historical rates. The direction and magnitude of the percentage change in unemployment is still however a valuable indicator of how policy simulations can be expected to affect the unemployment rate.

Historical unemployment data comes from the BLS Local Area Unemployment Statistics program.

The unemployment rate equation is shown below. It consists of the combination of five effects: (1) the region's unemployment rate in the previous year, (2) the current national unemployment rate, (3) the change in the national unemployment rate, (4) the relative change in real personal income per capita, and (5) the relative change in residence adjusted employment. The resulting rate from these five effects is additionally scaled by the ratio of actual[2] national unemployment to calculated national unemployment.

$$UR_t^k = \left[\beta_1 \cdot UR_{t-1}^k + \beta_2 \cdot UR_t^U + \beta_3 \cdot (UR_t^U - UR_{t-1}^U) + \beta_4 \cdot \left(\frac{\frac{RYP_t^k}{POP_t^k} - \frac{RYP_{t-1}^k}{POP_{t-1}^k}}{\frac{RYP_{t-1}^k}{POP_{t-1}^k}}\right) + \beta_5 \cdot \left(\frac{RAE_t^k - RAE_{t-1}^k}{RAE_{t-1}^k}\right) + C\right] \cdot \frac{UR_t^U}{UR_t^{UC}}$$

where:

UR_t^k = Region *k*'s Unemployment Rate at time *t*.

UR_t^U = The actual National Unemployment Rate at time *t*.

UR_t^{UC} = The National Unemployment Rate at time *t*, as calculated by our Unemployment Rate equation.

$\frac{RYP_t^k}{POP_t^k}$ = Region *k*'s Real Personal Income per Capita at time *t*.

RAE_t^k = Region *k*'s Residence Adjusted Employment at time *t*.

$\beta_1 = 0.8855264$ is the weight for the previous year effect.

$\beta_2 = 0.100795$ is the weight for the national effect.

$\beta_3 = 0.6533743$ is the weight for the national change effect.

$\beta_4 = 0.0011838$ is the weight for the change in Real Personal Income per Capita effect.

$\beta_5 = -8.244455$ is the weight for the change in Residence Adjusted Employment effect.

$C = 0.1568915$ is the Unemployment Rate Constant.

[1] Employment is a place of work concept, while labor force is counted by place of residence; therefore, we must adjust for commuter flows. Additionally, our employment definition counts jobs, not employed people, so we must account for multiple job holders.

[2] For history years, actual unemployment rates are historical values. For forecast years, values are from REMI's standard national forecast or from user-specified unemployment data included in a Macroeconomic Update.

Per Capita Income

Personal Income Results

The SEI module's per capita income results are calculated using the following methods to determine personal income, disposable income, and real disposable income. To determine these income measures per capita, total income in the region is divided by total population in the region.

Real Disposable Income Equations

Real disposable income *(RYD)* in the region equals personal income *(YP)* adjusted for taxes *(TAX)* and the PCE-Price Index, which represents the cost of living *(*$\bar{P}$*)*. Total personal income *(YP)* depends on compensation (*COMP)*, and proprietors' income *(YPI)*, property income *(YPROP)*, employee and self-employed contributions for government social insurance *(TWPER)*, employer contributions for government social insurance *(EGSI*), transfer payments *(V)*, and an adjustment to account for the difference between place-of-work and place-of-residence earnings *(RA)*.

$$RYD_t^k = \frac{\left(YP_t^k - TAX_t^k\right)}{\bar{P}_t^k}$$

$$YP_t^k = COMPT_t^k + YPIT_t^k + YPROP_t^k - TWPER_t^k - EGSI_t^k + V_t^k + RA_t^k$$

Total compensation, *COMPT*, is an aggregation of individual industry wages and salaries and supplements to wages and salaries. Thus,

$$COMPT_t^k = \sum_{i=1}^{ns}\left(E_{i,t}^k * CR_{i,t}^k + WBPVA_{i,t}^k + WSDAPV2_{i,t}^k + EPIFPVA_{i,t}^k + EGSIPVA_{i,t}^k\right)$$

where:

$COMPT_t^k =$ Total compensation aggregated across all industries.

$E_{i,t}^k =$ Employment in industry *i*.

$CR_{i,t}^k =$ The compensation rate of industry *i*.

$WBPVA_{i,t}^k =$ The policy variable for Wage and Salary Disbursements.

$WSDAPV2_{i,t}^k =$ The policy variable for Compensation.

$EPIFPVA_{i,t}^k =$ The policy variable for Employer Contributions for Employee Pension and Insurance Funds.

$EGSIPVA_{i,t}^k =$ The policy variable for Employer Contributions for Government Social Insurance.

The self-employed generate proprietors' income,

$$YPI_{i,t}^{k} = YLP_{i,t}^{k} - COMP_{i,t}^{k} + YPIPVA_{i,t}^{k}$$

where:

$YPI_{i,t}^{k} =$ Proprietors' income for industry *i*.

$YLP_{i,t}^{k} =$ Labor and proprietors' income for industry *i*.

$COMP_{i,t}^{k} =$ Compensation for industry *i*.

$YPIPVA_{i,t}^{k} =$ The policy variable for Proprietors' Income.

Total labor and proprietors' income, *YLP*, (also referred to as earnings by place of work) for all industries in the region can be calculated as

$$YLPT_{t}^{k} = \sum_{i=1}^{ns}\left(E_{i,t}^{k} * ER_{i,t}^{k} + WBPVA_{i,t}^{k} + WSDAPV2_{i,t}^{k} + EPIFPVA_{i,t}^{k} + EGSIPVA_{i,t}^{k}\right)$$

where:

$YLPT_{t}^{k} =$ Total labor and proprietors' income aggregated across all industries.

$E_{i,t}^{k} =$ Employment in industry *i*.

$ER_{i,t}^{k} =$ The earnings rate of industry *i*.

$WBPVA_{i,t}^{k} =$ The policy variable for Wage and Salary Disbursements.

$WSDAPV2_{i,t}^{k} =$ The policy variable for Compensation.

$EPIFPVA_{i,t}^{k} =$ The policy variable for Employer Contributions for Employee Pension and Insurance Funds.

$EGSIPVA_{i,t}^{k} =$ The policy variable for Employer Contributions for Government Social Insurance.

Wage and salary disbursements, *WSD*, are predicted as

$$WSDT_{t}^{k} = \sum_{i=1}^{ns}\left(E_{i,t}^{k} * WR_{i,t}^{k} + WBPVA_{i,t}^{k} + WSDAPV2_{i,t}^{k}\right)$$

where:

$WSDT_{t}^{k} =$ Total wage and salary disbursements aggregated across all industries.

$E_{i,t}^{k} =$ Employment in industry *i*.

$WR_{i,t}^{k} =$ The wage rate of industry *i*.

$WBPVA_{i,t}^{k} =$ The policy variable for Wage and Salary Disbursements.

Property income, *YPROP*, is split into its major components of Dividends *(YDIV)*, Interest *(YINT)*, and Rent *(YRENT)*, which each depend on the population and its age distribution, as well as historical regional differences in the type of property income received.

$$YDIV_t^k = \left(\lambda_{DIV,T}^k * YPROP\_POPPCT_t^k * YDIV_t^u\right) + YPRPOL_{j,t}^k \quad \text{(1-12a)}$$

$$YINT_t^k = \left(\lambda_{INT,T}^k * YPROP\_POPPCT_t^k * YINT_t^u\right) + YPRPOL_{j,t}^k \quad \text{(1-12b)}$$

$$YRENT_t^k = \left(\lambda_{RENT,T}^k * YPROP\_POPPCT_t^k * YRENT_t^u\right) + YPRPOL_{j,t}^k \quad \text{(1-12c)}$$

$$YPROP_t^k = YDIV_t^k + YINT_t^k + YRENT_t^k \quad \text{(1-12d)}$$

where:

$YDIV_t^k =$ Dividend income in region *k* for year *t*.

$\lambda_{DIV,T}^k =$ Adjustment for regional differences in dividend income based on the last history year.

$YPROP\_POPPCT_t^k =$ Age-weighted population in region *k* for year *t* as a share of the nation.

$YINT_t^k =$ Interest income in region *k* for year *t*.

$\lambda_{INT,T}^k =$ Adjustment for regional differences in interest income based on the last history year.

$YRENT_t^k =$ Rental income in region *k* for year *t*.

$\lambda_{RENT,T}^k =$ Adjustment for regional differences in rental income based on the last history year.

$YPROP_t^k =$ Total property income in region *k* for year *t*.

$YPRPOL_{j,t}^k =$ The policy variable for each type of Property Income.

and

$$NP_t^k = L65_t^k + m65^u * G65_t^k \quad \text{(1-13a)}$$

$$NP_t^u = L65_t^u + m65^u * G65_t^u \quad \text{(1-13b)}$$

$$YPROP\_POPPCT_t^k = \left(\frac{NP_t^k}{NP_t^u}\right) \quad \text{(1-13c)}$$

where NP_t^k is the age-weighted population in region *k* for year *t*, *m65* is the national ratio of per capita property income received for persons 65 years and older *(G65)* relative to property income received by persons younger than 65 *(L65)*, and $\lambda_{j,T}^k$ adjusts for regional differences and is calculated in the last historical year by solving equations (1-12) and (1-13).

Employee and self-employed contributions for government social insurance, *TWPER*, are predicted as

$$TWPER_t^k = \lambda_{TWPER,T}^k * WSDT_t^k * \left(\frac{TWPER_t^u}{WSDT_t^u}\right) + TWPPOL_t^k$$

where:

$\lambda_{TWPER,T}^k =$ Coefficient calculated in the last historical year to adjust for regional differences in the *TWPER* per dollar of wage and salary disbursements.

$WSDT =$ Total wage and salary disbursements.

$TWPPOL_t^k =$ The policy variable for Employee and Self-Employed Contributions for Government Social Insurance.

Employer contributions for government social insurance, *EGSI*, are predicted as

$$EGSI_t^k = \lambda_{EGSI,T}^k * WSDT_t^k * \left(\frac{EGSI_t^u}{WSDT_t^u}\right) + EGSIPVA_t^k$$

where:

$\lambda_{EGSI,T}^k =$ Coefficient calculated in the last historical year to adjust for regional differences in the *EGSI* per dollar of wage and salary disbursements.

$EGSIPVA_t^k =$ The policy variable for Employer Contributions for Government Social Insurance.

The residence adjustment, *RA*, is used to convert place-of-work income (compensation, proprietors' income, and contributions for government social insurance) to place-of-residence income. Residence adjustment is calculated as the net of the gross commuter flows in, *GI*, and the gross commuter flows out, *GO*.

$$RA_t^k = GI_t^k - GO_t^k + RAPOL_t^k$$

where:

$RAPOL_t^k =$ The policy variable for Residence Adjustment.

$$rs_t^{k,l} = \frac{LFA_t^l * \left[P_t^l * \frac{YP_t^l}{YD_t^l}\right]^{(1-\sigma)} * \left(D^{k,l}\right)^{-\beta}}{\sum_{k \neq l}^{n} LFA_t^j * \left[P_t^j * \frac{YP_t^j}{YD_t^j}\right]^{(1-\sigma)} * \left(D^{k,j}\right)^{-\beta}}$$

where:

$rs_t^{k,l} =$ The share of commuters who live in region *l* and work in region *k* in time period *t*.

$LFA_t^l =$ A geometrically declining moving average of the labor force in region *l* in time period *t*.

$P_t^l =$ The consumer price index including housing price in region *l* in time period *t*.

$YP_t^l =$ Total personal income in region *l* in time period *t*.

YD_t^l = Total disposable income in region *l* in time period *t*.

$D^{k,l}$ = The commute distance from region *l* to region *k*.

σ = Sigma value, the estimated parameter for consumer price.

β = Beta value, the estimated parameter for distance decay.

$$CI\_FLOW_t^{k,l} = \left(\sum_{k \neq l}^{n} rs_t^{k,l} * \left(COMPT_t^k - COMP_t^{nFM,k} - TWPER_t^k - EGSI_t^k\right)\right) + CommuterIncome\_PV_t^{k,l}$$

where:

$CI\_FLOW_t^{k,l}$ = The commuter income flow from commuters who live in region *l* and work in region *k* in time period *t*.

$CommuterIncome\_PV_t^{k,l}$ = The policy variable for commuter income flow from commuters who live in region *l* and work in region *k* in time period *t*.

$$GI_t^k = \sum_{k \neq l}^{n} CI\_FLOW_t^{l,k} + GROSSEARN\_PV_{IN,t}^k$$

where:

GI_t^k = Gross inflow of commuter dollars for residents of region *k* who work in all other areas.

$$GO_t^k = \sum_{k \neq l}^{n} CI\_FLOW_t^{k,l} + GROSSEARN\_PV_{OUT,t}^k$$

where:

GO_t^k = Gross outflow from region *k* to all other areas.

$GROSSEARN\_PV_{IN,t}^k$ = The policy variable for gross inflow of commuter earnings to region *k* from all other areas.

$GROSSEARN\_PV_{OUT,t}^k$ = The policy variable for gross outflow of commuter earnings from region *k* to all other areas.

Transfer payments by component, V_j, 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 who are not working. The components of transfer payments also are adjusted for historical regional differences.

$$V_{j,t}^{k} = \left(\lambda_{j,T}^{k} * V\_POPPCT_{m,t}^{k} * V_{j,t}^{u}\right) + VTRANSPOL_{j,t}^{k}$$

where:

> $\lambda_{j,T}^{k}$ = Adjustment for regional differences in individual components of transfer payments based on the last history year.
>
> $V\_POPPCT_{m,t}^{k}$ = Age-weighted population in region *k* for year *t* as a share of the nation (by four major types).
>
> $VTRANSPOL_{j,t}^{k}$ = The additive policy variable for individual components of Transfer Payments.
>
> V_t^k = Total transfer payments.

and

$$NV_{m,t}^{k} = VG_m^u * G65_t^k + VL_m^u\left[L65_t^k - EMPD_t^k\right] + \left[N_t^k - EMPD_t^k\right]$$
$$NV_{m,t}^{u} = VG_m^u * G65_t^u + VL_m^u[L65_t^u - EMPD_t^u] + [N_t^u - EMPD_t^u]$$
$$V\_POPPCT_{m,t}^{k} = \left(\frac{NV_{m,t}^{k}}{NV_{m,t}^{u}}\right)$$

where $NV_{m,t}^{k}$ is the age-weighted population in region *k* for year *t* (by four major types), *VG* are per capita transfer payments (by four major types) for persons 65 years and older relative to per capita transfer payments (by four major types) for all persons not working, *VL* are per capita transfer payments (by four major types) for persons younger than 65 who are not working, relative to per capita transfer payments for all persons not working (by four major types), $\lambda_{j,T}^{k}$ adjusts for regional differences and is calculated in the last historical year, and *EMPD* and *N* are, respectively, total employed (scaled from residence adjustment) and population in the region.

The variable *TAX* depends on net income after subtracting transfer income. It is adjusted for regional differences by λ_T^k and changes as national tax rates change.

$$TAX_t^k = \lambda_T^k * \left(YP_t^k - V_t^k\right) * \left[\frac{TAX_t^u}{(YP_t^u - V_t^u)}\right] + TPOL_t^k$$