“PWBM Microsimulation Model”

Penn Wharton Budget Model

April 11, 2022

The PWBM Microsimulation Model (Microsim) is a simulated population and economy on more than 60 demographic and economic attributes of the U.S. population, built using survey micro-data.

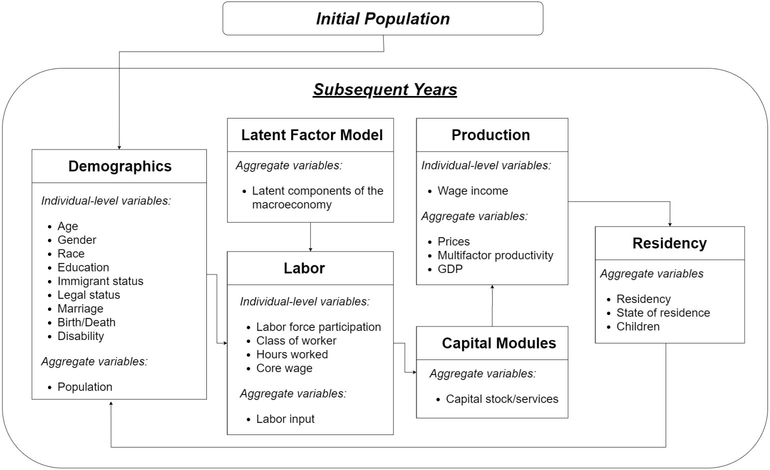
The motion of this population over time, beginning in the mid 1990s, is calibrated and validated to capture the momentum of demographic and economic forces in the United States.

Aggregating the population's labor and capital supplies and combining with productivity growth factors yields conventional estimates of several economic aggregates such as employment and output.

Continuing the simulation forward in time generates projections of alternative paths that the economy may take in the future given its current features and momentum.

These paths help to estimate the U.S. economic outlook and the range of uncertainty associated with it.

Figure 1: Concept Map of the Microsim



Some generic notation used throughout this document:

t	a year
i or j	an individual
`res`	a residency
f or F	a function
$P(E \mid C)$	probability of event E given conditions C
X_i	an array of demographic variables for individual i
Δx_t	growth of variable x from $t-1$ to t
$\overline{x}$	the mean of variable x
$\widehat{x}$	fitted values for variable x
$\tilde{x}$	observed (i.e., not simulated) values of x
$s(x)$	a spline on variable x

Table of Contents

Initial Population - Introduction

To construct an initial population, including all individual-level demographic and economic traits, the Microsim randomly picks residencies with replacement from a micro-survey data sample of U.S. households.

This "bootstrap" process may be set to collect between 100,000 and 1 million residencies from a pool that typically contains residencies from 5 annual surveys centered around a chosen initial simulation year.

Residencies are a cohabitating family concept consisting of, at minimum, a "head" individual, and might also include the head's spouse, children, and children's own spouses and children.

Within each bootstrapped residency, each individual enters the initial population with their observed demographic and economic attributes.

Initial Population - Overview and Key

Module Inputs	
Survey data for all micro-level variables in the start year, including:	
$\omega_i, \omega_{\texttt{res}}$	survey weight for an individual i or residency `res`
Intermediate Variables	
$n_{\texttt{res}}$	number of people in `res`
T_n	bootstrap target for size n
Module Outputs	
Simulated year 1 population and all micro-level variables. Main macro-level variables initialized as indices $= 1$.	

Initial Population Bootstrap

The bootstrap for the initial population proceeds in four steps:

1. Starting with survey data, groups residencies by size (# of members)
2. Each residency receives a weight $\omega_{\texttt{res}}$ equal to the average of its members' individual survey weights. I.e.,

$$\omega_{\texttt{res}} = \frac{1}{n_{\texttt{res}}} \sum_{i=1}^{n_{\texttt{res}}} \omega_i$$

where a residency `res` has n members $\{i_1, i_2, \ldots, i_n\}$.

Initial Population Bootstrap, cont.

3. The Microsim creates target shares T_n for each unique value of residency size $n \in \{1, 2, \ldots, \max\{n_{\texttt{res}}\}\}$. For each n,

$$T_n = \frac{1}{\sum_{n=1}^{\max\{n_{\texttt{res}}\}} \sum_{\{\texttt{res} | n_{\texttt{res}=n}\}} \omega_{\texttt{res}}} \cdot \frac{1}{n} \sum_{\{\texttt{res} | n_{\texttt{res}}=n\}} \omega_{\texttt{res}}$$

4. The Microsim then iterates through each potential residency size n—sizes for which $T_n < 0.05$ are treated as a single group. For each size, the Microsim draws a sample from residencies of that size in the survey data, with replacement, until the target T_n has been met. Each residency's probability of being chosen *for each draw* is proportional to its weight. I.e., for each value of n,

$$P(\texttt{res} \mid n_{\texttt{res}} = n) = \frac{\omega_{\texttt{res}}}{\sum_{\{\texttt{res} | n_{\texttt{res}}=n\}} \omega_{\texttt{res}}}$$

Initial Immigrant Bootstrap

The Microsim follows this same bootstrapping process to simulate foreign-born residencies of those who immigrated to the United States in the starting year.

These immigrant residencies are also drawn from a pool containing survey data surrounding the initial year.

For example, for an initial year 1996 the process pulls immigrant residencies who arrived in the United States anytime from 1994 to 1998 who might appear in survey data from 1992 to 2000.

This same bootstrap process is used in subsequent years.

Subsequent Years

Each year after the start, the Microsim implements transitions for each individual or family attribute through various modules.

Each variable in the simulation is controlled by a single module (though some modules control multiple variables).

For example, the "Age" module increases each individual's age by 1 each year.

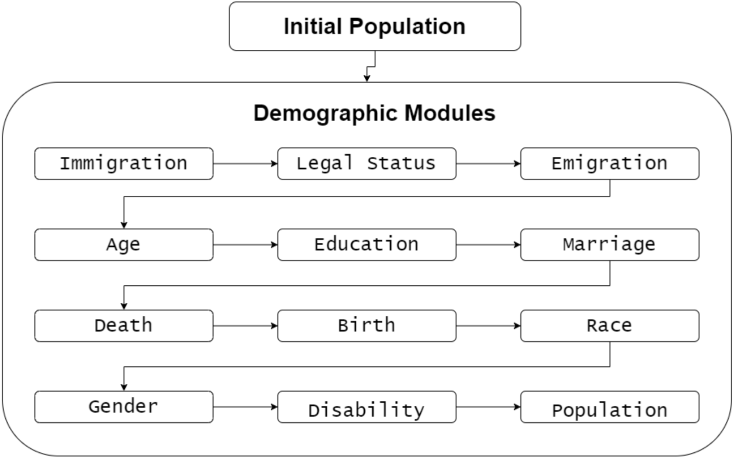
Table of Contents

Demographics - Overview and Key

Inputs	
Previous-year demographic variables	
Intermediate Variables	
pb	previous birth(s)
race+	race with distinct category for Mexican immigrants
Outputs	
a	age
imm	immigrant
e	education level
r	race
g	gender
m	married
$\texttt{add\_child}_t$	whether a woman gives birth to a child this year
pop	total population

Demographics - Run Order

Figure 2: Run Order for Demographic Modules



Gross Lawful Immigration

Lawful and unlawful immigration rates for each year t are estimated from historical data. Groww lawful immigration is estimated via:

$$\begin{aligned}\texttt{gross\_lawful}_t = (\texttt{n\_lawful}_t - \texttt{n\_lawful}_{t-1}) \\ - \texttt{emigration\_lawful}_{t-1} \\ - \texttt{deaths\_lawful}_{t-1} \\ - \texttt{naturalizations}_{t-1} \\ - \texttt{visa\_overstays}_{t-1}\end{aligned}$$

where `n_lawful` is the number of lawful noncitizens,[1] the variable `emigration_lawful` is total emigration of lawful noncitizens, and `deaths_lawful` is total deaths of lawful noncitizens.

[1] Comprising lawful permanent residents and nonimmigrants.

Gross Illegal Immigration

Similarly, gross illegal immigration is estimated as:

$$\begin{aligned}\texttt{gross\_illegal}_t = (\texttt{n\_unauth}_t - \texttt{n\_unauth}_{t-1}) \\ - \texttt{emigration\_unauth}_{t-1} \\ - \texttt{deaths\_unauth}_{t-1} \\ + \texttt{visa\_overstays}_{t-1}\end{aligned}$$

where `n_unauth` is the number of unauthorized immigrants, `emigration_unauth` is total emigration of unauthorized immigrants,[2] and `deaths_unauth` is total deaths of unauthorized immigrants.

Future gross legal and illegal immigration flows are projected based on recent trends and anticipated changes in policy.

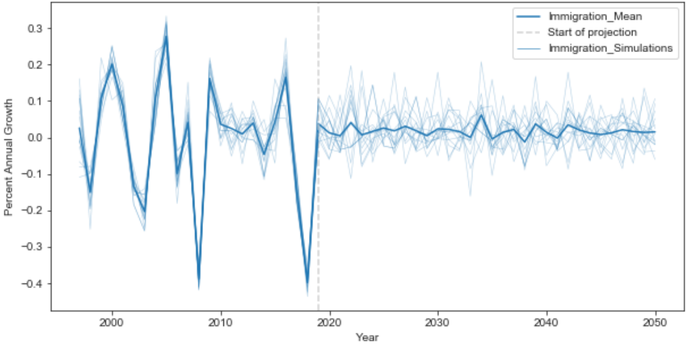
[2]Including removals.

New Immigrant Bootstrap

New immigrants are bootstrapped via the same process described in section initial immigrant bootstrap in order to hit the estimated target rates for each year.

For projection years, the bootstrap draws from the latest valid pool of potential immigrants.

Figure 3: Simulated Gross Immigration by Year



Legal status of foreign-born persons is assigned in survey microdata based on observable attributes. Broadly:

1. *Naturalized citizens* are identified primarily through a questions about citizenship status, with some adjustments for overreporting.
2. *Nonimmigrants* on temporary visas are identified by matching their characteristics such as occupation, year of arrival, and family relationships.
3. *Lawful permanent residents* (LPRs) are identified probabilistically based on country of origin, state of residence, age, gender, marital status, and occupation.
4. *Unauthorized immigrants* are identified probabilistically based on characteristics related to legal status (e.g. receipt of government benefits), country of origin, and state of residence.

Finally, foreign-born people not initially assigned to other categories are classed as nonimmigrants.

Naturalization

Each year, each person at least 18 and either a lawful permanent resident (LPR) for 5 years OR LPR for 3 years and married to a US citizen has a chance to naturalize.

Naturalization rates by country of origin, age, marital status, and gender are estimated based on the number of naturalizations each year from Department of Homeland Security data. The eligible population each year is based on the legal status assignment described above.

Figure 4: Simulated Naturalization Rates by Year, Gender, and Marital Status

Visa Overstays

A visa overstay occurs when a nonimmigrant remains in the U.S. past the date authorized by their entry visa.

Nonimmigrants who overstay short-term travel visas enter the model as new unauthorized immigrants.

Nonimmigrants on longer-term visas who are at least 18 years old, not married to a U.S. citizen, and have resided in the United States at least 1 year have a chance to overstay their visa and become unauthorized immigrants.

Visa overstay rates by visa type and country of origin are estimated based on the number of visa overstays each year from Department of Homeland Security data.

Figure 5: Simulated Visa Overstay Rates by Year, Gender, and Marital Status

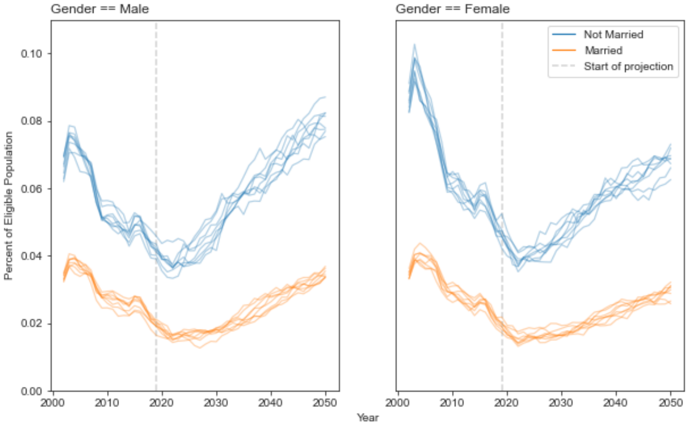
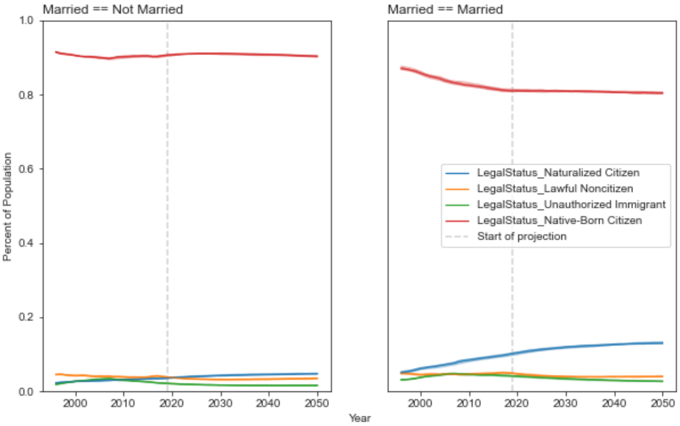


Figure 6: Legal Status Shares by Year and Marital Status



Emigration

Historical emigration rates are based on several sources:

- Emigration by age, gender, and county of origin from Van Hook et al (2006).
- Emigration of lawful immigrants from Schwabish (2011).
- Emigration of unauthorized immigrants from Warren and Warren (2013).
- Information on removals and adjustments of status of unauthorized immigrants from the Department of Homeland Security.
- Information on removals of unauthorized immigrants from the Transactional Records Access Clearinghouse.

These sources are combined with estimates of the foreign-born population to estimate emigration probabilities by legal status, country of origin, age, gender, and years of residence in U.S.

Education

The Microsim tracks both *total years of schooling*, ranging from 0 to 18, and *highest degree attained*, ranging from "less than high school" to "advanced degree".

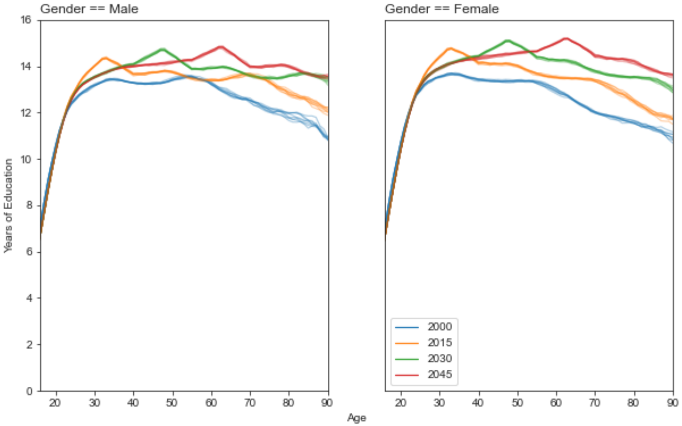
Only individuals ages 6 through 89 can add years of education in the Microsim.

The Microsim population is split into 10 subgroups by gender and race. For each year in the historical sample, we estimate a probability distribution function $f(t, a_t, g_t)$ and a corresponding cumulative distribution function $F(e_t \mid t, a_t, g_t)$.

In a year t, each individual's chance of adding a year of education is a function P estimated by:

$$P(e_{t+1} = e_t + 1 \mid t, a_t, g_t) = \frac{F(e_t \mid t, a_t, g_t) - F(e_{t+1} \mid t+1, a_{t+1}, g_{t+1})}{f(e_t \mid t, a_t, g_t)}$$

Figure 7: Simulated Education Population Shares by Age, Gender, and Year



Marriage - Encounters

Each year of the simulation, all single women ages 16-84 enter the marriage market and meet up to 9 potential male marriage-eligible partners per year.

With 50 percent probability, the partner is drawn from the pool of eligible men of the same race as the woman seeking to marry. And with the remaining 50 percent likelihood, the partner is drawn from the entire population of eligible males, including the race of the female in question.

At each meeting, the woman either rejects the man, moving on to her next date, or marries the man and removes herself from the marriage market. For a woman i and male j, the probability of acceptance conditional on a date in year t is given by:

$$P(accept_{i,j} \mid date_{i,j}) = F(t, r_i, e_i, a_i, r_j, e_j, a_j - a_i)$$

estimated via logit regression.

Divorce

Each year, each married couple (a woman i and man j) has a chance of divorce in year t given by,

$$P(divorce_{i,j} \mid married_{i,j}) = F(t, r_i, e_i, a_i, r_j, e_j)$$

estimated via logit regression.

Marriages may also end due to the death of a spouse, in which case the surviving spouse enters the marriage market in the following year.

Final marriage and divorce rates are calibrated over a large number of Microsim runs to ensure accuracy. Rates are algorithmically adjusted between runs to achieve a final set of targets.

Marriage and Divorce Calibration - Categories

The Microsim's marriage and divorce rates are further calibrated along several categories:

1. All females ages 18-84
2. All females, by:
   - Each of 12 age groups between 16 and 85
   - Education categories
   - Race categories
   - Immigrant Status categories
3. All males by each of the sub-categories under (2)
4. Those in interracial marriages

The calibration process iterates through this list of categories twelve times to produce the final marriage and divorce rates.

Marriage and Divorce Calibration - Process

In each iteration of the calibration process, the Microsim generates a simulated population using the current marriage and divorce rates.

For each of the listed categories ς the Microsim calculates:

$$\mathtt{ratio}_{\varsigma} = \frac{1}{9} \sum_{t=2005}^{2013} \frac{\mathtt{m\_rate}(\mathtt{SIM}_{t\varsigma})}{\mathtt{m\_rate}(\mathtt{CPS}_{t\varsigma})}$$

where $\mathtt{m\_rate}(\mathtt{SIM})$ is the marriage rate calculated from the simulated data and $\mathtt{m\_rate}(\mathtt{CPS})$ is calculated from survey data.

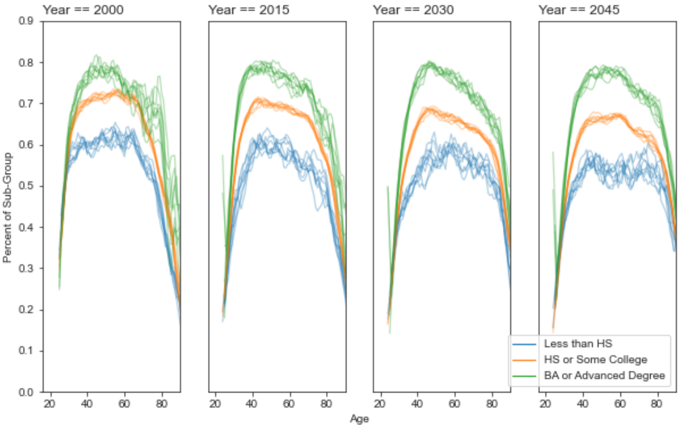
Finally, the Microsim updates its acceptance and divorce rates:

$$\mathtt{acceptance\_updated}_{\varsigma} = \mathtt{acceptance\_original}_{\varsigma} \sqrt{1/\mathtt{ratio}_{\varsigma}}$$

$$\mathtt{divorce\_updated}_{\varsigma} = \mathtt{divorce\_original}_{\varsigma} \sqrt{\mathtt{ratio}_{\varsigma}}$$

and repeats this process twelve times.

Figure 8: Simulated Married Population Share by Year and Education



Death

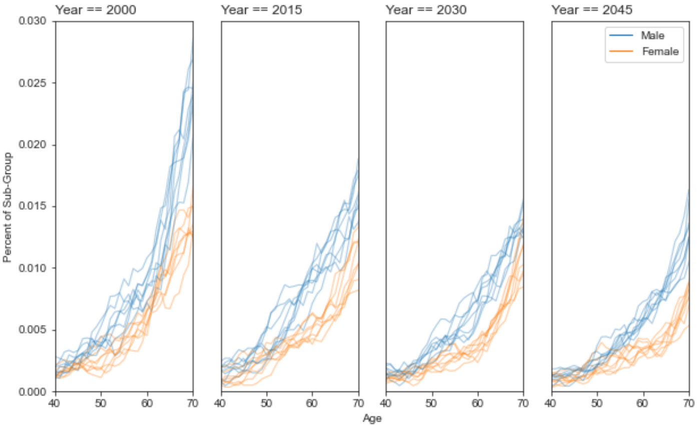
Each individual's chance of death in a given year is estimated via a weighted logit model:

$$P(\texttt{death}_t) = F(t, a_t, a_t^2, [a_t < 2], m_t, e_t, r_t, g_t)$$

As the Microsim draws from actual death records, in certain small demographic subgroups and years there may be more recorded deaths than there are simulated individuals.

The Microsim distributes these death records across surrounding years.

Figure 9: Simulated Mortality Rates By Age, Year, and Gender



Births (Married)

Each year, each woman in the the Microsim has some probability of having a child. This probability distribution is modeled separately for single and married adult women (ages 20-45) and for teens (ages 19 and under).

Each year, each married woman's chance of having a new child is given by:

$$P(\texttt{add\_child}_{it} \mid m_{it} = 1, a_{it} \geq 20) = \ldots$$
$$F(t, a_{it}, e_{it}, r_{it}, \texttt{pb}_{it}, \texttt{imm}_{it}, \texttt{imm}_{it}, \texttt{initial\_decade})$$

estimated via a weighted logit regression, where `initial_decade` is an indicator for years in the first decade of the simulation.

Births (Single)

Each year, each single adult woman's chance of having a new child is given by:

$$P(\texttt{add\_child}_t \mid m_t = 0, a_{it} \geq 20) = \ldots$$
$$F(t, a_{it}, e_{it}, r_{it}, \texttt{pb}_{it}, [\texttt{pb}_{it} = 0], \texttt{imm}_{it}, \texttt{initial\_decade})$$

estimated via a weighted logit regression.

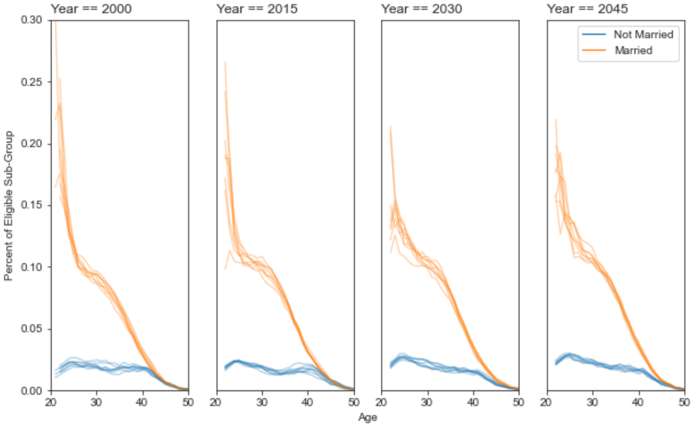
Births (Teen)

Each year, each teen woman's chance of having a new child is given by:

$$P(\texttt{add\_child}_{it} \mid a_{it} < 20) = F(t, a_{it}, m_{it}, \texttt{race+}_{it})$$

estimated via a weighted logit regression.

Figure 10: Simulated Birth Rates by Year, Mother's Marital Status and Age



Gender and Race

In the simulated population, each individual's gender and race do not change over time.

Additionally, the gender and race of new immigrants is determined by the bootstrapping process.

Each newborn's gender is randomly assigned, with

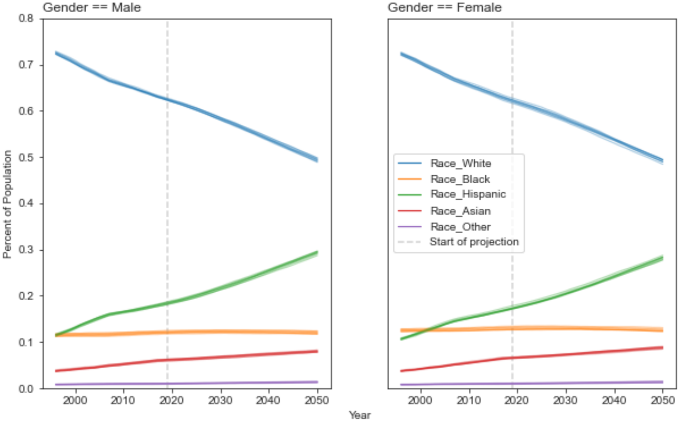
$$P(g_i = \mathit{female}) \approx 0.512$$

and

$$P(g_i = \mathit{male}) = 1 - P(g_i = \mathit{female}) \approx 0.488$$

Similarly, newborns have equal odds of inheriting the race of each of their parents (children born to single mothers inherit the race of their mother).

Figure 11: Simulated Race and Gender Population Shares by Year



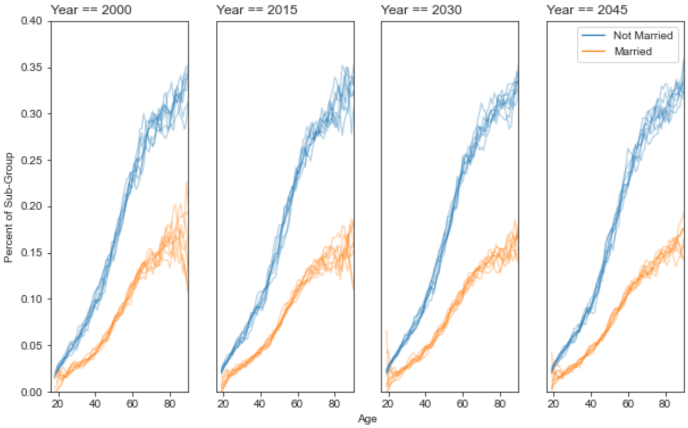
Disability

Disability is estimated through a logit regression:

$$P(d_{it}) = F(s(a_{it}), d_{i,t-1}, r_{it}, m_{it})$$

where $s(a_{it})$ is a cubic spline on age.

Figure 12: Simulated Disability Rates by Age, Year, and Marital Status



Each year, the change in the total population is the sum of estimated immigration and births minus emigration and deaths:

Figure 13: Simulated Population by Year

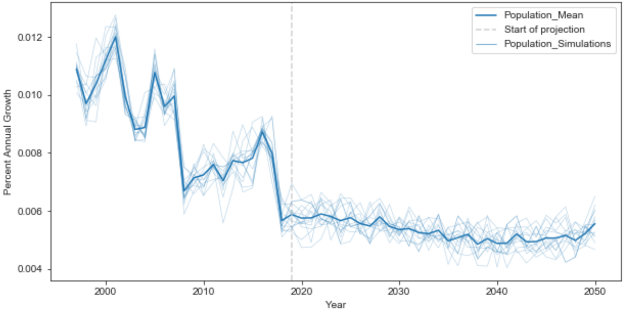


Table of Contents

For use in later modules, we estimate a dynamic factor model on historical data, based on Stock and Watson (2016).

This model estimates the first eight principal components of 95 dis-aggregated macroeconomic data series which include various measures of production (e.g., real GDP, industrial production), prices (e.g., PCE, CPI-U, PPI), housing starts, labor markets (e.g., unemployment, payrolls), etc.

Latent Factor Model - Key and Overview

Inputs	
Γ	an array of 95 macroeconomic data series
Intermediate Variables	
λ, Ψ	regression coefficients
PC	vector of principal components
ε, ϵ	regression error terms
Outputs	
PC	the principal components of Γ

Latent Factor Model (cont.)

For our vector Γ_t of macroeconomic variables, the dynamic factor model is represented by:

$$\Gamma_t = \lambda C_t + \varepsilon_t$$

where λ is a coefficient matrix for our vector of principal components C_t and ε_t represents an idiosyncratic component term.

The principal components themselves are modeled as first-order autoregressive processes:

$$PC_t = \Psi PC_{t-1} + \epsilon_t$$

Latent Factors Over Time

Figure 14: Estimated Latent Factors by Year

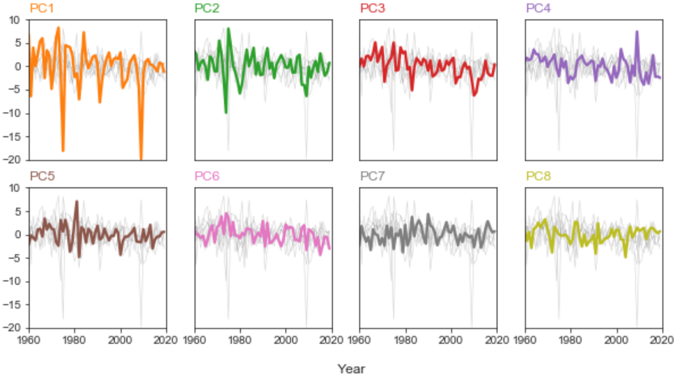


Table of Contents

Labor - Introduction

The Microsim classifies workers along three dimensions:

1. Labor force participation (`worker`)
2. Employee vs. self-employed and private vs. government (`class`)
3. Part-time vs. full-time and part-year vs. full-year (`status`)

The Microsim then estimates each worker's hours worked h and "core" wage z, i.e., the portion of wages explained by individual characteristics (as opposed to macroeconomic factors).

Finally, the Microsim estimates growth in overall labor input L.

Labor - Key and Overview I

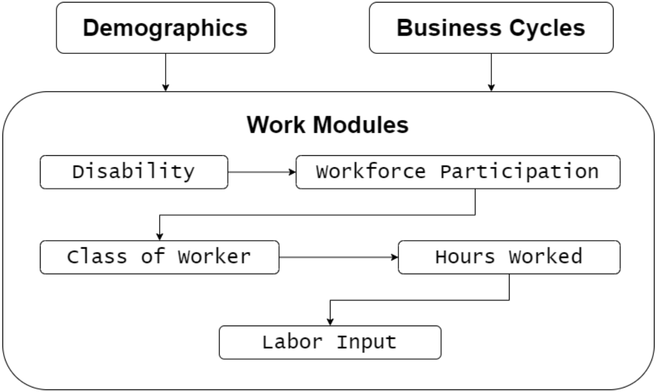
Inputs	
a	age
r	race
g	gender
m	married
e	education level
$\mathtt{PC}$	the principal components of Γ

Intermediate Variables	
β	wage regression coefficients for each year
μ	core wage shock
Φ	core wage shock persistence
ν	transitory core wage shock

Labor - Key and Overview II

Outputs	
d	work disability
`worker`	workforce participation
`class`	employer type (govt or private, self-employed or not)
`status`	work status (part- or full-time, part- or full-year)
h	hours worked
z	core wage
L	labor input

Figure 15: Run Order for Work Modules

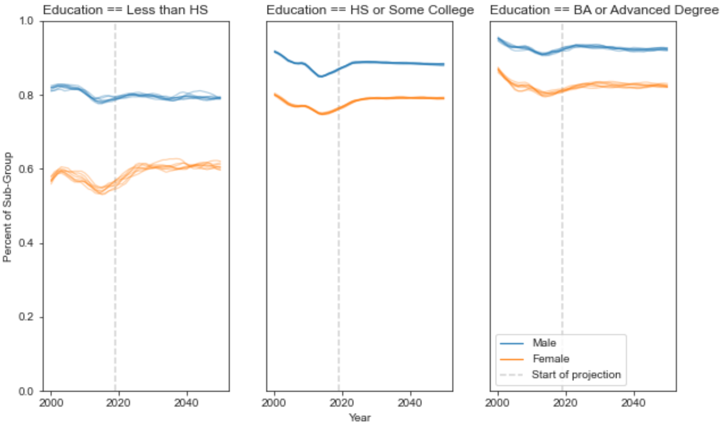


Individual labor force participation, `worker`, is estimated via classification tree with the following predictors:

$$\begin{aligned}\texttt{worker}_{it} = F(&\texttt{worker}_{i,t-1}, \texttt{class}_{i,t-1}, a_{it}, g_{it}, r_{it}, e_{it}, d_{it}, m_{it}, \ldots \\ &\texttt{has\_children}_{it}, \texttt{legal\_status}_{it}, \ldots \\ &\texttt{years\_of\_residence}_{t}, \mathrm{PC}_{t,t-1,t-2})\end{aligned}$$

where the vector `PC` is the macroeconomic principal components estimated in section 3.

Figure 16: Simulated Percent Working by Gender, Education, and Year

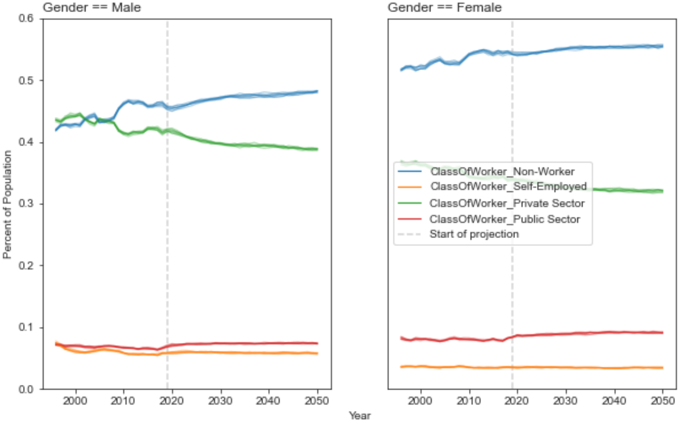


Similarly, we want to classify workers along two dimensions: employed vs. self-employed and private vs. government. This is also estimated via classification tree using the same predictors as for `worker`:

$$\begin{aligned}\texttt{class}_{it} = F(&\texttt{worker}_{i,t-1}, \texttt{class}_{i,t-1}, a_{it}, g_{it}, r_{it}, e_{it}, d_{it}, m_{it}, \ldots \\ &\texttt{has\_children}_{it}, \texttt{legal\_status}_{it}, \ldots \\ &\texttt{years\_of\_residence}_{it}, \mathrm{PC}_{t,t-1,t-2})\end{aligned}$$

where the vector `PC` is the macroeconomic principal components estimated in section 3.

Figure 17: Simulated Worker Class Shares by Year and Gender



Hours Worked

Workers are classified into full- or part-time and full- or part-year via classification tree:

$$\begin{aligned} \texttt{status}_{it} = F(&\texttt{status}_{i,t-1}, \texttt{class}_{i,t-1}, \ldots \\ &a_{it}, g_{it}, r_{it}, e_{it}, d_{it}, m_{it}, \texttt{has\_children}_{it}, \texttt{PC}_{t,t-1,t-2}, \ldots \\ &\texttt{legal\_status}_{it}, \texttt{years\_of\_residence}_{it}) \end{aligned}$$

where the vector PC is the macroeconomic principal components estimated in section 3.

Subsequently, hours worked h_{it} is estimated via regression tree:

$$\begin{aligned} h_{it} = F(&\texttt{status}_{it}, \texttt{class}_{it}, \ldots \\ &a_{it}, g_{it}, r_{it}, e_{it}, d_{it}, m_{it}, \texttt{has\_children}_{it}, \texttt{PC}_{t,t-1,t-2}, \ldots \\ &\texttt{legal\_status}_{it}, \texttt{years\_of\_residence}_{it}) \end{aligned}$$

Figure 18: Simulated Aggregate Hours Worked by Year

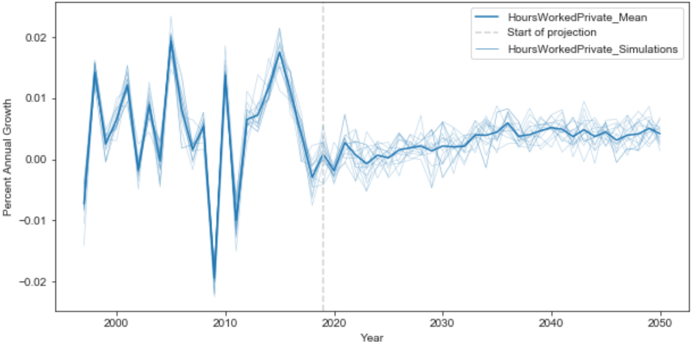
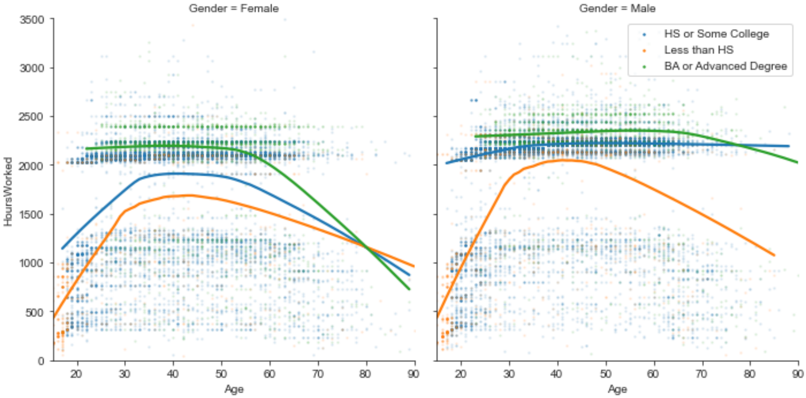
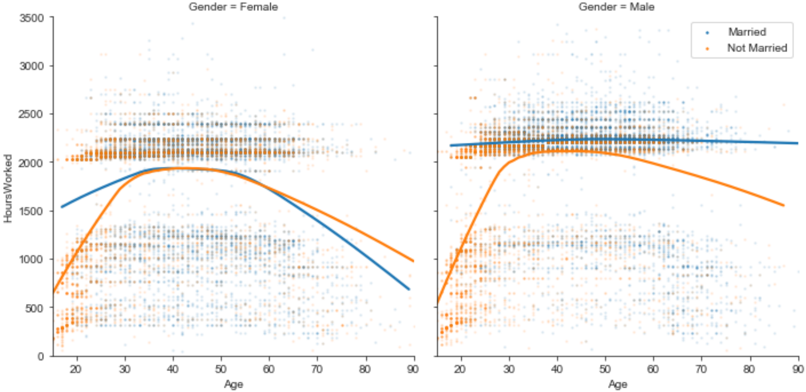


Figure 19: Simulated Hours Worked for Workers, by Age, Gender, and Education



Shown with LOWESS lines for each sub-group.

Figure 20: Simulated Hours Worked for Workers, by Age, Gender, and Marital Status



Shown with LOWESS lines for each sub-group.

Core Wage

Wages are estimated via the "core" wage, i.e., the portion of hourly wages explained by a worker's demographic attributes.

The core wage consists of persistent (across years) and transitory components.

These components are then estimated separately, allowing us to recompose and project core wages and nominal wage income for all workers.

Later, on slide 91, the core wage will be combined with information about the macroeconomy to get nominal hourly wages.

Historical Wage Regressions

Estimation of the historical wage regressions starts with survey data, restricted to workers whose primary employment was as private wage and salary workers.[3]

Workers with extreme hourly wages are removed, i.e., anyone with a normalized hourly wage of less than \$1 or greater than \$500 (about 0.3-0.4% of the sample each year).[4]

[3] This historical estimation includes employees of businesses and nonprofits, but excludes primarily self-employed and government employees. All simulated workers, however, will have wages assigned through the same process.

[4] Normalized wages for each year equal the actual wage in the last sample year inflated by cumulative growth in the average hourly wage.

We then run a regression on the hourly wage w as a function of demographic variables X:

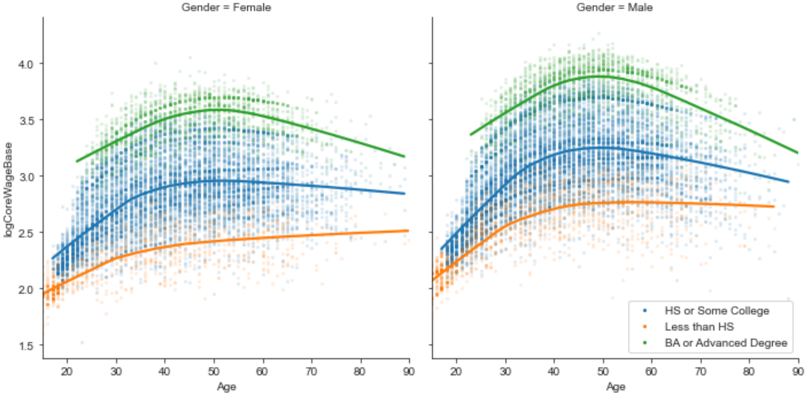
$$w_{it} = \beta_t X_{it} + \varepsilon = F(a_{it}, g_{it}, e_{it}, r_{it}, m_{it}, d_{it}, \texttt{class}_{it}, \texttt{children}_{it})$$

Additional predictors are only included in the regression when available in later years:

- After 1992, X_{it} also includes indicators of origin from various Hispanic countries.
- After 1993, X_{it} also includes additional information on immigrant status, years of residence, and birthplace.
- After 1994, X_{it} also includes legal status.

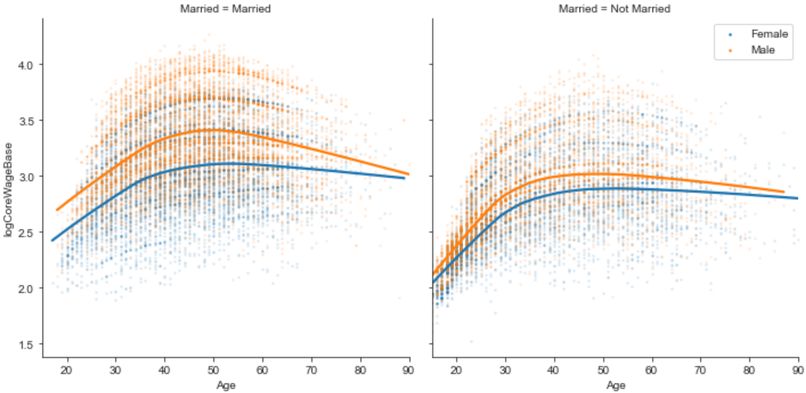
Simulated Core Wage Base

Figure 21: Simulated Core Wage Base by Age, Gender, and Education



Shown with LOWESS lines for each sub-group.

Figure 22: Simulated Core Wage Base by Age, Gender, and Marital Status



Shown with LOWESS lines for each sub-group.

Calculation of Core Wage Shocks

Once the core wage fitted values $\widehat{z} = \beta X_{it}$ are estimated, the historical core wage shocks μ are calculated directly:

$$\mu_{it} = \tilde{z}_{it} \left(\frac{1}{\widehat{z_{it}}} \right) = \frac{\tilde{w}_{it}}{\gamma_t} \left(\frac{1}{\widehat{z_{it}}} \right)$$

where $\tilde{z}$ and $\tilde{w}$ are observed core wages and wages, respectively, and γ is a macroeconomic deflator discussed on slide 90.

μ is then further decomposed into a persistent component and a transitory shock.

Persistence of Core Wage Shocks

The persistence of wage shocks is estimated as an autoregressive process.

For each year after 1996 and each value in age, we estimate this process through a rolling 5-year, age-centered, weighted regression. I.e., for each age a, we subset to the sample of $\{a, a \pm 1, a \pm 2\}$ and estimate the regression:

$$\log(\mu_{it}) = \Phi_a \mu_{i,t-1} + \varepsilon_{it}$$

weighted by hours worked h_{it}. Weights for $a \pm 1$ are scaled by $\frac{2}{3}$ while weights for $a \pm 2$ are scaled by $\frac{1}{3}$.

Transitory Core Wage Shocks

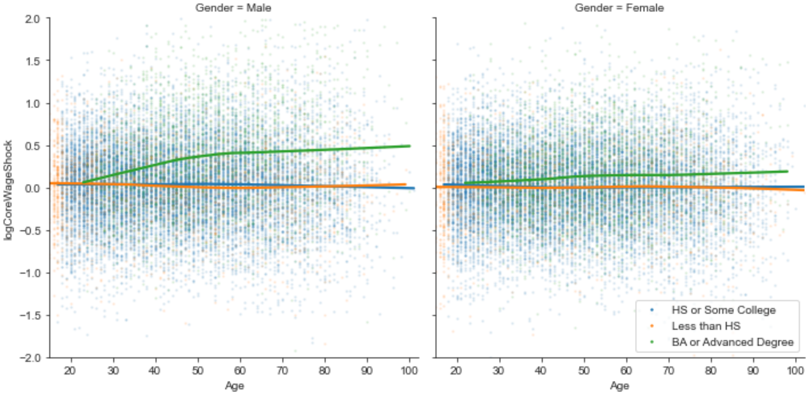
Using the core wage shock persistance regression, we calculate fitted core wage shocks $\widehat{\mu}_{it}$. Transitory core wage shocks for each individual i and year t the difference between fitted and actual shocks:

$$\nu_{it} = \widehat{\mu}_{it} - \mu_{it}$$

The largest 1 percent and smallest 1 percent of these transitory shocks are dropped. The remaining transitory shocks form a pool for the Microsim to draw from at each age a.

The probability of each transitory shock in this pool being chosen is proportional to its weight (hours worked, h_{it}) from the wage shock regression.

Figure 23: Simulated Core Wage Shocks by Age, Gender, and Education



Shown with LOWESS lines for each sub-group.

Core Wage Projections

Now, using the estimated wage regression coefficients β, persistence of shocks Φ, and pool of transitory shocks, the Microsim projects core wages z.

For each individual i and year t, we draw a transitory shock ν_{it} from the pool for age a_{it}. Then their core wage for that year is given by:

$$z_{it} = \beta_t X_{it} + \Phi_{a_{it}} \mu_{i,t-1} + \nu_{it}$$

The macroeconomic variable labor input L is an index, initialized in the start year with value $= 1$.

We estimate the growth of this index using the estimated coefficients β from the core wage regressions discussed previously:

$$w_{it} = \beta_t X_{it} + \varepsilon_{it}$$

Labor Input (cont.)

For years in the historical sample, we estimate the growth of L via:

$$\Delta L_t = \sqrt{\frac{\beta_t X_{it}}{\beta_t X_{i,t-1}} \cdot \frac{\beta_{t-1} X_{it}}{\beta_{t-1} X_{i,t-1}}}$$

i.e., a Fisher index growth based on lagged rates of return β_{t-1} and growth based on current-year rates of return β_t.

For years outside the historical sample, the coefficients $\beta_t = \beta_{t-1}$. So for projection years t:

$$\Delta L_t = \frac{\beta_t X_{it}}{\beta_t X_{i,t-1}}$$

Figure 24: Estimated Labor Input by Year

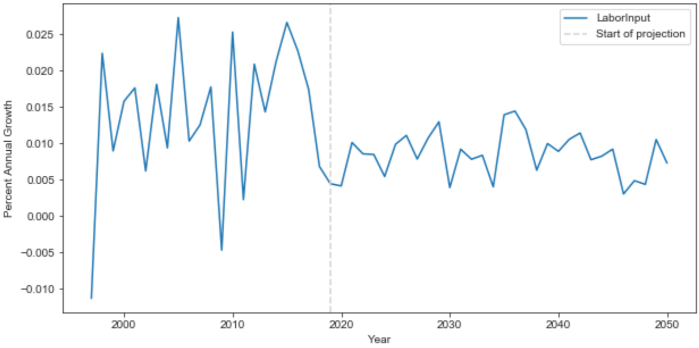


Table of Contents

With labor input determined, the Microsim makes projections for other important private sector macroeconomic aggregates, including prices, capital services, productivity, and production.

As mentioned in section 1, these macroeconomic variables are indices initialized in the start year with value $= 1$.

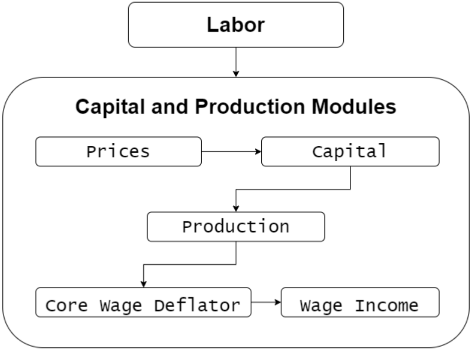
As with labor input, each of these modules includes separate processes for historical years and projection years. The separate, more detailed historical process allows for more accurate parameter calibration.

Finally, these macroeconomic variables are used to

Capital and Production - Key and Overview

Inputs	
L	labor input
z	core wage
Intermediate Variables and Parameters	
α	capital share of production
η	capital elasticity
γ	core wage deflator
Outputs	
P	price index
K	capital services
A	multifactor productivity
Y	production
w	hourly wage
inc	wage income

Figure 25: Run Order for Capital and Production Modules

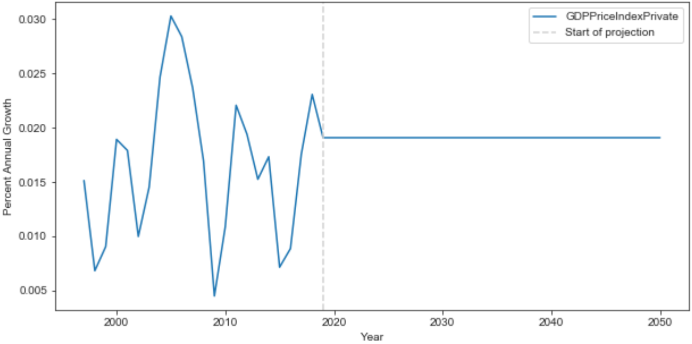


Prices

For historical years, the Microsim tracks two price indices: a total GDP price index P_{total} and a specific price index for the private sector (businesses and non-profits), P.

For projection years, the Microsim only tracks private GDP. Inflation ΔP in each projection year t is set to its observed historical sample mean.

Figure 26: Estimated Inflation by Year



Capital

Growth in capital services is calculated as:

$$\Delta K_t = \Delta L_t \cdot \overline{\Delta K / \Delta L}$$

For projection years, $\overline{\Delta K / \Delta L}$ is the historical observed mean growth of the capital-labor ratio.

For historical sample years, the capital elasticity $\eta_t = \frac{1}{2}(\alpha_t + \alpha_{t-1})$, where α is the capital share. For projection years, η equals its mean value on the historical sample.

Capital Services Over Time

Figure 27: Estimated Capital Services by Year

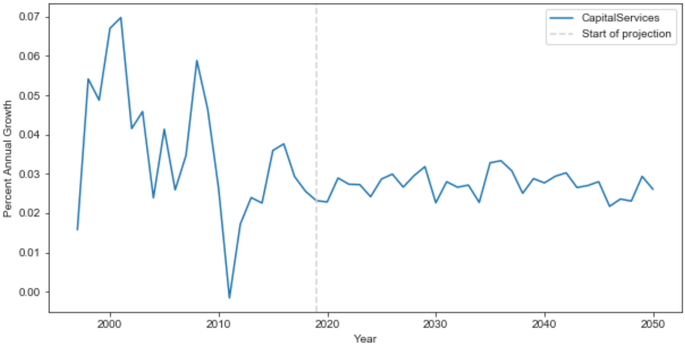
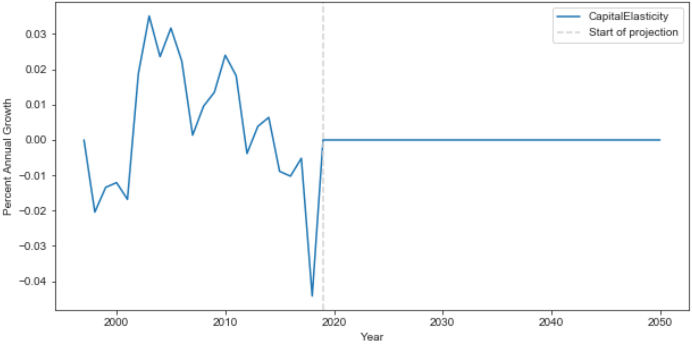


Figure 28: Estimated Capital Elasticity by Year



Multifactor Productivity

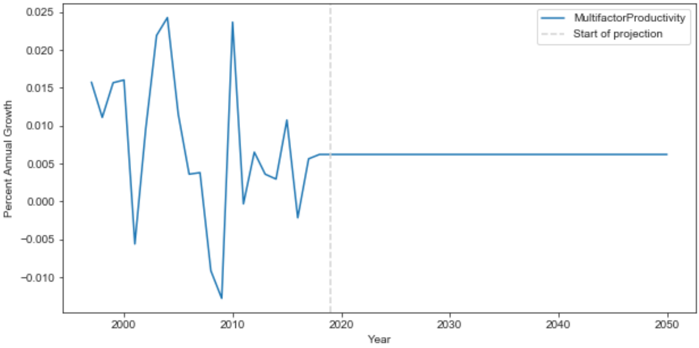
For the historical sample, the growth of multifactor productivity A is estimated via:

$$\Delta A_t = \frac{\Delta Y_{private,t}}{(\Delta K_t)^{(\eta_t)}(\Delta L_t)^{(1-\eta_t)}}$$

For projection years, ΔA_t is set to its historical sample mean value.

Multifactor Productivity Over Time

Figure 29: Estimated Multifactor Productivity by Year



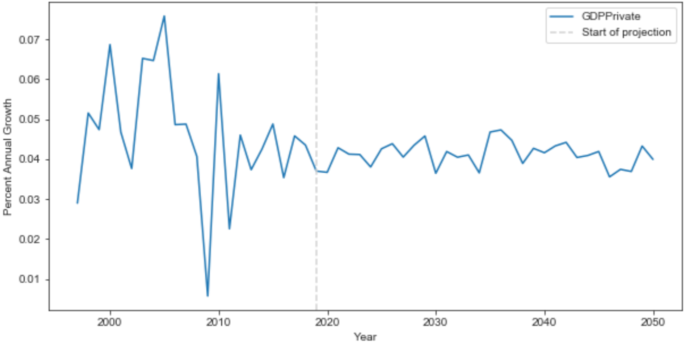
Production

For projection years, the Microsim only tracks private nominal production (nonprofit and business sector).

Growth in nominal private GDP is calculated by:

$$\Delta Y_t = \Delta P_t \Big[\Delta A_t (\Delta K_t)^{\eta_t} (\Delta L_t)^{(1-\eta_t)} \Big]$$

Figure 30: Estimated GDP by Year



Core Wage Deflator

The difference between the actual wage and the core wage estimated on slide 73 is represented by the "core wage deflator" γ_t that accounts for macroeconomic conditions.

The growth of this deflator is estimated via:

$$\Delta\gamma_t = (\Delta P_t)(\Delta A_t)(\Delta K_t)^{(\eta_t)}$$

where ΔP is inflation, A is multifactor productivity, K is capital services, and η is the capital elasticity.

Nominal Wages and Wage Income

Using core wages z, estimated in section 73, nominal hourly wages w are given by:

$$w_{it} = \gamma_t z_{it}$$

and so total wage income is given by:

$$\texttt{inc}_{it} = w_{it} h_{it}$$

where h is hours worked, estimated in section 60.

Table of Contents

Residency - Introduction

As stated previously, residencies consist of, at minimum, a "head" individual. They might also include the head's spouse, children, and children's own spouses and children.

The Microsim tracks the formation of new residencies, the movement of individuals between residencies, and the movement of residencies between U.S. states.

Residency - Overview and Key

Inputs	
a	age
imm	immigrant
e	education level
r	race
g	gender
inc	wage income
Intermediate Variables	
moving	whether an individual leaves to form a new residency
Outputs	
res	residency I.D.
children	whether an adult has a child of age < 18 in their household
state	U.S. state of residency

Residency - Run Order

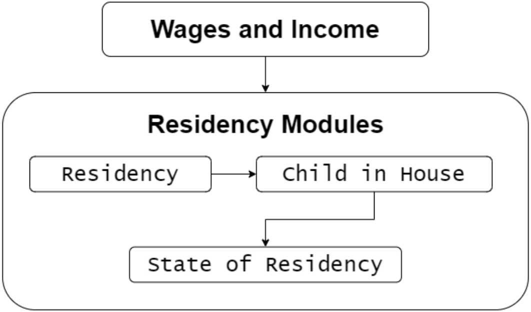


Figure 31: Run Order for Residency Modules

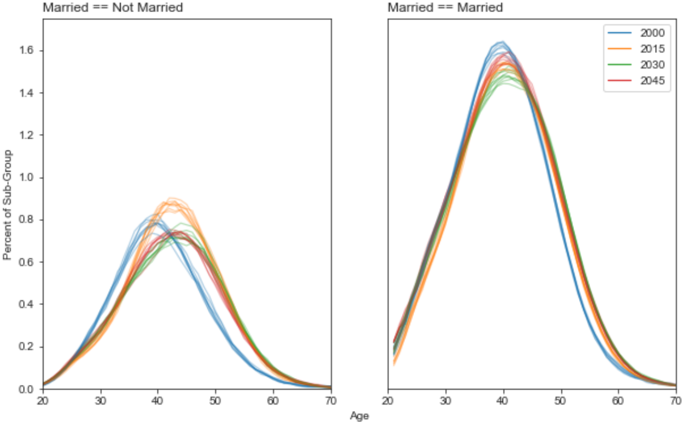
Starting at age 16, each year each child has a chance of moving away and becoming the head of their own residency. A logit regression on survey data yields the conditional probability of moving away in each year:

$$P(\mathtt{moving_t}) = F(s(a_t), e_t, r_t, g_t, \mathtt{imm}_t, \mathtt{inc}_t, t)$$

where $s(a_t)$ is a spline of degree 2 on age.

Compared to traditional models in which dependents automatically move out at 18, this better captures the distribution of adults who still live with their parents.

Figure 32: Simulated Shares With Children in Residence by Age and Year



Family Formation - Other Movements

The Microsim also tracks other potential family movements:

- At marriage, one newlywed in the couple is randomly chosen to join the other's residence.
- For divorcees, one spouse is randomly chosen to split off into a new residence.
- Other special cases: divorcees living with one of the partner's parents, children whose parents die, etc.

Transitions between states are applied at the residency level, based on the characteristics of the head resident.

A matrix of transition probabilities from each state to each other state are estimated directly from survey data and are adjusted based on marriage, education levels, and race.

Therefore, each year a residency's probability of moving from state k to state ℓ is a function given by:

$$P(\texttt{state}_{t+1} = k \mid \texttt{state}_t = \ell) = F_{k,\ell}(m_{it}, e_{it}, r_{it})$$

where i is the head individual of that residency.

Table of Contents

References

Stock, J. H. and M. W. Watson (2016): "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," in *Handbook of Macroeconomis, vol. 2*, Elsevier.

Table of Contents

Survey Weight Scaling

Because the initial population bootstrap works at the household level, it may underrepresent demographic groups with high individual survey weights.

The initial population bootstrap process therefore needs to be calibrated on gender:

1. A population is constructed according to the bootstrap process described previously
2. PWBMsim calculates the difference between the frequency of $g = \texttt{male}$ in the bootstrapped population and the original survey data
3. A binary search algorithm chooses weight scalars to minimize this difference
4. Original survey weights w_i are multiplied by these scalars before the bootstrap for the initial population.

This process is repeated to calibrate based on the longest consecutive streak of undercounted ages.