# DE-US-Sim

Demographic and Economic Microsimulation: Methods

Penn Wharton Budget Model

This Version: 2026-11-11

**Abstract**

DE-US-Sim projects a population of people linked through families, residences, and tax units. Demographic events, schooling, work, and earnings evolve through conditional statistical rules. Household donor sampling supplies the initial population and immigrant entrants; births generate new life histories. An aggregate labor-input index connects the changing workforce to the earnings scale and to private-output projections. This account specifies the economic rules of the assembled model and the inputs needed to reproduce its behavior.

## Contents

**1 Economic purpose and model boundary** **3**

1.1 Population, timing, and notation . . . 3

1.2 Economic dependencies and closure . . . 4

1.3 Inputs and policy comparisons . . . 4

**2 Population representation and the initial population** **5**

2.1 Donors, calibration and household sampling . . . 5

2.2 Annual stocks and event timing . . . 6

**3 Births and survival** **7**

3.1 Fertility and newborn attributes . . . 7

3.2 Mortality and competing emigration . . . 8

**4 International migration, legal status and location** **8**

4.1 Arrival levels and household composition . . . 8

4.2 Legal status and emigration . . . 9

4.3 State location . . . 10

**5 Partnerships, living arrangements, and tax units** **11**

5.1 Children and household predictors . . . 11

5.2 Partnership formation . . . 12

5.3 Cohabitation and partnership dissolution . . . 13

5.4 Leaving the parental home . . . 14

5.5 Residence formation and family dependence . . . 14
5.6 Tax-unit construction . . . 15
5.7 Social Security starting information . . . 16

**6 Education and family background** **17**
6.1 Qualifications, schooling and entrant records . . . 17
6.2 Childhood inputs and the score at age 17 . . . 17
6.3 Conditional qualification transitions . . . 18
6.4 Graduate type and subsequent enrollment . . . 20

**7 Work disability, employment and hours** **21**
7.1 Work disability . . . 21
7.2 Fitted work transitions and dependence . . . 21
7.3 Annual hours and unemployment duration . . . 22
7.4 Occupation and entrant work histories . . . 23

**8 Earnings histories and aggregate labor input** **23**
8.1 Predictable log earnings . . . 24
8.2 Persistent heterogeneity and earnings shocks . . . 24
8.3 Real and nominal earnings . . . 25
8.4 Quality-adjusted labor input . . . 26

**9 Macro paths and reported economic quantities** **27**
9.1 Prices, productivity, and the earnings scale . . . 27
9.2 Counting and expansion . . . 27
9.3 Aggregate labor, capital, and private output . . . 28
9.4 Fertility, mortality, and period life tables . . . 28

**A Implementation components** **29**
A.1 Assembly, inputs and aggregate outputs . . . 29
A.2 Population, vital events and migration . . . 29
A.3 Partnerships, residences and tax units . . . 30
A.4 Education, work and earnings . . . 30

# 1 Economic purpose and model boundary

DE-US-Sim asks how changes in the composition and life histories of the United States population change employment, earnings, households, and the inputs to long-run economic and budget projections. It is PWBM's annual dynamic microsimulation: a person retains education, kinship, migration history, and earnings components as the population advances through time. Its economic objects include people, couples, residences, and tax units. A residence is a co-resident group; biological kinship and a couple's union are distinct relationships and can persist across changes of residence.

The selected model combines household sampling, exogenous demographic rate tables, fitted conditional education and work predictors, an earnings process, and an aggregate labor-input calculation. The distribution of demographic and economic outcomes responds to the simulated population. Price growth, multifactor-productivity growth, capital–labor-ratio growth, and production elasticities are supplied externally. The model does not solve household utility maximization or a market-clearing capital accumulation problem.

The microeconomic outputs describe demographic histories, labor supply, earnings, and tax-unit membership. Tax-unit assignment supplies a grouping of people rather than tax liabilities. The aggregate interface supplies population, employment, wages and salaries, and private-output and productivity indices to PWBM's EconomicProjections component. Base-year Social Security record attributes are input characteristics, rather than a simulated benefit entitlement or payment system.

## 1.1 Population, timing, and notation

Let $t$ denote calendar year, $b$ the base year, $i$ a person, and $h$ a residence. Age is $a_{it} = t - \text{birth year}_i$: age attained during the calendar year, without adjustment for the birthday. The base population is sampled in year $b$; a run of $T$ transitions represents years $b, \ldots, b+T$. The selected run specification sets $b = 2020$, $T = 5$, and a sample of 1,000 donor households. These values describe a run specification, not national population totals.

The year's population $\mathcal{P}_t$ includes continuing residents, immigrant arrivals, and newborns, including people who die or emigrate during that year. Departures are excluded from $\mathcal{P}_{t+1}$. This convention preserves within-year earnings and demographic events for people who leave. Let $\mathcal{C}_t$ denote continuing residents, $\mathcal{I}_t$ arrivals, and $\mathcal{B}_t$ newborns. Arrivals enter before continuing residents' transitions; newborns enter after those transitions. Arrivals carry donor work and earnings in their arrival year and do not undergo the continuing work transition. They may form unions and die in that year. Continuing women can give birth in a year in which they subsequently die or emigrate. Infant deaths are drawn after births, and death-ended unions close after the annual calculations.

| Symbol | Meaning |
|---|---|
| $b, t, T$ | Base calendar year, current calendar year, and number of transitions. |
| $i, h, a_{it}$ | Person, residence, and calendar-year age. |
| $\mathcal{P}_t$ | All people represented during year $t$, including that year's departures. |
| $\mathcal{C}_t, \mathcal{I}_t, \mathcal{B}_t$ | Continuing people, immigrant arrivals, and newborns. |
| $w_j, m_h(c)$ | Original donor person weight and calibrated household sampling weight. |
| $W$ | Common expansion weight for exported microeconomic records. |
| $U, \epsilon$ | Fresh uniform $[0, 1)$ and standard-normal innovations unless a shared draw is specified. |
| $\mathbf{1}\{A\}$ | One if condition $A$ is true, zero otherwise. |
| $\text{expit}(x)$ | $1/(1+e^{-x})$, the logistic probability function. |
| $P_t, E_t$ | Price level and earnings adjustment index. |

## 1.2 Economic dependencies and closure

Figure 1 shows economic links, rather than a sequence of software operations. Schooling changes work probabilities and earnings; parental earnings and household composition condition children's schooling; earnings influence independent living and tax-unit assignments. These links make a demographic change persist into later cohorts. The aggregate labor-input index feeds the nominal earnings scale, which in turn enters future household income and earnings histories.

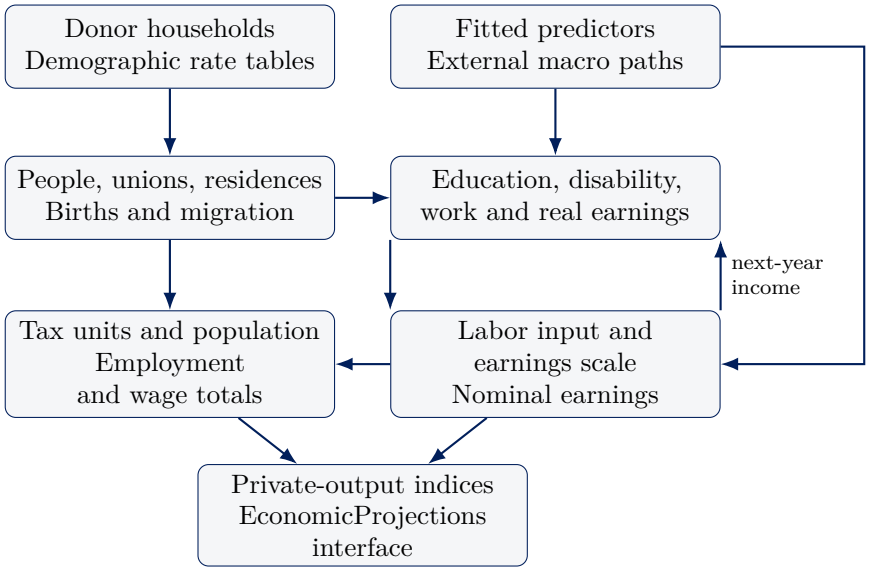


Figure 1: Economic inputs, within-model dependencies, and aggregate outputs.

The stochastic specification uses independent innovations across people and years conditional on the modeled state, except for the household and couple draws, shared uniforms across repeated schooling choices, and persistent earnings components specified below. Independent innovations do not imply independent outcomes: household sampling, partner selection, parental background, common aggregate paths, and inherited state generate dependence.

## 1.3 Inputs and policy comparisons

The empirical inputs are augmented donor records; rates of births, deaths, migration, naturalization, union formation and dissolution; fitted education, disability and work models; earnings coefficients and residual parameters; and macro paths. Their fitted coefficients and classification probabilities are inputs to this account, not quantities inferred from the simulation itself. The donor interfaces identify the U.S. Census Bureau's American Community Survey (ACS); labor and education inputs identify the Current Population Survey (CPS), jointly sponsored by Census and the Bureau of Labor Statistics (BLS), and BLS's National Longitudinal Surveys of Youth (NLSY79 and NLSY97). These producers' survey descriptions establish their demographic and longitudinal roles.[1] Augmentation supplies the model-specific histories, links, latent earnings components, and fitted-model covariates required by the mechanisms below. The supplied divorce coefficients identify the University of Michigan's Panel Study of Income Dynamics (PSID); parental-home departure coefficients identify the Census Bureau's Survey of Income and Program Participation (SIPP). Both are longitudinal household surveys.[2]

[1] U.S. Census Bureau, About the ACS and About the CPS; BLS, National Longitudinal Surveys: Handbook of Methods. External descriptions checked 2026-11-11.

[2] PSID, Institute for Social Research; U.S. Census Bureau, About SIPP. Checked 2026-11-11.

For a comparison, a scenario specifies changed rates, demographic targets, or external paths. Parameters outside that intervention remain supplied inputs. Mortality and fertility multipliers default to no adjustment; immigration target scalars also default to no adjustment. Immigrant weight calibration, state migration conditional on demographic attributes, and the education pandemic adjustment are enabled; the optional log-earnings dispersion cap is off. An intervention that changes donor targets changes the calibrated sampling distribution. The labor-input regressions are fitted to each simulated year's workforce, including in each scenario; the externally fitted education and work predictors are reused unless the scenario supplies replacements. Changes in births, education, family links, and earnings states propagate into later years through those states. A scenario is therefore defined by its inputs and population evolution, not by a statutory rate change alone.

# 2 Population representation and the initial population

The unit carried through time is a person linked to parents, a partner, a residence and a tax unit. These links make population changes consequential for education, earnings and family composition. A donor household is the unit sampled to construct the initial population, whereas an immigrant donor residence is the unit sampled for arrivals. Neither sampling unit is a tax unit. Within a sampled copy, kin and partner links are rebuilt between its members; repeated draws produce separate families rather than additional weight on a single family. Persons outside the same initial household copy cannot supply a linked parent or partner. Missing parent identifiers denote no linked parent in the simulated population.

Let $t$ denote calendar year, $b_i$ person $i$'s birth year and $a_{it} = t - b_i$ the person's age attained during that year. A birthday month is retained separately; it does not change this annual age definition. For a foreign-born person arriving in year $r_i$, years of US residence are $t - r_i + 1$, so the arrival year counts as one. US-born persons have no immigration year and no foreign-residence duration. For initial foreign-born donors, an immigration year after the base year is capped at the base year. The initial birth year is the base year minus the donor's reported age.

## 2.1 Donors, calibration and household sampling

The base population uses a supplied person-level donor pool for the selected base year $b$. Its inputs include household membership, residence roles, partner and parent links, age, sex, two race classifications, education and enrollment, prior births, nativity, birthplace, state, legal status, disability, labor and earnings records, educational background and initial Social Security receipt. The export interface selects an American Community Survey (ACS) population pool; ACS is produced by the US Census Bureau.[3] The pool is a prepared and augmented empirical input, rather than the raw survey alone. Its additional histories, imputations and fitted quantities are inputs to the mechanisms described below.

Write $w_i > 0$ for donor $i$'s survey weight, $h(i)$ for the household containing $i$, $n_h$ for its member count and $c_i$ for a calibration factor. Define the household sampling weight

$$m_h(c) = \frac{1}{n_h} \sum_{i:h(i)=h} w_i c_i. \tag{2.1}$$

Whole-household draws preserve family dependence but give each member the household's sampling probability. Calibration therefore reconciles survey-weighted person shares with the shares implied by

[3] US Census Bureau, *Understanding and Using the American Community Survey Public Use Microdata Sample Files*, https://www.census.gov/programs-surveys/acs/library/handbooks/pums.html, accessed 2026-11-11.

household mean weights. For each level $\ell$ of a calibration dimension, define

$$q_\ell = \frac{\sum_i w_i \mathbf{1}\{i \in \ell\}}{\sum_i w_i}, \qquad \widehat{q}_\ell(c) = \frac{\sum_i m_{h(i)}(c) \mathbf{1}\{i \in \ell\}}{\sum_i m_{h(i)}(c)}. \tag{2.2}$$

Starting from $c_i = 1$, multiply each member's factor by $q_\ell / \widehat{q}_\ell(c)$ for the level it occupies, and recompute household means after each dimension. The dimensions are visited in the order sex, bridged race, education and age. Sex has male and female categories; bridged race has White, Black, Hispanic, Asian/Pacific Islander and American Indian/Alaska Native categories; education has less than high school, high school, some college, bachelor's and graduate categories. Age uses five-year bands with a common 90-and-older band.

Calibration stops when every marginal relative discrepancy is below 0.001. Every declared level must have positive survey-weight mass. The factor is restricted to $|\log c_i| \leq 4$, and convergence must occur within 150 complete cycles. As a joint-distribution check, observed joint cells are ordered by survey-weight share, the smallest quarter of cells is discarded (the number discarded is the nearest integer to one quarter, rounding halves upward), and no retained cell may have a relative discrepancy above 0.20. These conditions define the admissible calibrated donor distribution; they do not match the particular random realization of the subsequent draw.

Let $H$ be the requested number of initial household draws. Households with a reference person are grouped by member count. If $k_*$ is the smallest observed size whose total household sampling weight is less than 5 percent of all households' sampling weight, every size at or above $k_*$ is pooled into a top stratum. If no size meets that condition, sizes remain separate. An unobserved size does not create a stratum or determine the threshold. Households without a reference person form a separate stratum; their sampling weight enters the denominator of the 5 percent comparison. For each resulting stratum $s$, set

$$H_s = \left\lfloor H \frac{\sum_{h \in s} m_h}{\sum_h m_h} \right\rfloor, \qquad \Pr(h \mid s) = \frac{m_h}{\sum_{g \in s} m_g}. \tag{2.3}$$

Draw $H_s$ households independently with replacement within the stratum and copy every member of each draw. Truncation is applied separately to each stratum; its remainder is not reassigned. Consequently the realized number of households can be below $H$, and the number of persons is determined by the households selected. An initial population with no household draw is inadmissible. A household drawn several times contributes that many complete copies.

The resulting simulation population is an unweighted set of person records. Survey weights determine donor selection and are not individual longitudinal expansion weights. Randomly realized demographic shares can differ from calibration targets. Initial parent pointers that do not name a member of the selected household copy become absent links. Initial unions and residences are reconstructed within each copy; the reserved ward residence groups persons who have no ordinary residence assignment without making them one economic household.

## 2.2 Annual stocks and event timing

Continuing persons are those carried from the preceding year after its departures. Arrivals enter before the continuing population's annual changes; births are realized after those changes. Deaths and emigrations remain represented in the year in which they occur, including that year's earnings calculations, and leave the continuing population before the following year. Accordingly a within-year record count includes entrants and that year's departures, while an end-of-year surviving resident count excludes deaths and emigrants. These universes must be distinguished when reporting a population level.

Births are drawn before the mother's death or emigration is determined. A continuing woman can therefore give birth in a year in which she departs. The mother's partner after that year's union formation

and dissolution supplies the newborn's father link. Death-ended unions close after births, so a partner who dies in the year can remain the newborn's linked father. Arrivals do not give birth or emigrate during their arrival year, but they can form unions and die. Newborns face infant mortality in their birth year. These timing conventions specify a joint annual process; they do not describe monthly exposure times.

# 3 Births and survival

## 3.1 Fertility and newborn attributes

Every continuing woman aged 15–49 inclusive receives an annual fertility intensity from a supplied table. Its conditioning variables are calendar year, age, bridged race, education, legal-marriage status and enrollment. Legal marriage here means an active union that is not cohabitation-only. Education and enrollment are taken from the preceding year; marriage is current after union formation, cohabitation assignment and dissolution. The fertility education category pools bachelor's and graduate degrees. Enrollment retains program code 3 or 4 only for a person in the some-college education category and is zero otherwise.

For a rate table ending at $T_F$, use the row for $\min(t, T_F)$. This rule holds the last supplied year's conditional rates constant thereafter; it does not interpolate a missing earlier year. Let $\lambda_{it}$ denote the selected intensity after multiplication by every applicable positive fertility-policy scalar. Scalar conditions use the actual calendar year $t$, even beyond $T_F$. Conditions within a scalar are conjunctive; scalar effects compound when several conditions apply. With independent Poisson innovations conditional on the state and inputs,

$$Z_{it} \sim \text{Poisson}(\lambda_{it}), \qquad B_{it} = \min\{Z_{it}, 8\} \tag{3.1}$$

for eligible women; $B_{it} = 0$ for everyone else. The supplied eligible intensity must be positive, finite and within the supported Poisson domain. The recorded fertility rate is the uncapped Poisson intensity; the expected capped birth count is $E[\min\{Z_{it}, 8\}]$, not exactly $\lambda_{it}$. Birth counts are not Bernoulli indicators: multiple births can occur within the annual model period. The selected defaults apply no fertility scalar.

Each birth becomes a separate person. Its birth year is $t$, its birth age is zero and its mother link is the woman whose count generated it. Sex is male with probability 105/205 and female otherwise. If a father is linked, one independent fair coin chooses the mother's or father's pair of race classifications; the bridged and survey race labels are inherited together from that same parent. Without a father link, both labels are the mother's. These sex and parent-side draws are independent across newborn records, including siblings.

An additional independent uniform draw assigns birthday month from a supplied twelve-month cumulative distribution $F_m$: the month is the first $m$ for which $U < F_m$, with the final month receiving the upper tail. The same month distribution applies to initial donors, arrivals and newborns and is retained throughout life. The January–December cumulative probabilities are

| | | | | | |
|---|---|---|---|---|---|
| 0.08142629 | 0.15728812 | 0.24040362 | 0.32015500 | 0.40366927 | 0.48682666 |
| 0.57473308 | 0.66406912 | 0.75077514 | 0.83602410 | 0.91658156 | 1.00000000 |

The displayed probabilities are rounded for readability; the declared input is the cumulative distribution itself.

Newborns start as US-born citizens with no foreign-residence duration or DACA eligibility. They have no prior births, disability or labor activity. Their educational and earnings starting states are specified with those mechanisms. After births, a continuing mother's cumulative prior-birth count increases by the number of her newborn records, including newborns that die in the birth year. Infant survival does

not reduce that birth-history count. Residence placement follows the available mother, then father, then ward rules in the household mechanism.

### 3.2 Mortality and competing emigration

Mortality uses an annual hazard table indexed by year, age, bridged race, sex, education, legal-marriage status and enrollment. Education and enrollment use the same recodes as fertility, but mortality reads the current year's values. For a child below age 15 who has positive recoded education or legal-marriage status, the education, marriage and enrollment keys are all set to zero. Otherwise the recoded keys are retained. Let $h_{it}$ denote the table hazard multiplied by every applicable positive mortality scalar, with table year capped at its final supplied year $T_D$ and scalar conditions evaluated at the actual year. Then

$$p_{it}^{D} = \begin{cases} 1, & a_{it} \geq 120, \\ 1 - \exp(-h_{it}), & a_{it} < 120. \end{cases} \tag{3.2}$$

All rated records require a usable table rate and a strictly positive scaled hazard, including those at the maximum age. The selected defaults apply no mortality scalar. A separate uniform innovation generates each person's death indicator. A continuing person already marked as emigrating in that year is excluded from death: with $E_{it}$ the emigration indicator and $U_{it}^{D}$ uniform on $[0, 1)$,

$$D_{it} = \mathbf{1}\{U_{it}^{D} < p_{it}^{D}\}(1 - E_{it}), \qquad X_{it} = \max\{D_{it}, E_{it}\}. \tag{3.3}$$

Here $X_{it}$ indicates departure before the next year's continuing population. Birth-year mortality applies the same hazard transformation at age zero to all newborn records. Arrival-year mortality applies it to arriving donor records after their education and arrival unions are established. Neither pass prorates hazards by birthday or arrival month. A death dissolves the affected union and affects which family members can provide a residence to dependents.

## 4 International migration, legal status and location

### 4.1 Arrival levels and household composition

Let $n_t^{\text{open}}$ be the simulated person count at the opening of year $t$, before arrivals, births and that year's departures. Supplied annual lawful and unauthorized arrival rates $\iota_{Lt}$ and $\iota_{Ut}$ set desired person counts

$$A_{kt}^{*} = n_t^{\text{open}}\, \iota_{k,\tau}\, s_{k,\tau}, \quad k \in \{L, U\}, \quad \tau = \min\{t, T_I\}, \tag{4.1}$$

where $T_I$ is the final supplied rate year and $s_{k,\tau}$ is that rate year's nonnegative scenario scalar, or one if none is supplied. Thus a scalar applying only after $T_I$ does not alter the terminal rate: both the rate and scalar are selected by the capped table year. This differs from fertility and mortality scalars, whose conditions use actual calendar time. The selected defaults have no arrival-count scalar.

The immigrant donor pool supplies year-specific households with their demographic and economic attributes. The selected pool year is $\min\{t, T_P\}$, where $T_P$ is its last supplied year; a missing earlier requested pool year is not interpolated. The ACS export interface has annual pools for 2007–2022 and assigns the projection pool to 2023, which is then reused for later simulation years. These are input-interface year labels, not a claim about a currently bound empirical run. Records with the no-residence sentinel are removed before sampling.

Each donor residence is classified as lawful-headed if its reference person is a naturalized citizen or lawful noncitizen; all its members enter that type's pool. Other residences enter the unauthorized-headed pool.

These labels classify the sampling pool by the head and do not overwrite an individual member's legal status. The entry rule admits lawful-noncitizen or unauthorized status on every actual arrival, copies the member's status and DACA flag, and sets arrival year to $t$. Each sampled residence must have exactly one reference person and resolve every nonmissing spouse or parent link inside its own copy.

Optional composition targets alter the donor distribution within each type in years explicitly named by the target table. They do not carry forward by a terminal-year rule. Targets can give marginal category shares or joint cells. Marginal targets assign every donor category to a specified share; grouped levels may be ranges. Joint targets assign shares to combinations of categories. Shares must be nonnegative and sum to one in the relevant distribution. A donor household is removed if any of its members receives a zero target. A joint target that names no donor category has zero mass; a positive target must have supporting donor records. For positive-target records, iterative proportional fitting multiplies each member's weight by the ratio of its target group share to its current weighted share. Dimensions are visited in target-table order. At most five rounds are performed, stopping when all target-to-current ratios differ from one by at most 0.001; a fifth-round discrepancy is retained in the resulting weights. Marginal levels named in overlapping grouped bins use the later bin's target; joint positive-target assignments must be unambiguous. These rules define the admitted target operation, although nonoverlapping marginal bins provide its usual interpretation.

The selected default then applies the household-mean calibration in Section 2 separately to each type's current-year pool. Targets are the marginal shares of the post-raking person weights. Dimensions are the target dimensions, followed by any of sex, partnered status, survey race, education and ten-year age bands not already represented; the top age band is 60 and older. The calibration excludes zero-weight records, carries them back with zero weight, and replaces positive input weights with $w_i c_i$. A zero-weight member of a positive-weight household is still copied when that household is selected. Its marginal tolerance, factor bound and cycle limit are the same as for the base population. The joint-distribution discrepancy is reported without imposing the base population's 20 percent ceiling.

For the resulting weights $\widetilde{w}_i$, define household size $n_h$, total weight $W_h = \sum_{i \in h} \widetilde{w}_i$ and mean weight $\overline{w}_h = W_h / n_h$. Immigration stratification uses person-weight shares $\sum_{h:n_h=k} W_h / \sum_h W_h$. Sizes at or above the smallest observed size with a share below 5 percent are collapsed; no collapse occurs if none is below 5 percent. Let $k_s$ be the exact size for an ordinary stratum and the lower size bound for the collapsed top stratum. With $Q_s = \sum_{h \in s} W_h / \sum_h W_h$, draw

$$H_{kst} = \text{round}_{\text{even}}\left(\frac{Q_s A^*_{kt}}{k_s}\right), \qquad \Pr(h \mid s, k) = \frac{\overline{w}_h}{\sum_{g \in s} \overline{w}_g}. \tag{4.2}$$

The rounding is to the nearest integer with half-integers rounded to the even integer. Strata with a zero draw target contribute no arrivals. Draws are independent with replacement within each type and stratum. Every draw contributes a full household copy, including repeated draws of one donor. Desired arrivals $A^*_{kt}$ are therefore not exact person quotas. In particular, the top stratum divides by its lower size bound while drawing households at least that large, so its realized person count can exceed its desired allocation. This is a household-composition approximation, not an additional behavioral immigration response.

### 4.2 Legal status and emigration

Legal status distinguishes native-born citizen, naturalized citizen, lawful noncitizen and unauthorized resident. Initial donors carry their supplied status and DACA eligibility; a native-born donor cannot be a lawful noncitizen or unauthorized resident. Foreign-born lawful or naturalized persons start with legal-residence years equal to years of US residence, whereas unauthorized persons start with zero legal-residence years. Arrival lawful noncitizens start at one legal year and unauthorized arrivals at zero.

A newborn starts as a native-born citizen. In subsequent years a positive legal-residence count increases by one; zero and absent values remain unchanged until another stated rule changes them.

The model's Deferred Action for Childhood Arrivals (DACA) flag is recomputed in calendar 2012 and otherwise carried. In 2012 it is true exactly for an unauthorized person no older than 30 in that year, who arrived before age 16, and who either has at least a high-school education or is at most age 20. This is the modeled eligibility classification used by the simulation. Arrival and base-year donor flags supply the corresponding initial state; the continuing-person 2012 test is not rerun on new entrants in another year.

Naturalization is drawn for a current lawful noncitizen aged at least 18 whose updated legal residence exceeds five years, or exceeds three years with a citizen spouse. A citizen spouse is a current legally married partner whose preceding-year status is native-born or naturalized citizen. A new partner without a preceding-year record supplies no citizen status to this test. The annual naturalization probability is supplied by year, sex, legal-marriage status, birthplace and age; it is used directly as a probability, not transformed as a hazard. The final table year's probabilities are reused beyond that year. A missing eligible lookup gives no naturalization in that year. One uniform innovation is compared to the probability for each person. A lawful noncitizen below 18 also naturalizes if either linked parent naturalizes in that same draw; co-residence is not required. The parental rule reads the parents' adult naturalization results once, rather than propagating through successive generations.

After naturalization, an adult who remains a lawful noncitizen can overstay if there is no citizen spouse and total years of US residence are greater than one and at most six. A direct Bernoulli probability supplied by year and birthplace determines overstay. Overstayers become unauthorized and their legal-residence count becomes zero. Unauthorized persons in the arrival pool receive a separate traveler-overstay flag from a year/birthplace probability; that flag does not change their already unauthorized status. A missing overstay lookup gives no overstay. Both probability tables reuse their last year after the supplied horizon. Naturalization and overstay have distinct annual uniform innovations. No amnesty or citizenship-act transition is part of this selected legal-status mechanism.

Emigration is drawn for every continuing person from a supplied probability indexed by year, current legal status, total years of US residence, sex and age. The probability's year is capped at the table's final year. A missing key, including a duration without a matching rate, gives probability zero. Independent uniform innovations determine departures. This calculation uses legal status after the year's naturalization and overstay decisions, so those decisions can alter same-year emigration. Arrivals and newborns have zero emigration in their entry year. Emigration precludes same-year death in the reported competing-exit indicators as specified in Section 3.

### 4.3 State location

The model records US states and the District of Columbia. The default migration distribution is a supplied cumulative destination distribution conditional on prior-year state, survey race, current active-union status and whether current education is at least some college. It has no calendar-year index and is reused over the projection horizon. The active-union indicator includes cohabitation, unlike the legal-marriage key in mortality, fertility and naturalization. Demographic conditioning here refers to the destination-state distribution; it does not add state to the birth, mortality, naturalization, or emigration rate-table keys. For a uniform draw, select the destination with the smallest supplied cumulative value at least as large as the draw. Ties select the first destination in the supplied order; if all cumulative values are below the draw, select that first destination. A missing conditioning cell has no defined destination distribution and stops the calculation.

Location is assigned after the annual residence rebuild. Each continuing reference person in a primary residence draws one destination; each person in the ward residence draws separately. Conditional on

these keys, destination innovations are independent across drawers. Every ordinary member follows the final reference person's state, so within-residence moves are shared. An arriving reference person retains the donor state in the arrival year; a continuing member who joins that person's residence follows it. The move indicator compares a continuing person's final state with that person's preceding-year state, so a changed household head can move a member even without that member receiving a separate destination draw. Departing persons remain in the year's record universe and receive the residence and location treatment stated by those rules. Newborns inherit their assigned reference person's state; the ward assignment uses the designated ward state, Idaho (Census code 82). State identifiers are translated to FIPS identifiers and Census regions through the fixed state crosswalk. The distinction between independently drawing heads and dependent members preserves the model's household dependence in geographic mobility.

# 5 Partnerships, living arrangements, and tax units

Household composition transmits demographic changes into fertility, schooling, labor outcomes, and the units supplied to tax analysis. A partnership, a residence, and a tax unit are distinct objects. A residence has a reference person, that person's partner, dependents, and partners of dependents. The common ward residence represents people without a private household; its residents are separate groups for the initial tax-unit assignment. Biological mother and father links persist when people leave the simulation. A missing parent link denotes an unknown parent, whereas a known parent who is no longer present cannot supply current attributes.

An active partnership supplies the model's married indicator even when a person is classified as cohabiting only. Cohabitation is an additional person-level draw; it does not define a separate type of partnership. Accordingly, a married covariate in another equation means membership in an active union unless that equation also conditions on cohabitation.

## 5.1 Children and household predictors

Three child variables enter different relationships. The count of children below 18 counts each represented child who names the person as mother plus each who names the person as father. The two parent-link slots contribute separately. This count has no residence restriction and includes children departing in the current year. In the within-year labor-input calculation, before newborn records exist, it also includes the person's current births and births to women naming that person as current partner. Once newborn records exist, their parent links supply those counts.

The modeled underage-child indicator is different. Its pre-birth component is true for a continuing person if a currently continuing child below 18 names that person as a parent and shared the person's preceding-year residence. Missing residences do not match; two ward assignments do match. Current-year moves and dissolutions do not alter that preceding-residence test. The annual indicator is this component OR giving birth this year OR being the active or death-ended partner of a woman giving birth this year. Births count even when the infant subsequently dies. Base and immigrant entrants supply their donor pre-birth indicators; newborns' pre-birth indicators are false. The terms "underage children" and "coresident minor children" in the fitted partnership, education, work and earnings relationships refer to this modeled indicator at the stated current or lagged date.

An education-only "any children" flag uses the preceding year's completed population and residences. Each child below 18 who did not depart in that year is assigned first to a nondeparting father sharing its residence, and only if that test fails to a nondeparting mother sharing its residence. A person is flagged when at least one child is assigned to that person. Missing residences do not match, while ward

assignments do. Thus a child living with both parents flags the father under this particular education predictor.

### 5.2 Partnership formation

The base population preserves donor partner pairs within each sampled household copy, including same-sex pairs. Partners cannot be younger than 15 under the donor contract. Each partnered base person begins with one counted marriage; everyone else begins at zero. Arriving households preserve valid reciprocal woman–man donor pairs but begin with zero counted marriages. A new market partnership adds one to each endpoint's count, including a partnership between a continuing person and an arrival.

The annual market pairs women aged 18–84 with men aged 19–85. Continuing people must have been single at the preceding year-end, must not already have formed a partnership this year, and must not have been widowed in the preceding year. The preceding year's partnership with an emigrating partner also excludes a person from the next year's market. Current ages, sex, race, nativity, and parent links are combined with preceding-year education, enrollment, residence, and the indicator for underage children.

There are two markets. The first includes arriving singles and eligible continuing singles, and retains only resulting pairs with at least one arrival. The second matches the still-eligible continuing singles. Arrival attributes come from their donor records; their sampled household copy substitutes for a preceding-year residence, and their supplied child indicator substitutes for its lag. Continuing-only pairs produced in the first market participate in its competition for partners but are not retained for the second market.

Each woman initially draws six meetings. For each meeting, independent uniform draws select the race pool, the suitor, and acceptance. With probability one-half the pool contains men of her survey race; otherwise it contains all eligible men. A woman with no eligible man of her race has no meetings, including no all-race meetings. Within either pool, order men by race and then stable person identifier, index them from $L$ to $H$, and select

$$j = \mathrm{round}_{\mathrm{even}}\{L + U(H - L)\}, \qquad U \sim \mathcal{U}[0, 1).$$

Thus, when the pool has at least three men, the endpoints each receive half the probability of an interior man. A one-man pool always selects that man.

A meeting is refused if the two people shared a residence, including the ward residence, if they share a mother or father present in the relevant population, or if either is the other's parent. For continuing people, the relevant parent population is the preceding year's population; the arrival market also includes arrivals' parents in their household copies. Unknown links and absent parent records do not create a shared-parent refusal.

Let $r_{ijtr}$ be the supplied acceptance rate for woman $i$ and man $j$ in meeting round $r$, numbered from one within each pass. Its conditioning set is calendar year, the woman's age, both partners' survey race, education and immigrant status, and the man's age minus the woman's age. The nine inclusive age-gap groups are

$$[-68, -21], [-20, -11], [-10, -6], [-5, -2], [-1, 1], [2, 5], [6, 10], [11, 20], [21, 68].$$

The final supplied year is held fixed afterward; missing cells, including years before available coverage, do not produce a partnership. The rate is successively multiplied by the following factors.

- An arriving person flagged as previously married activates a factor $1.14r/\bar{r}_r$ (using meeting number $r$) when either partner has that flag and their age-gap group is one of $[-5, -2], [-1, 1], [2, 5], [6, 10]$. Here $\bar{r}_r$ is the mean unadjusted acceptance rate across all meetings in that round of the current pass involving a flagged person, omitting missing rates. Continuing persons do not activate this arrival-specific factor.

- For each partner aged 18–24, college enrollment in either of the two undergraduate enrollment categories multiplies the rate by $c_1 = 0.572$; otherwise the factor is $c_0 = 1.165$.
- For each partner aged 25–28 with some college but no bachelor's degree, the factor is $1 + (c_0/c_1 - 1)/2$.
- For a woman aged 18–44 with underage children the factor is $k_1 = 0.93$. For one without children it is $k_0 = (1 - k_1 s)/(1 - s)$, where $s$ is the unweighted share with children among eligible women aged 18–44 before the race-pool exclusion. If that population is empty or all have children, neither child factor is applied.

A meeting is willing when its acceptance uniform is below the adjusted rate. Nonfinite adjusted rates give no willing meeting. Finite rates above one imply certain willingness, and negative rates imply none. Willing meetings are ordered by woman identifier and round; process rounds in their first-appearance order in that ordered list. Within a round, an already-matched woman takes no further partner. An unmatched man is allocated to the lowest-identifier woman seeking him. A conflict with an already allocated man, or a refused meeting, earns one replacement meeting for a woman who remains unmatched. Redraw passes repeat the same rules over the original eligible suitor population, retaining men already allocated as unavailable for a new pair. The selected setting permits 1,000 redraw passes after the initial pass; a woman still unmatched then stays single. These competition and normalization rules create dependence across people's partnership outcomes even though the primitive uniforms are independent.

### 5.3 Cohabitation and partnership dissolution

Every continuing person receives a fresh cohabitation draw each year. Its supplied probability is indexed by current age, sex, bridged race, calendar year, and preceding-year education, combining bachelor's and graduate education in the highest category. The last table year is held fixed after its end, and a missing cell has probability zero. The resulting cohabiting-only flag equals one only if the person is in an active partnership and an independent uniform falls below that probability. Partners' flags can therefore differ. Arrivals use their own current donor education in the same rule, with another draw if they form a market partnership. Base people and newborns begin with the flag false.

Dissolution applies to unions that began before the current year and remain active. Each female endpoint receives a draw, so a woman–woman base union has two opportunities to dissolve and a man–man union has none. The fitted logit uses her survey race, immigrant status, grouped age, and preceding-year education. Education has three categories: less than high school; high school or some college; and bachelor's or graduate degree. Age groups are 16–24, 25–34, 35–54, and 55–84, with younger and older ages assigned to the endpoint group. Writing $x_{it}^U$ for the resulting design vector, including the categorical contrasts and interactions specified by the supplied coefficient table,

$$p_{it}^U = \frac{\exp(x_{it}^{U\prime}\beta^U)}{1 + \exp(x_{it}^{U\prime}\beta^U)}.$$

For a cohabiting-only woman replace this by $\min(1, 1.85p_{it}^U)$, denoted $\widetilde{p}_{it}^U$. Let $S_{tn}$ be the supplied year shift for her immigrant indicator $n$, $J_{tn}$ the corresponding spouse-immigrant shift, $n_j$ her partner's immigrant indicator, and $E_{e_i e_j}$ the supplied adjustment for their five-category preceding-year educations. Dissolution is drawn by comparing a fresh uniform with

$$\widetilde{p}_{it}^U \left[1 + (S_{tn} - 1) + (J_{tn} - 1)n_j + (E_{e_i e_j} - 1)\right].$$

The shift table holds its final year after coverage ends. A missing adjustment cell means no dissolution draw. The threshold is not clipped after adjustment; a threshold at least one gives certain dissolution, and one at most zero gives none. A union ends if any eligible endpoint draws dissolution, and both endpoints' cohabitation flags are cleared.

A death ends an otherwise-active union after the year's births, residence changes, and tax-unit assignment. Thus, mechanisms earlier in that year can still read the dying person's partnership. Continuing endpoints' cohabitation flags are cleared at closure. An arrival whose continuing partner dies retains its drawn cohabitation flag for that arrival year, although the union itself ends. A partner's emigration removes the union when the following year's continuing population is formed.

### 5.4 Leaving the parental home

Continuing persons aged at least 15 who were residence dependents in the preceding year may leave home. The monthly probability is a fitted logit of current education, survey race, sex, immigrant status, current earnings decile, $\log(t - 1995)$, and a quadratic age B-spline with interior knot 29 and bounds 0 and 120. Its clamped knot sequence repeats each boundary three times and drops the first basis function, leaving three spline regressors. Categorical contrasts and interactions are those of the supplied coefficient table. The year covariate requires $t > 1995$.

Earnings deciles are unweighted and use positive nominal earnings of everyone present in the current pre-birth population, including arrivals and persons departing this year. Sort those positive earnings $y_{(k)}$, $k = 1, \ldots, m$, and interpolate the earnings function through probability positions $(k - 1/2)/m$. Let $q_p$ be its value at $p/100$ for $p = 0, \ldots, 99$, extending endpoint earnings beyond their positions. A person's decile is

$$d_{it} = \left\lceil \frac{\#\{p : y_{it} \geq q_p\}}{10} \right\rceil .$$

If nobody has positive earnings, all deciles are zero. This definition includes ties in the lower-bound comparison and allows zero below the lowest cutoff. For an eligible person's monthly probability $p_{it}^M$, the annual draw threshold is

$$p_{it}^L = \rho_L\{1 - (1 - p_{it}^M)^{12}\}, \qquad \rho_L = 2.55.$$

Ineligible people have threshold zero. A fresh uniform below this threshold marks a departure from the parental home; thresholds above one imply a certain departure. Arrivals and newborns do not draw a parental-home departure in their entry year. The residence rules below determine the new household, so this draw does not itself select a destination.

### 5.5 Residence formation and family dependence

Living arrangements follow partner and parent links rather than a housing-market equilibrium. Each base donor residence is copied separately within its sampled household. A base person younger than 15 with both parent links unknown is placed in the ward residence. Base dependents with neither parent present are then treated by the same age-15 founding rule as subsequent residents. Arrivals begin with one residence per sampled household copy and donor residence roles.

A newly formed woman–man couple shares one partner's home, selected by a fair coin. If the husband joins a wife who heads a residence or is a head's partner, he takes the reference-person role; a wife who is the reference person becomes the reference person's partner. A husband joining a dependent wife becomes a dependent's partner. A wife joining a husband who is the reference person becomes his partner, and one joining a dependent husband becomes a dependent's partner. Otherwise the mover retains the supplied role. The same coin is used for an arrival–continuing couple: the arrival's move is made on entry if selected, or the continuing person's move is made during the annual residence update. A two-arrival couple is placed on entry and retains that arrangement.

After the newlywed adjustment, the following rules establish new or retained residences. They are applied in the listed order; a later applicable assignment replaces an earlier one.

1. After dissolution, a man founds a residence unless he was a dependent. A woman who was a dependent's partner founds a residence; a woman who was the reference person's partner heads that residence in place. For a dissolved woman–woman couple, the endpoint holding either partner role founds a residence, and the other remains.
2. A person who was a dependent and is at least 15 founds a residence if both parents are absent or the person drew a parental-home departure. The founding role is reference person, except that a woman with a present, nondeparting partner, or a partnered man who himself departs this year, takes the reference person's partner role for the subsequent family reconciliation.
3. A dependent under 15 with neither parent present, any parentless age-zero person, and a person marked adopted enters the ward residence as a dependent. A ward dependent aged at least 15 instead founds a residence as its reference person; this outflow rule takes precedence over ward inflow.
4. When the reference person's partner loses that partner through death or emigration, the surviving partner becomes reference person in place. The departing counterpart is assigned the partner role in its own preceding residence for the remaining year. A dependent's partner who loses that dependent founds a residence, except when the lost partner is already a founder assigned the exceptional partner role in the preceding rule.
5. Unplaced reference persons and their partners retain their adjusted preceding residences and roles. A residence containing only this year's arrivals retains its arrival-year placement.

Here a parent or partner is absent if the link is unknown, names no current person, or names someone departing this year. Departing people retain their own residence membership through year-end, so they remain available to tax grouping.

Remaining persons are placed by alternating a parents pass and a partners pass. In a parents pass an unplaced person, other than a dependent's partner, follows an already-placed mother outside the ward residence; if the mother's placement is still unknown, the person follows an already-placed father outside the ward. Otherwise an already-placed ward parent places the person in the ward. These assignments give the dependent role. In a partners pass an unplaced person follows an already-placed partner, taking the complementary role: reference person versus partner, or dependent versus partner of dependent. Following a ward partner gives the dependent role. Each pass reads placements as they stand at its start. A mother whose residence is established only in a later pass does not displace an earlier assignment to the father. A departing mother can retain a placement to which a continuing child is assigned for the remainder of the year. The model allows twelve alternating passes and requires every person to have a placement, exactly one reference person per private residence, and compatible partner placements.

A newborn joins the ward if neither parent remains present. If its mother departs and its father remains in a different private residence, it joins the father. Otherwise it joins the mother's private residence, then the father's private residence if the mother has no residence, then the ward if either parent is placed there. Every newborn is a residence dependent. These birth-year rules are applied after the older population's residence assignments.

### 5.6 Tax-unit construction

Tax units are constructed annually from nominal earnings, work disability, schooling, age, legal status, partnership, and residence. The pre-birth assignment includes continuing people, arrivals, and people who die or emigrate during the year. A departed person can therefore remain the household's highest earner. This grouping is an input to a tax calculator; it does not calculate tax liabilities or optimize filing choices.

Let $G_t$ be the model's nominal dependent-income limit. The supplied historical schedule covers 1994–2026. For the selected base year and subsequent supplied years, the model amounts are

| Year | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 |
|---|---|---|---|---|---|---|---|
| $G_t$ (dollars) | 4,300 | 4,300 | 4,400 | 4,700 | 5,050 | 5,200 | 5,300 |

For $t > 2026$ the rule is

$$G_t = 50 \operatorname{trunc}\left\{\frac{G_{t-1}}{50}\left(\frac{P_{t-1}}{P_{t-2}} + 0.001004\right)\right\},$$

where $P_t$ is the externally supplied GDP price index. The recursion uses the preceding two price levels, with downward truncation to a multiple of \$50. Base-year assignment requires a year within the historical schedule.

Within each residence, select the person with maximum nominal earnings $y_{it}$ as reference person $r$; ties go to the highest stable person identifier. Each ward, and each person without a residence, begins as a separate group. A person is a potential dependent if not unauthorized and at least one condition holds: earnings below $G_t$, age at most 18, work disability, or full-time schooling status as defined here. At ages 19–23, the base-year schooling flag comes from donor full-time high-school or college attendance. In later years it requires an increase in attained years of education over the preceding year; an arrival without a preceding observation does not satisfy this test. The initial dependency indicator and unit are

$$D_i^0 = \mathbf{1}\{\text{potential dependent}_i\}\mathbf{1}\{2y_{it} \le y_{rt},\ y_{rt} > 0\}, \qquad T_i^0 = \begin{cases} r, & D_i^0 = 1, \\ i, & D_i^0 = 0. \end{cases}$$

The earnings threshold is strict, whereas the half-reference-income condition includes equality. With no positive reference earnings nobody becomes a dependent through this initial assignment.

Partner reconciliation then reads one fixed snapshot of each partner's initial unit, dependency status, and earnings. First, a dependent in a different unit from its partner takes that partner's initial unit and ceases to be a dependent. Next, among partners treated as nondependents in different units, the lower-earning person's whole current unit, including its dependents, moves one step to the partner's snapshot unit. For equal earners, the larger current unit identifier moves to the smaller snapshot identifier. Own assignments are updated after each pass, while partner reads remain fixed at the initial snapshot. Finally, a person whose partner was initially a nondependent is not a dependent. A couple initially dependent in the same unit remains dependent; two dependents initially assigned to different units can exchange units and both become nondependents. The reconciliation uses active partners, including cohabiting partners under the model's union convention.

Newborns are assigned after this reconciliation. A newborn in a private residence whose pre-birth reference person has positive earnings becomes a dependent in that reference person's final tax unit. Otherwise it is its own nondependent unit. Newborns' zero earnings do not alter the pre-birth reference-person selection.

### 5.7 Social Security starting information

The base population carries each donor's Social Security receipt indicator and annual income as starting information for linked analysis. For people younger than 15, the model sets receipt to false and income to zero; for older people it copies the supplied values. Receipt must be a binary code and income a nonnegative integer. These are start-year attributes, distinct from a calculation of future eligibility, claiming, or benefits.

# 6 Education and family background

Education links childhood household circumstances to adult work, earnings, fertility and mortality. The model draws qualifications and enrollment from supplied fitted relationships, rather than solving a schooling investment problem. The education state comprises highest qualification, current program, most recent program, cumulative years in each postsecondary program, and graduate degree type. A continuing person's qualification is updated before that year's enrollment decision; a new qualification can therefore change the program the person enters in the same year.

## 6.1 Qualifications, schooling and entrant records

Let $K_{it} \in \{0, 1, 2, 3, 4, 5\}$ denote no secondary qualification, General Educational Development (GED) credential, high-school diploma, associate degree, bachelor's degree, and graduate degree, respectively. Let $e_{it} \in \{0, 3, 4, 5\}$ denote no postsecondary enrollment, associate, bachelor, and graduate enrollment. Let $c_{it}^k$ count years enrolled in program $k = 3, 4, 5$, and let $c_{it}^C = c_{it}^3 + c_{it}^4$ be undergraduate enrollment years.

The attainment category used elsewhere has five values: less than high school when $K = 0$; high school when $K \in \{1, 2\}$ and $c^C = 0$; some college when $K \in \{1, 2\}$ and $c^C > 0$, or when $K = 3$; bachelor's when $K = 4$; and graduate when $K = 5$. Its numerical codes are $0, 1, 2, 3, 4$. The corresponding schooling-years index equals 11 for $K = 0$, 12 for a secondary credential without college enrollment, 14 for an associate degree without recorded college enrollment, and 13, 14 or 15 for a sub-bachelor's credential with respectively one, two, or at least three recorded undergraduate years. It equals 16 for a bachelor's degree and 17 for a graduate degree. This is the model's educational index, including for children, rather than an independently accumulated count of school attendance.

Base-population members and immigrants copy their donor qualification, enrollment, recent enrollment, program histories and graduate type. Histories are nonnegative; at ages through 30 each program count is capped at six, while above 30 the bachelor and graduate counts are capped at 12 and 10. An entrant with a graduate degree must have a supplied type: master's, professional or doctoral. Non-graduates have no graduate type. Newborns start with qualification, enrollment and histories zero and no graduate type.

## 6.2 Childhood inputs and the score at age 17

Each continuing person aged 17 or younger obtains maternal attainment and an indicator that the biological father holds at least a bachelor's degree from the parents' preceding-year attainment, provided those parents remain in the current continuing population. An absent mother's category is less than high school; an absent father's indicator is false. Above 17 these two background variables persist. A separate indicator is refreshed at every age: it is true when the person and both currently present biological parents shared the same residence in the previous year. Thus the latter variable measures preceding-year parental coresidence when used in later education transitions; its name does not make it a permanently frozen age-16 measurement.

Mother's age at the person's birth is a permanent personal attribute. Every base and immigrant donor supplies an integer from 15 through 49, including when no mother is linked in the simulated population. A newborn receives the mother's current annual age clipped to that interval. The value persists after the mother leaves; it has no missing-mother fallback. The distinct child-count and child-indicator definitions used below are in Section 5.1.

For people reaching 17, household earnings are the sum of previous-year nominal earnings of currently continuing members sharing their previous-year residence. Ward residents and people without such a

residence receive household earnings zero. Quartiles are computed across continuing 17-year-olds, with one observation per person and no additional survey weight. With $H_{it}$ denoting this household sum and $q_j$ the linearly interpolated sample quantile at $j/4$,

$$g_{it} = \sum_{j=1}^{3} \mathbf{1}\{H_{it} \geq q_j\}, \qquad g_{it} \in \{0, 1, 2, 3\}. \tag{6.1}$$

Equality belongs to the upper group. A single observation, or a cohort whose incomes are all equal, is consequently assigned group 3. This group is retained after 17 and is undefined before 17. The income sum excludes people who departed before the current year, even if they appeared in the preceding-year household. Specifically, for ordered incomes $H_{(1)}, \ldots, H_{(n)}$ and quantile probability $p$, set $u = (n-1)p$, $k = \lfloor u \rfloor + 1$ and $q(p) = (1 - u + \lfloor u \rfloor)H_{(k)} + (u - \lfloor u \rfloor)H_{(\min(k+1,n))}$. With one observation every quantile equals that income; an empty tested cohort requires no assignment.

A fitted linear model supplies the mean of a latent score at age 17:

$$z_{it} = x_{it}^{S}\beta^{S} + \sigma_S \epsilon_{it}, \qquad \epsilon_{it} \sim N(0, 1), \qquad s_{it}^{0} = \Phi(z_{it}), \tag{6.2}$$

where $\Phi$ is the standard normal distribution function. The supplied coefficient terms select an intercept, levels and interactions from sex, survey race, paternal college, maternal attainment, preceding-year Census region, the refreshed parental-coresidence indicator, the person's preceding-year indicator for coresident minor children, income group, and the mother's age at the person's birth and its square. The four-region transformation maps original region codes $\{30, 31\}$ to 1, $\{20, 21\}$ to 2, $\{40, 41, 42\}$ to 3, and $\{10, 11\}$ to 4; missing or other codes become $-1$. Sex is male 0, female 1, and survey-race codes are white 0, black 1, Hispanic 2, Asian 3 and other 4. The input contains $\beta^S$, its actual term definitions and $\sigma_S$. Income group 0 and region 1 are the reference levels in the supplied-model specification; the income effects distinguish groups 1, 2 and 3 and the region effects distinguish regions 2, 3 and 4.

With the selected education pandemic adjustment enabled, the model adds a normally distributed learning-loss term to $s^0$ for cohorts tested in 2025–2030. The means by test year and income group are

| Test years | Groups 0–1 | Groups 2–3 |
|---|---|---|
| 2025–2027 | −0.0342 | −0.0292 |
| 2028–2030 | −0.0650 | −0.0192 |

In each band the standard deviation is half the absolute sum of its two means. The loss draw is independent of the latent-score normal. The adjusted score is bounded by the minimum and maximum of $s^0$ among that year's tested people. Outside those years, the adjustment is zero. The stored score is $S_{it} = \lfloor 100{,}000 s_{it} \rfloor$ and persists after 17; people below 17 have no score. Education regressions consume $S/100{,}000$, using $-1$ when the score is absent.

Donor entrants aged 17 or older retain their donor score and background record. Their supplied survey-source identifier distinguishes the Bureau of Labor Statistics' National Longitudinal Survey of Youth 1979 and 1997 cohorts (NLSY79 and NLSY97). A donor NLSY79 income-group marker of $-1$ becomes an absent group; other entrants aged 17 or older require a group 0–3. Younger entrants have no score or income group. Newborn maternal attainment starts at the lowest category and the parental indicators start false before subsequent annual updates. The NLSY labels identify the supplied empirical models, not a rule restricting the simulation to the original survey birth cohorts.

### 6.3 Conditional qualification transitions

At ages through 30 the younger-survey fitted models apply only when the stored income group is nonmissing. At ages 31–60 the older-survey models apply without that condition. All completion risk

sets use preceding-year qualification and enrollment history. The following table supplies the remaining eligibility and ordered outcomes; "none" means no new qualification.

| Prior qualification | Ages, younger model | Required prior enrollment | Outcomes |
|---|---|---|---|
| No secondary credential | 18–30 | None | None, GED, high school |
| GED or high school | 20–30 | $c^3 + c^4 > 0$ | None, associate, bachelor |
| Associate | 21–30 | $c^4 > 0$ | None, bachelor |
| Bachelor | 23–30 | $c^5 > 0$ | None, graduate |

Each row also has a separate age-31–60 fitted model. For the older GED/high-school row the prerequisite is $c^4 > 0$. At exactly age 31, nonnegative associate and bachelor histories are merged into the bachelor count and the associate count is set to zero before evaluating these histories. Outside all risk sets, qualification is retained. Within them, a nonzero completion replaces the prior qualification; the risk sets are disjoint, so one person acquires at most one qualification in a year.

The first two rows use multinomial logits with no-completion as the base outcome: $p_j = \exp(x\beta_j)/(1 + \sum_{k>0} \exp(x\beta_k))$ for $j > 0$ and $p_0 = 1/(1 + \sum_{k>0} \exp(x\beta_k))$. The other rows use binary logits. Each input coefficient table declares its own terms and interactions. The common conditioning vocabulary consists of sex; original survey race; a three-way race recoding that joins Asian and other to white; a second race group joining black, Hispanic and other; native-born status and Hispanic-immigrant status; preceding-year four-region geography; mother's age at birth; birth month; preceding-year indicators for any children and coresident minor children; prior qualification; prior and most recent programs; cumulative program counts; age and separate indicators at ages 18–26; the normalized score; income group; current parental coresidence; and the retained maternal and paternal educational background.

Program features distinguish prior enrollment in associate or bachelor programs, graduate enrollment, and bachelor-or-graduate enrollment. For the last feature an associate program is treated as bachelor from age 31. The undergraduate-prior-program feature recodes graduate enrollment to bachelor; other unrecognized values become $-1$. The bachelor-history indicators identify exactly two and exactly four years, and three complete-history indicators equal one. Both age-16 and age-18 labels for parental coresidence use the current refreshed indicator; the alternative maternal and paternal background labels use the same retained variables. A recent graduate program is recoded to bachelor for the younger GED/high-school completion sample. These transformations define the economic regressors; the supplied term lists determine which of them enter each fitted equation.

The model adjusts one outcome's weight multiplicatively, preserving the relative weights of the others. For outcome $j$ and adjustment $f$, define

$$\widetilde{p}_j = p_j(1 + f), \qquad \widetilde{p}_k = p_k \frac{1 - p_j(1 + f)}{1 - p_j}, \quad k \neq j. \tag{6.3}$$

Write $r_C$ for the supplied NLSY97–CPS college comparison ratio, $r_G$ for its graduate ratio, and $r_H$ for the Hispanic over-17/under-17 education ratio. CPS is the Current Population Survey, jointly sponsored by the Census Bureau and Bureau of Labor Statistics. Adjustments are rounded to four decimal places. For the no-secondary-credential row the adjusted outcome is no completion and $f = \mathbf{1}\{\text{Hispanic immigrant}\}\,\text{round}(r_H - 1, 4)$. For the GED/high-school row the adjusted outcome is bachelor, and for the associate row it is completion; both receive $\text{round}(r_C - 1, 4)$ at ages 22–25 and zero at other ages. The bachelor row adjusts graduate completion by $\text{round}(r_G - 1, 4)$ at ages 27–28 and zero otherwise. The older-survey models receive no age-specific college or graduate adjustment. Supplied race-by-year education trend tables are evaluated at the earlier of their last available years once the simulation passes that year; the selected model multiplies every trend by zero, so these adjustments have constant levels. Years preceding the available trend coverage require the corresponding inputs.

There is one independent uniform draw for each eligible person's completion. For a binary model, outcome 1 occurs exactly when the adjusted weight for that outcome exceeds the uniform. For more than two outcomes the selected outcome is the first cumulative adjusted weight strictly exceeding the uniform, or outcome 0 if none does. Adjusted weights are used without clipping to $[0, 1]$. Undefined cumulative comparisons do not select an outcome; the no-exceedance rule then selects zero. These draw conventions also define the rule at a saturated input probability in equation (6.3).

### 6.4 Graduate type and subsequent enrollment

A person acquiring a graduate qualification draws its type from cumulative thresholds $t_0, t_1, t_2$ supplied by sex and survey race. For a fresh uniform $u$, professional is selected for $t_0 \leq u < t_1$, doctoral for $t_1 \leq u < t_2$, and master's otherwise. An absent cell consequently yields master's. Existing graduates keep their type and non-graduates have no type.

Enrollment uses this year's qualification with the same background and lagged history features as completion. The risk sets and ordered options are

| Current qualification | Ages | Options |
|---|---|---|
| GED, high school or associate | 18–30 | None, associate program, bachelor program |
| GED, high school or associate | 31–60 | None, bachelor program |
| Bachelor | 20–30 | None, graduate program |
| Bachelor | 31–60 | None, graduate program |

The two younger samples require a nonmissing income group. Others have no postsecondary enrollment that year. The first row uses a multinomial logit and the remaining rows binary logits. The first row initially adjusts bachelor enrollment by $\text{round}(r_C - 1, 4)$ at ages 18–21. It is then adjusted a second time on the non-enrollment option and redrawn with the same uniform. The second adjustment includes $\mathbf{1}\{\text{Hispanic immigrant}\}\,\text{round}(r_A - 1, 4)$, where $r_A$ is the distinct supplied immigrant/nonimmigrant college-attendance ratio. The age-31–60 bachelor-enrollment model applies this same immigrant term to non-enrollment in a single draw. A positive $r_A - 1$ therefore increases the non-enrollment weight. The graduate-enrollment model at ages 20–30 adjusts the enrolled outcome by $\text{round}(r_G - 1, 4)$ at ages 25–26; its older counterpart has no such adjustment.

The enabled pandemic enrollment adjustment applies in 2020 and 2021. For the younger undergraduate sample, the first draw defines group enrollment share $v$. Groups with no previous undergraduate enrollment are split by income quartile; their 2020 relative enrollment changes $\kappa$ are $-0.101, -0.101, -0.049, -0.049$. People with positive undergraduate history are split by latest associate or bachelor program, with changes $-0.052$ and $-0.006$. The latest-program rule takes bachelor if either prior or recent program is bachelor, otherwise associate if either is associate, otherwise none. The 2021 changes are 0.85 times these values, rounded to three decimals. The second non-enrollment adjustment adds

$$f^{\text{pandemic}} = \text{round}\left(\frac{1 - v(1+\kappa)}{1 - v} - 1, 4\right). \tag{6.4}$$

Shares count people in each simulation group without survey weights. A saturated group has the corresponding infinite adjustment and follows the draw conventions above. Graduate enrollment at ages 20–30 instead adds 0.036 to the enrolled adjustment in 2020 and zero in 2021. Outside those years all enrollment pandemic adjustments vanish; resulting schooling histories and scores can affect later outcomes.

Each enrolled year increments the corresponding cumulative count, subject to the age-specific caps already stated. Nonzero current enrollment replaces the recent-program field; non-enrollment preserves its prior value. A missing recent program is replaced by zero for people without a secondary qualification or with a graduate degree. A recent graduate program is recoded to bachelor in the younger undergraduate-enrollment sample. Qualification, program histories and test scores then enter subsequent years as state variables.

The selected scalar inputs are $r_C = 1.1470119069179148$, $r_G = 1.3817283302950532$, $r_H = 1.2327214579875942$, and $r_A = 1.2333229925834976$. They are model calibration inputs, not statutory parameters. Their economic labels identify comparisons across supplied education samples; an independent implementation requires the fitted coefficient objects and their coding, as well as these ratios.

# 7 Work disability, employment and hours

Labor outcomes are conditional annual draws. Disability, family status and education affect employment; current employment determines worker class and hours; these outcomes determine earnings. Household links enter through partner education and employment. Persistence comes from lagged characteristics, fitted transition probabilities and the additional persistence rules below.

## 7.1 Work disability

Every continuing person aged at least 15 receives a Bernoulli work-disability draw with probability

$$\Pr(W_{it}^{\text{dis}} = 1 \mid \mathcal{H}_{it}) = \left[1 + \exp\{-x_{it}^{\text{dis}}\beta^{\text{dis}}\}\right]^{-1}. \tag{7.1}$$

Here $W^{\text{dis}}$ denotes work disability, $\mathcal{H}_{it}$ is the available conditioning information, and $x^{\text{dis}}$ contains the supplied model's terms in lagged disability, survey race, current partnership and age. The age basis is a clamped cubic B-spline with interior knots 18, 36, 50 and 62, boundary knots 0 and 120 each repeated four times, and the first basis function omitted; the model's separate intercept is included when its term list specifies one. Partnership is the model's current union indicator, after this year's formation and dissolution. Prior disability is joined to the same person. Below 15 disability is false. Base-year donor disability uses 0 and 1, with the donor not-in-universe code $-1$ treated as false. Immigrants and newborns start with disability false; immigrants join the annual disability draw in their next continuing year. Ages outside the spline domain are not assigned an extrapolated probability.

## 7.2 Fitted work transitions and dependence

The model takes five sets of fitted classification forests as inputs: labor-force status, worker class, full-time status, hours worked and unemployment duration. Each supplies its feature list, class labels and conditional class probabilities. A probability column maps to its supplied class label; the column position is not itself the outcome. A year-specific fitted object is used when available. Beyond the last fitted year a supplied pooled projection object is used; an internal missing year is not filled from an adjacent year. When an object declares a first available year, earlier years are outside its domain.

Unless a certainty or persistence rule below applies, one uniform is inverted against the cumulative probabilities in their supplied class order, taking the first cumulative probability strictly greater than the draw. A rounding shortfall at the upper endpoint selects the last class with positive probability. Ties for the most likely class select its first occurrence in the supplied class order. Random draws for

separate people and separate mechanisms are independent conditional on the current and lagged states, except for explicitly reused draws and the household and aggregate dependencies described here.

For labor-force status, worker class and full-time status, age 16 uses a model without lagged predictors whose vocabulary is sex, current work disability and immigrant status. Above 16 a separate model may use those variables plus age, current five-level attainment, survey race, current schooling, coresident minor children, partnership and partner attainment. “In school” means age 16–24 and current enrollment in program 3, 4 or 5. An absent partner is assigned the lowest attainment and not-in-labor-force status. The fitted object’s feature list selects among the permitted variables; its trained splits remain an empirical input.

Labor-force status has categories not in the labor force, employed and unemployed, with respective codes 0, 1 and 2. Below 16 it is not in the labor force. At 16 the most likely class is used only when its probability is at least 1; above 16 the corresponding threshold is 0.97. Otherwise the categorical draw applies. The older model also permits the person’s lagged labor-force status, lagged annual hours, usual weekly hours and weeks, the current partner’s lagged employed indicator, and age indicators 55–61, 62, 63–64, 65, 66–70 and at least 65. Regardless of the draw, ages at least 90 are outside the labor force.

An employed person draws one of six worker classes: unpaid family worker, unincorporated self-employed, incorporated self-employed, private wage and salary, federal government, or state/local government, with respective codes 0 through 5. Nonworkers have no current class. The older fitted model additionally reads the most recently held class, lagged full-time status and hours, and the current partner’s current labor-force status. If its most likely class is unpaid family work, incorporated self-employment or federal employment, matches the person’s most recently held class, and a second independent uniform is below 0.1, that class is retained. Otherwise the ordinary categorical draw applies. The age-16 model has no such persistence override. The ever-worked flag is the logical union of its previous value and current employment; the most recently held class updates only on employment.

Employed people next draw full-time versus part-time status. The older model also reads current worker class and partner employment, own lagged full-time status and hours, and whether the current partner was full-time last year. A nonworking or absent partner’s lagged full-time indicator is false. The certainty threshold is 1 at age 16 and 0.955 above 16. Nonworkers have no full-time status. Missing lagged class and full-time values are represented by $-1$ when a fitted model consumes them.

### 7.3 Annual hours and unemployment duration

Hours-worked forests are selected jointly by current full-time status and whether prior annual hours were positive. Their no-lag vocabulary contains the current demographic and household predictors above, current class and current partner labor-force status. Their lagged vocabulary additionally contains own lagged disability, schooling, coresident minor children, class, full-time status, annual hours, usual weekly hours and weeks worked. Each fitted forest returns separate marginal probabilities for usual weekly hours and weeks worked.

For an employed person, draw usual weekly hours twice using independent uniforms from the same marginal distribution. Use the second value only if the worker is incorporated self-employed and that value is at least 60; otherwise use the first. Independently draw weeks from its marginal distribution, except that the most likely weeks value is taken when its probability reaches the following class-specific threshold:

| Worker class | Threshold |
|---|---|
| Unincorporated self-employed | 0.95 |
| Incorporated self-employed | 0.75 |
| Private wage and salary | 0.93 |
| Federal government | 0.80 |
| State/local government | 0.93 |

Unpaid family workers have no weeks certainty override. Usual hours must lie in 35–99 for full-time workers and 1–34 for part-time workers, and weeks in 1–52. Annual hours are their product. A positive product below 40 is replaced by 40 hours: usual hours/weeks become 40/1 for full-time or 20/2 for part-time. A zero-hours employed draw is outside the rule's domain. All three hours measures are zero for nonworkers.

An employed or unemployed person with fewer than 52 weeks worked and fewer than 2,080 hours worked draws weeks unemployed, from 0–52. The age-16 model reads sex, disability, immigrant status, current labor-force and full-time status, and current annual hours. Above 16 the vocabulary additionally includes the other current demographic and household variables, current worker class and partner employment, lagged hours unemployed and lagged hours worked. There is no certainty override. Drawn unemployment weeks are capped so that weeks worked plus weeks unemployed are at most 52. Multiplication by 40 produces unemployment hours, which are further capped so that work plus unemployment hours are at most 2,080. A person with both work and unemployment hours zero must be unemployed and receives a 40-hour unemployment floor. Those outside the draw population have unemployment hours zero.

### 7.4 Occupation and entrant work histories

Among employed continuing people, a supplied logit determines employment in science, technology, engineering and mathematics (STEM). Its vocabulary is sex, survey race, immigrant status, lagged STEM indicator, age group, separate indicators for the lowest attainment, some college and bachelor's, and separate master's, professional and doctoral indicators for graduates. Its supplied calendar-year interval terms enter at their stated inclusive endpoints. The age groups have edges

$$16, 17, 22, 24, 29, 34, 39, 44, 49, 54, 59, 64, 69, 120.$$

The first interval includes both endpoints and subsequent intervals are open on the left and closed on the right. A fresh Bernoulli draw is made for each worker; nonworkers have STEM status false.

Donor entrants copy labor-force status, class, full-time status, usual weekly hours, weeks and annual hours worked, unemployment hours and STEM status. Their ever-worked indicator initially means employed at entry, and their most recent class equals the supplied current class. Immigrant labor-force status is set to not in the labor force from age 90, and the resulting donor work record must satisfy the same employment/class/hour consistency conditions. Base-year donors are copied without this age override. Newborn work states have zero hours, no worker class or full-time status, no prior work and no STEM occupation.

## 8 Earnings histories and aggregate labor input

The earnings calculation separates observable characteristics, persistent individual differences, serially correlated shocks, annual hours, and an aggregate earnings adjustment index. A paid worker is an employed person in any worker class except unpaid family work. The model assigns zero earnings when no hours are worked or the worker is unpaid. Earnings are still calculated for continuing people who die or emigrate during the current year; the annual departure flags do not truncate the year's work record.

## 8.1 Predictable log earnings

For paid continuing workers the predictable component of log real full-time equivalent earnings is

$$m_{it} = x_{it}^{Y} \beta_t. \tag{8.1}$$

Full-time equivalent (FTE) means 2,080 annual hours. The coefficients and their term list are supplied by fitted calendar-year model, with a pooled projection model after the last supplied year and no interpolation across an internal missing year. They describe real earnings in the price basis of the empirical input. The predictor vocabulary comprises current sex, disability, immigrant status, attainment, survey race, schooling, coresident minor children, partnership, partner attainment and labor-force status, worker class, full-time status, age group and annual-hours group. Age groups are those defined for STEM occupation. The hours-group edges are

$$40, 80, 320, 780, 1040, 1560, 1820, 2079, 2080, 2340, 2600, 3120, 4000, 5148,$$

with the same endpoint convention. Additional terms distinguish immigrant ages through the supplied age-group indicators, residence of at most two years, 3–5, 6–10 or 11–15 years, and birthplace code 13 or groups $\{12, 50\}$, $\{20, 21, 22, 23\}$ and $\{60, 61, 62, 63\}$. Residence and birthplace indicators in this earnings predictor apply only to immigrants. The model's term list supplies every included level and interaction. Paid predictions must be positive in log units; other people receive $m = 0$.

## 8.2 Persistent heterogeneity and earnings shocks

Write a paid person's residual log earnings as

$$r_{it} = A_{it} + P_{it} + Q_{it} + V_{it}, \tag{8.2}$$

where $A$ is persistent individual heterogeneity, $P$ an autoregressive component, $Q$ a component that sometimes persists unchanged, and $V$ a fresh annual bonus component. Each sex has supplied parameters $\sigma_A, \sigma_P, \sigma_T, v_T, \omega, \rho$. Define $\sigma_V = \sigma_T\sqrt{v_T}$, $\sigma_Q = \sigma_T\sqrt{1 - v_T}$ and $\pi = \exp(-\omega)$, with positive standard deviations and $0 < v_T < 1$. For a continuing paid worker who has previously entered paid work,

$$A_{it} = A_{i,t-1}, \tag{8.3}$$

$$P_{it} = \rho P_{i,t-1} + \sigma_P \epsilon_{it}^{P}, \tag{8.4}$$

$$Q_{it} = \begin{cases} Q_{i,t-1}, & u_{it}^{Q} < \pi, \\ \sigma_Q \epsilon_{it}^{Q}, & u_{it}^{Q} \geq \pi, \end{cases} \tag{8.5}$$

$$V_{it} = \sigma_V \epsilon_{it}^{V}. \tag{8.6}$$

The $\epsilon$ variables are independent standard normals and $u^Q$ is an independent uniform. For people outside paid work, $A, P, Q$ retain their previous values while the earnings residual used in that year is zero. A person returning to paid work with an existing permanent component resumes this process rather than drawing a new donor residual. The annual bonus is not carried into the next year.

Every newborn receives $A_i = \sigma_A \epsilon_i^A$ at birth, with $P = Q = 0$. Every donor entrant also draws $A$ independently; a paid donor entrant additionally receives initial $P, Q, V$ draws. Let $a$ be age and let $a^E$ be the assigned entry age: 17 for the lowest attainment, 19 for high school, 21 for some college, and 23 for bachelor's or graduate attainment. With $n = \max(a - a^E + 1, 1)$, the initial permanent-component variance is

$$V_P(a) = \begin{cases} \sigma_P^2(1 - \rho^{2n})/(1 - \rho^2), & \rho^2 < 0.9999, \\ n\sigma_P^2, & \rho^2 \geq 0.9999. \end{cases} \tag{8.7}$$

Initial $P$ is $N(0, V_P)$, initial $Q$ is $N(0, \sigma_Q^2)$ and initial $V$ is $N(0, \sigma_V^2)$, independently. These component draws are then reconciled to the donor residual $r^* = \log(Y^{F,\text{donor}}) - m^{\text{donor}}$, where the two donor earnings measures are supplied inputs.

The same reconciliation is used on first entry into paid work. Such an entry means current paid class, previous class absent or unpaid, and previous $P = 0$. A residual $r^*$ is drawn uniformly with replacement from positive-earnings CPS donors in the exact sex, attained education, current partnership and age cell, with age capped at 70 for both simulated person and donor. A missing cell does not borrow from another cell. The person keeps the previously drawn $A$, obtains fresh $P$ with variance $V_P$, fresh $Q$ and $V$, and is reconciled to $r^*$.

For either entry calculation, let $A^0, P^0, Q^0, V^0$ be those initial draws and $e = r^* - (A^0 + P^0 + Q^0 + V^0)$. Define

$$\Omega = \sigma_A^2 + V_P + \pi\sigma_Q^2. \tag{8.8}$$

The final components are

$$A = A^0 + e\sigma_A^2/\Omega, \qquad P = P^0 + eV_P/\Omega, \tag{8.9}$$

$$Q = Q^0 + e\pi\sigma_Q^2/\Omega, \qquad V = V^0 + e - e(\sigma_A^2 + V_P + \pi\sigma_Q^2)/\Omega. \tag{8.10}$$

The three adjustments exhaust $e$, so the last expression leaves $V = V^0$ apart from arithmetic rounding. The four components sum to the selected donor residual. Entry matching therefore changes persistent heterogeneity as well as the two persistent stochastic components; subsequent earnings histories inherit that reconciliation.

## 8.3 Real and nominal earnings

The selected model leaves the residual-variance cap off. If a positive target variance $v^*$ is explicitly selected, the paid positive-hours population in the current calculation supplies residual mean $\bar{r}$ and sample variance $v$ with denominator one less than its person count. When $v > v^*$, the residual used for that year's conversion is $\widehat{r} = \bar{r} + \sqrt{v^*/v}(r - \bar{r})$; otherwise $\widehat{r} = r$. Base entrants, arrivals and continuing people supply separate cap populations. The transformed residual does not replace the stored components, so future shocks evolve from the uncompressed history.

For positive-hours paid work, real FTE and annual earnings are

$$Y_{it}^F = \text{round}_{\text{even}}(\exp(m_{it} + \widehat{r}_{it}))\,, \qquad Y_{it}^R = \text{round}_{\text{even}}\Big(Y_{it}^F h_{it}/2080\Big)\,, \tag{8.11}$$

where $h$ is annual hours and ties in whole-dollar rounding go to the even integer. Real hourly earnings are $Y^R/h$. With zero hours or unpaid family work, real FTE and annual earnings are zero; hourly earnings are zero at zero hours and otherwise annual earnings divided by hours. The residual cap is applied before exponentiation.

Let $E_t$ denote the aggregate earnings adjustment index. Nominal FTE and annual earnings multiply their corresponding real-dollar values by $E_t$ and truncate the nonnegative result to whole dollars. Nominal hourly earnings multiply the real hourly rate by $E_t$ without whole-dollar truncation; consequently they need not equal truncated nominal annual earnings divided by hours. The base population uses the base-year supplied index. Arrivals use the preceding stored index grown by the current year's exogenous terms alone. Continuing people use the index after the current labor-input adjustment. These arrival amounts remain in the year's population used to calculate aggregate labor input.

### 8.4 Quality-adjusted labor input

The labor quantity index values annual hours at fitted real hourly earnings. It includes private wage and salary workers and both self-employment classes, and also includes unpaid family workers in the quantity totals. Federal and state/local government workers are excluded. Separate unweighted least-squares regressions are fitted in the current and prior simulation populations using private-sector observations with positive real hourly earnings:

$$\widehat{\gamma}_s = \arg\min_{\gamma} \sum_{i \in \mathcal{V}_s,\, w_{is}^R > 0} [\log w_{is}^R - z_{is}\gamma]^2, \qquad s \in \{t-1, t\}, \tag{8.12}$$

where $\mathcal{V}_s$ is the private-sector population just defined. The minimum-norm least-squares solution is used when the design is rank deficient. There is no lognormal residual correction when converting fitted values to levels.

The design vocabulary contains an intercept; age powers 1–4; sex, attainment and survey-race indicators; disability, full-time and partnership indicators; indicators for one, two or at least three children below 18; hours bands 40–1,039, 1,040–1,820, 1,821–2,079 and above 2,080; and immigrant status. Each age power interacts with sex, each attainment category and each race category. Sex interacts with attainment, race, disability, full-time status, each child-count category and partnership. Each child-count category interacts with partnership and with sex times partnership. Immigrant status interacts with attainment. Both native and immigrant indicators separately interact with residence bands 1–2, 3–5, 6–10, 11–15 and 16–20 years, and with birthplace 13 or groups $\{12, 50\}$, $\{20, 21, 22, 23\}$ and $\{60, 61, 62, 63, 64\}$. Missing residence duration activates none of its bands. Treatment indicators use the lowest observed level as the reference within each year's fitted sample, and terms requiring an absent level are omitted.

Each fit retains its own reference categories and term list when valuing either year's workers. Write its design evaluated on person $i$ in year $v$ as $z_{iv}^{(s)}$, where $s$ is the fitting year. Each retained indicator compares the same fixed category in both valuation years. A category absent from the fit acquires no new indicator or coefficient when it appears in the other population. The two fitted coefficient vectors can therefore have different dimensions.

Current child counts follow the biological-link and pre-birth credit rules in Section 5.1. The current population includes arrivals and people departing during the year. The prior population uses its complete preceding-year records and treats people widowed in that preceding year as partnered for the prior regression. These conventions align the two annual earnings snapshots.

Define the hours total valued using year-$s$ regression prices for year-$v$ people by

$$T_{sv} = \sum_{i \in \mathcal{V}_v} \exp(z_{iv}^{(s)} \widehat{\gamma}_s) h_{iv}. \tag{8.13}$$

Positive predicted rates are applied also to unpaid family workers, who are excluded from the fit but included in this sum. The growth factor and chained level are

$$d_t = \sqrt{\frac{T_{tt}}{T_{t,t-1}} \frac{T_{t-1,t}}{T_{t-1,t-1}}}, \qquad L_t = L_{t-1} d_t. \tag{8.14}$$

Thus a common proportional change in all hourly wages does not itself change the quantity index, while changes in worker numbers, characteristics and hours do. The regressions are re-estimated within each scenario's simulated population. If the growth calculation yields a nonpositive or nonfinite total or factor, or $|\log d_t| > 0.25$, its labor-growth contribution is set to zero ($d_t = 1$). Other missing-design or empty-fit conditions stop the calculation. The labor growth feeds the earnings adjustment index alongside the exogenous macro inputs described in the aggregate-accounting section.

# 9 Macro paths and reported economic quantities

## 9.1 Prices, productivity, and the earnings scale

Let $\pi_t$ be supplied log price growth, $g_t^A$ supplied log multifactor productivity growth, $g_t^{K/L}$ supplied log capital–labor-ratio growth, and $\alpha_t$ and $\beta_t$ supplied capital and labor elasticities. These are separate series; $\beta_t$ is not defined as $1 - \alpha_t$. The price index has $P_b = 1$ and

$$P_t = P_{t-1} e^{\pi_t}. \tag{9.1}$$

For dates before the base year, use $P_{t-1} = P_t e^{-\pi_t}$. The price-input domain is 1930–2175 for growth, yielding levels from 1929–2175. The earnings index starts at the supplied historical value $E_b$ for a base year in 1993–2024. Its subsequent recursion is

$$E_t = E_{t-1} \exp\{\pi_t + g_t^A + \alpha_t[\Delta \log L_t + g_t^{K/L}]\}, \tag{9.2}$$

where $\Delta \log L_t$ is the in-year labor-input growth specified in the earnings section. Arrivals' first-year conversion uses the same recursion with the current labor term set to zero and the preceding year's stored $E_{t-1}$. The externally supplied index $E_t^0$ uses a zero labor term in every year, anchored at $E_b$; it is distinct from $E_t$.

The growth inputs for the earnings recursion extend through 2174. A projection requires every intervening year's growth terms and the base-year historical anchor. There is no terminal extrapolation beyond that domain. Demographic tables instead retain their last supplied year when their mechanism-specific lookup rules prescribe it. These different horizon conventions matter when a common scenario horizon is selected.

## 9.2 Counting and expansion

Let $n_t = |\mathcal{P}_t|$ and let $D_t$ and $M_t$ be the numbers who die and emigrate during year $t$. Let $B_{t+1}$ and $I_{t+1}$ be births and arrivals in the next year. The full-year counting identity is

$$n_{t+1} = n_t - D_t - M_t + B_{t+1} + I_{t+1}. \tag{9.3}$$

Death and emigration are mutually exclusive departures. $n_t$ is a simulated record count, not a resident population at December 31. The latter would exclude that year's departures.

The microeconomic export assigns one expansion weight to all records and years:

$$W = \text{round}_{\text{even}} \left( \frac{\sum_{j \in \mathcal{D}_b} w_j}{n_b} \right), \tag{9.4}$$

where $\mathcal{D}_b$ is the raw base-year donor population, $w_j$ its person survey weight, and rounding chooses the nearest integer with ties to even. An expanded total for a person-level quantity $x_{it}$ is $W \sum_{i \in \mathcal{P}_t} x_{it}$; a weighted mean under this common weight equals the unweighted mean. The aggregate output interface reports the unexpanded counts and sums described here. Indices are normalized quantities, not national dollar levels. The distinction lets a downstream model use growth without mistaking the sample's dollar sum for national receipts.

Employment counts people whose current work status is employed. Private employment further selects unincorporated self-employment, incorporated self-employment, or private wage and salary work. Wages and salaries sum nominal annual earnings for employed private, federal, and state/local wage and salary workers; private wages select private wage and salary workers only. Each person's dollar amount is truncated toward zero before summation. Private hours sum hours worked by the three private classes, with no separate employment filter. The year's deaths, emigrants, and arrivals remain in these summation populations.

## 9.3 Aggregate labor, capital, and private output

For aggregate reporting, the labor-input calculation uses successive full-year populations. The married indicator includes a union active at year-end or ended by a death during the year; children are counted through year-end biological-parent links. Denote the resulting log growth by $\ell_t$ to distinguish it from the within-year growth in equation (9.2). The same degenerate-growth hold rule described in the labor-input mechanism applies. Let $\mathcal{L}_t$, $\mathcal{K}_t$, $A_t$, $Y_t$ and $Q_t$ denote reported labor input, capital services, multifactor productivity, real private output, and labor efficiency, respectively. Their base values are one, and

$$\mathcal{L}_t = \mathcal{L}_{t-1} e^{\ell_t}, \qquad \mathcal{K}_t = \mathcal{K}_{t-1} e^{\ell_t + g_t^{K/L}}, \tag{9.5}$$

$$A_t = A_{t-1} e^{g_t^A}, \qquad Q_t = Q_{t-1} \frac{e^{\ell_t}}{H_t^p / H_{t-1}^p}, \tag{9.6}$$

$$Y_t = Y_{t-1} \exp\{g_t^A + \beta_t \ell_t + \alpha_t (\ell_t + g_t^{K/L})\}. \tag{9.7}$$

Here $H_t^p$ is private hours, which must be positive in both years for its growth rate to be defined. Nominal private output is $P_t Y_t$. The capital path therefore expands with labor input and an external capital–labor-ratio path. These equations are a growth-accounting closure, not an endogenous saving or investment decision. The reported earnings index is the in-year index $E_t$ used for continuing people's nominal earnings, with the arrival-specific conversion described above. It is not recomputed from the separately measured reporting growth $\ell_t$.

The EconomicProjections interface uses private nominal and real output, wages and salaries, multifactor productivity, employment, and population. Its dollar levels and price measures are downstream constructs; the normalization here specifies the growth inputs delivered to it.

## 9.4 Fertility, mortality, and period life tables

Current-population mortality statistics are arithmetic means of the assigned mortality hazards over all represented people, those below 65, and those aged 65 or above. They are not the realized number of deaths divided by the population. A mean over no records is defined as zero. Total fertility is

$$\mathrm{TFR}_t = \sum_{a=15}^{49} \frac{\sum_{i \in \mathcal{P}_t : i \text{ female}, a_{it} = a} \lambda_{it}}{\#\{i \in \mathcal{P}_t : i \text{ female}, a_{it} = a\}}, \tag{9.8}$$

where $\lambda_{it}$ is the assigned birth intensity. Every age from 15 through 49 must be represented. Arriving women have zero arrival-year intensity and remain in the denominator. Completed fertility is mean prior births among women aged 49, after the current year's birth accounting. All vital-statistics values are zero by convention in the base year.

For sex $s$ and represented age $a < 96$, let $\overline{m}_{sat}$ be the mean assigned mortality hazard; a third series pools sexes before averaging. The period life table defines annual survival as

$$p_{sat} = 1 - \frac{\overline{m}_{sat}}{1 + \overline{m}_{sat}/2}. \tag{9.9}$$

This reporting transformation differs from the exponential hazard-to-death conversion used for simulated deaths. If $a_s^*$ is the oldest represented age below 96, extend the table through 120 by $p_{sat} = p_{sa_s^* t}(120 - a)/(120 - a_s^*)$ for $a \geq a_s^*$. Thus survival reaches zero at 120. A sex must have at least one observation below 96. Missing ages below $a_s^*$ are skipped, rather than interpolated.

The output reports expected *age at death*, at birth and at age 65. For a complete age grid and base age $a_0 \in \{0, 65\}$, it is

$$\mathcal{E}_{s,t}(a_0) = \sum_{a=a_0}^{120} a(1 - p_{sat}) \prod_{k=a_0}^{a-1} p_{skt}. \tag{9.10}$$

With a missing-age grid, use the sorted observed ages plus the extrapolated tail. Shift survival by one row on that complete table before selecting ages at least $a_0$, replace the previous-survival factor by one at an observed $a_0$, and cumulatively multiply the remaining previous-row factors. An absent previous-row factor contributes no term and leaves the cumulative product unchanged. This specifies the sparse-table convention as well as the economically usual complete-grid case. Subtracting 65 from the age-65 output gives remaining years of life when the full conditional life table is represented.

# A Implementation components

## A.1 Assembly, inputs and aggregate outputs

**Model boundary and selected settings.**
The model follows people and their family relationships through annual changes in population, schooling, work and earnings. Its components combine household sampling with conditional demographic and economic transitions. Scenario assumptions specify the demographic rates, composition targets and external economic paths used in those transitions.

**Empirical inputs.**
Augmented survey records supply the households used to construct the initial population and immigrant arrivals. Rate tables, fitted statistical models and earnings-process parameters determine subsequent outcomes. Each component applies these inputs to the relevant population and conditioning information.

**Prices and the earnings scale.**
External paths for prices, productivity and the capital–labor ratio establish the aggregate economic environment. Growth in the simulated labor input also affects the scale used to convert individual real earnings into nominal amounts. Those nominal earnings enter later household and schooling decisions.

**Population totals and aggregate reporting.**
Individual records supply population, employment, hours and earnings totals. The reporting calculations construct labor, capital and private-output indices for use in economic projections. A common expansion weight relates the exported individual records to the size of the base-year donor population.

**Demographic statistics.**
Assigned fertility intensities and mortality hazards provide measures of the population's demographic conditions. Period life tables translate the mortality schedule into expected ages at death. These summaries describe the annual population and rates rather than following one birth cohort over its lifetime.

## A.2 Population, vital events and migration

**Initial population and calibration.**
Household sampling preserves relationships among people in the same donor household. Calibration adjusts household selection probabilities to reconcile the donor population's demographic shares with the sampling distribution. Repeated household draws create separate families in the simulated population.

**Births, survival and population continuity.**
Births create new people with parental links and starting demographic attributes. Mortality and emigration determine who remains in the following year's population. Records retain events and earnings for the year in which a person departs, allowing annual totals to include that person's contribution.

**Immigrant arrivals and legal status.**
Arrival rates determine the scale of immigration, while household donor pools and composition targets

determine the characteristics of arriving families. Legal-status transitions affect subsequent naturalization, overstay and emigration outcomes. Arrivals bring education and work histories into the population before joining the continuing-person transitions in later years.

**State location.**
Geographic transitions assign a destination to household reference persons using their prior state and demographic characteristics. Other household members follow the resulting residence assignment. This treatment preserves dependence among family members' locations while allowing changes in household composition to alter a person's state.

### A.3 Partnerships, residences and tax units

**Partnership formation and dissolution.**
The partnership market brings eligible women and men together through meetings and conditional acceptance decisions. Existing partnerships can dissolve, and cohabitation status distinguishes unions for mechanisms that require legal marriage. Partnership histories connect demographic events to schooling, employment and household formation.

**Living arrangements and family links.**
Parental-home departures, new partnerships and the loss of family members change residential arrangements. The residence component places people using their partner and parent relationships. Biological kinship, shared residence and the child indicators used in fitted relationships remain distinct concepts.

**Tax units and starting benefit information.**
Tax-unit assignment groups people using residence, partnership, earnings and dependency conditions. The resulting units provide inputs for tax analysis; the assignment itself does not calculate liabilities. Donor Social Security receipt and income attributes supply starting information for linked analysis.

### A.4 Education, work and earnings

**Family background and schooling.**
Parental attainment, childhood household circumstances and the score assigned at age 17 help determine educational transitions. Qualification and enrollment decisions update a person's schooling history. The resulting attainment affects later work, earnings and demographic outcomes.

**Disability, employment and hours.**
Work disability and fitted labor transitions determine participation, employment class, full-time status, hours and occupation. Current family circumstances and prior work histories condition these outcomes. Employment and hours then define the population and work exposure used in the earnings calculation.

**Earnings histories.**
Predictable earnings reflect personal characteristics and work circumstances. Persistent individual differences and serially dependent shocks generate variation across otherwise similar workers and continuity within each person's history. Annual hours and the aggregate earnings scale translate these components into real and nominal earnings.

**Aggregate labor input.**
The labor-input calculation values workers' hours using fitted hourly earnings relationships from successive years. Its chained quantity index captures changes in the workforce's size, composition and hours. The within-year index feeds the earnings scale, while full-year reporting supplies the labor measure used in aggregate production accounting.