

Populace dynamics: an open, scored longitudinal layer for policy microsimulation

Max Ghenis

July 2026

Abstract

Dynamic microsimulation — the longitudinal aging of a person-level population through earnings, family structure, disability, mortality, and program participation — underpins retirement and social-insurance policy analysis, yet the benchmark United States models — DYNASIM, MINT, and CBOLT — are closed: internal to government, tied to restricted administrative records, or accessible only through institutional relationships. This design paper specifies an open alternative built as an extension of Populace, PolicyEngine’s open-source microdata stack, whose cross-sectional layer is the certified default United States microdata in PolicyEngine after outperforming its predecessor on held-out administrative targets. The design contributes four elements: a trajectory-weighted kernel in which multi-period calibration cannot silently destroy panel structure; a Dynamics operator that treats state transitions as conditional models, mixing deterministic demographic hazards with machine-learned earnings processes; an explicit domains-of-validity framework that refuses point forecasts where parameter uncertainty dominates — including the 75-year actuarial balance — and instead publishes sensitivity surfaces; and a scoring protocol under which every claim resolves against administrative publications, backtests with leakage control, or computes exactly from statute, with contributions merging only when they improve held-out scores. United States Social Security is the first validation domain; the layer itself is country-agnostic.

1 Introduction

Analysts who want to model taxes can run open, calibrated models directly: Tax-Calculator (Policy Simulation Library 2026) and PolicyEngine (PolicyEngine 2026) are openly callable on public data, alongside source-available models with restricted inputs (The Budget Lab at Yale 2026) and proprietary models used for outside-facing analysis (Tax Policy Center 2025; Institute on Taxation and Economic Policy 2025; Tax Foundation 2025; Penn Wharton Budget Model 2025). Analysts who want dynamic microsimulation — the longitudinal modeling that retirement and social-insurance policy requires — find the benchmark models closed: SSA projects retirement income and the distributional effects of policy proposals with MINT (Social Security Administration 2024), CBO produces its long-term Social Security projections with CBOLT (Congressional Budget Office 2018, 2024), the Urban Institute projects retirement income and long-term care with DYNASIM (Favreault et al. 2015; Urban Institute 2024), and Morningstar studies retirement adequacy with its Model of US Retirement Outcomes (Look and VanDerhei 2024). Outside users reach each only through an institutional relationship. The nearest open analogue, the Cato Social Security model (Chanwong 2026), simulates roughly 10,000 households from the 2007 CPS ASEC under the Social Security Administration’s assumptions and reports trust-fund metrics and reform scores, without published validation against administrative benchmarks.

This paper specifies the design of an open longitudinal layer — *Populace dynamics* — as an extension of Populace, PolicyEngine’s open-source microdata stack. It is a design paper: the cross-sectional

foundation is in production; this paper specifies the longitudinal layer, which the project builds in the open behind the scoring protocol of Section 5.

The design makes four contributions:

1. **A trajectory-weighted kernel with explicit alignment.** One weight per trajectory: multi-period calibration stacks constraint rows over the same weight vector, so that hitting cross-sectional totals in multiple periods cannot silently destroy panel structure — combined with event-selection alignment for period-by-period control, since weights alone cannot separate a correct life course from a correctly timed one (Section 4).
2. **Transitions as conditional models.** A single operator interface covers deterministic demographic hazards and machine-learned earnings processes, so candidate architectures compete under one evaluation standard (Section 4).
3. **Domains of validity as shipped metadata.** The model states which questions it will not answer with false precision — led by the 75-year actuarial balance, where input uncertainty dominates model fidelity — and publishes long-horizon results as sensitivity surfaces rather than point forecasts (Section 3).
4. **A scoring protocol in place of fidelity-only validation.** Claims resolve against administrative publications on an annual calendar, backtest against realized history with leakage control, or compute exactly from statute; contributions merge only when they improve held-out scores (Section 5).

United States Social Security is the first validation domain because benefit adequacy and reform incidence depend jointly on lifetime earnings, marriage and survivorship, disability, differential mortality, and claiming. The layer itself is country-agnostic and extends to other pension and benefit systems as PolicyEngine’s country coverage grows.

2 Terminology and scope

This paper uses “dynamic” in the microsimulation field’s standard sense, following Orcutt et al. (1961) and the tradition carried by DYNASIM, MINT, CBOLT, and SimPaths (Bronka et al. 2025): a longitudinal model that ages a person-level population through time (Li and O’Donoghue 2013). It does not mean “dynamic scoring” — the tax-policy usage denoting macroeconomic feedback in revenue estimation — and the design is not an overlapping-generations general-equilibrium model of the Auerbach–Kotlikoff kind, such as the Penn Wharton Budget Model operates (Penn Wharton Budget Model 2025). The model claims neither macroeconomic feedback nor equilibrium closure. Behavioral responses enter as labeled scenario inputs with documented ranges, not as point estimates carrying model authority.

3 Domains of validity

All models are wrong; a model earns its keep only where it improves predictions. The design therefore begins by bounding its own claims.

3.1 Parameter uncertainty dominates the long horizon

The 75-year actuarial balance of a pension system is a function of a small set of exogenous assumptions — fertility, mortality improvement, net immigration, real wage growth, interest rates — whose uncertainty dominates the microsimulation machinery that processes them. The public record makes this concrete for United States Social Security. The Trustees’ own low- and high-cost scenarios bracket a range of 75-year balances wider than the intermediate deficit itself (Board of Trustees, Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds 2025). The Congressional Budget

Office’s long-term projections differ from the Trustees’ — chiefly for assumption reasons, though method differences such as CBO’s micro-founded projection of the taxable share of earnings also contribute to the divergence (Congressional Budget Office 2024). Assumption variance dominates the headline; model structure matters for distribution and the near term, which is the division of labor the output tiers encode. Successive Technical Panels convened by the Social Security Advisory Board have recommended revising fertility and mortality-improvement assumptions as realized values ran persistently outside the projected path (Technical Panel on Assumptions and Methods 2023). Two institutions with administrative data and decades of refinement disagree with each other and have both missed realized demographic trends; a better microsimulation does not repair that, because the variance lives in the inputs.

3.2 Three output tiers

The model sorts its outputs into tiers, each carrying the strongest claim it can support.

Tier 1: distributional analysis under fixed assumptions. Reform analysis is a difference — outcome under reform minus outcome under baseline, holding the population and assumption path fixed — and much of the unforecastable demographic uncertainty is common to both arms and cancels in the difference. The required ingredients — a calibrated joint distribution of lifetime earnings, family structure, and differential mortality, plus an exact rules engine — are the components the scoring protocol validates directly. The design labels rather than buries the slice of a reform delta that does not cancel: reform-induced claiming responses, and interactions between the reform and uncertain dynamics — the mortality gradient enters many deltas as a covariance with benefit position, not a level, so it does not difference out. Deltas evaluated past trust-fund depletion also require an explicit scheduled-versus-payable baseline convention, which the model states with every such output. For claiming, scenario ranges anchor to the quasi-experimental record on retirement-age responses (Mastrobuoni 2009; Behaghel and Blau 2012), and the model publishes them as a scenario library rather than embedding them as point estimates.

Tier 2: near-term components that resolve. Over roughly a ten-year horizon, mechanics rather than demographic extrapolation dominate outputs: beneficiary counts by type, average benefits, covered earnings and taxable payroll, claiming-age distributions, disability incidence. These resolve against administrative publications each year, and the protocol scores them (Section 5).

Tier 3: the long horizon as a sensitivity surface. The model publishes long-horizon outputs as surfaces over documented assumption ranges — how the balance, cohort replacement rates, or distributional outcomes move as fertility, mortality improvement, and immigration vary — never as point forecasts. The distinction is between computing and blessing: the model computes 75-year balances and depletion dates *conditional on named assumption paths*, including the Trustees’ intermediate path, so the numbers the policy debate runs on remain available as labeled conditional outputs; what the model declines is presenting any single path as its own forecast. Incumbent practice publishes the point estimate up front and the sensitivity analysis in an appendix; an open model can invert that and make the sensitivity the interface.

Every API response carries its tier, assumption path, and calibration history as metadata, so a downstream consumer — human or machine — weights the output by demonstrated reliability rather than by the producer’s reputation.

4 Architecture

4.1 The cross-sectional foundation

Populace builds a calibrated synthetic population entirely from primary-source government data — the Current Population Survey ASEC, the IRS Public Use File, the Survey of Consumer Finances, SIPP, CPS outgoing-rotation groups, MEPS, and the ACS — synthesizing missing variables with weight-aware conditional models and calibrating to administrative aggregates treated as uncertainty-weighted facts. In June 2026 it replaced PolicyEngine’s enhanced CPS as the certified default United States microdata in PolicyEngine, after a matched, symmetric-refit comparison on 41,314 households with a 739-target holdout (Table 1).

Table 1: Certification comparison from the Populace release manifest. The enhanced CPS wins more individual targets by small margins while its largest misses dominate the loss; the scoring protocol requires publishing the count that cuts against the headline.

Metric (lower is better)	Populace	enhanced CPS
Holdout loss (739 held-out targets)	0.038	0.317
Training loss	0.190	1.089
Full loss	0.228	1.405
Per-target wins	1,040	2,613 (51 ties)

4.2 Trajectory weights and population accounting

The longitudinal extension follows two kernel rules set in Populace’s charter. First, **one weight per trajectory**: multi-period calibration targets stack as (target, period) constraint rows over a single trajectory-level weight vector, so that calibrating cross-sections independently — which severs the trajectory-level consistency a panel exists to provide — is a kernel-level error rather than a modeling temptation. Second, **population is not closed**: trajectories carry entry and exit markers (birth, death, immigration, emigration), and a trajectory’s weight contributes to a period only while the person is present. Household and couple links are period-scoped, so family recomposition preserves per-period accounting identities; couples carry a shared unit weight derived from their trajectory weights so that spousal and survivor benefits have a well-defined representation.

Weights alone are one layer of alignment, not the whole answer: a weight cannot distinguish a correct life course from a correctly timed one, and reweighting trajectories to hit future cross-sectional cells risks selecting on entire correlated life courses. Period-by-period control therefore also operates through event selection — ranking individual transition probabilities and selecting the number of events an external control demands, the mechanism CBOLT and DYNASIM use — with trajectory weights reserved for base-year representation and slow-moving composition (Li and O’Donoghue 2013; Dekkers and Cumpston 2012). The kernel solves the stacked constraint system as uncertainty-weighted penalized least squares against target standard errors, so infeasible combinations resolve by SE-weighted compromise rather than silent failure, and projected demographic controls pin cohorts born after the base year rather than leaving them free.

4.3 The Dynamics operator

Dynamics is an operator from a population and a transition specification to a population with extended periods. A transition is a conditional distribution — the probability of next-period state given current state and covariates — which is the same interface Populace’s synthesis models already implement. The shipped baseline for earnings is a regime-gated, sequentially chained, weighted quantile-regression-forest imputer (Meinshausen 2006) whose zero-inflation gate doubles as a nonemployment model; richer

architectures (zero-inflated neural distribution models, normalizing flows) are candidates that must beat the baseline on held-out longitudinal moments to merge.

The design is hybrid. Where the evidence base is tabular — mortality from official life tables with published income gradients, fertility from vital statistics, marriage and divorce from ACS- and CPS-based rates (federal collection of detailed marriage and divorce statistics ended in the 1990s), disability incidence from program statistics — transitions are deterministic hazards, auditable row by row. Marriage requires a matching model, not only a hazard: spousal and survivor benefits depend on assortative matching over lifetime earnings, which the design treats as a first-class estimation target rather than an afterthought. The design reserves machine learning for processes with rich conditional structure, led by earnings dynamics — where the process is not stationary: volatility and mobility vary by cohort and period in administrative data (Sabelhaus and Song 2010; Kopczuk et al. 2010), so transition models carry cohort and period conditioning rather than pooling across decades. Chained one-period models also understate long-spell persistence unless spell structure enters the model explicitly, and backcasting is a distinct conditional object from forward simulation rather than the same operator reversed; both enter the evaluation as targets in their own right, disciplined by held-out panel moments such as higher-order earnings-change distributions (Guvenen et al. 2021).

The design states one measurement caveat rather than hiding it: the most demanding earnings-dynamics moments come from administrative records that public panels understate, and survey- and administrative-based estimates disagree on volatility levels and trends. Where the project cannot recompute a published administrative moment on held-out public data, matching it is calibration, not validation, and the scorecard labels it as such.

4.4 Rules and delivery

Statute is the deterministic slice of any policy forecast. Core retirement benefit formulas — average indexed monthly earnings, primary insurance amounts, actuarial adjustments — and benefit taxation compute exactly today through PolicyEngine’s rules engine via Populace’s rules-adapter protocol, vectorized over person-periods. Auxiliary, spousal, and survivor benefit formulas are explicit build items in the validation program: the current engine carries them as calibrated aggregates rather than person-level formulas, and the scorecard treats “computes exactly” as a per-rule status each formula earns, not a blanket claim.

The rules adapter is engine-agnostic. PolicyEngine-US implements it today; Axiom — an open project that encodes statute declaratively and compiles it to Rust — is the next adapter, and the performance headroom matters when benefit formulas run over person-periods across the full trajectory panel and many reform scenarios. In that architecture, PolicyEngine is a composition: Axiom supplies the rules, Populace supplies the population, and behavioral responses enter as the labeled scenario layer.

Data governance is a design requirement, not an afterthought. Estimating transition models on restricted-use panels and publishing the estimated parameters is settled practice — open models such as OG-USA, and DYNASIM itself, estimate on the PSID and publish what they learn. Releasing a synthetic *microdata* artifact informed by such panels is a stricter problem, because donor-based samplers can emit observed training values. The design answers it structurally: released records originate from Populace’s public-use cross-section, restricted panels train processes rather than donate records, samplers smooth or noise their draws so no verbatim donor value ships, and every release passes nearest-neighbor disclosure checks alongside its accuracy scorecard — and the project engages data producers directly where their terms of use warrant it. Uncertainty budgets therefore attach only to the components statute does not fix. The deliverable is a versioned artifact — a longitudinal population release with a manifest and scorecard, certified through the same path as the cross-sectional release — exposed through a Python library, a REST API, and a Model Context Protocol server so that AI agents can run baseline distributions and reform analyses with validity metadata attached.

5 Scoring and resolution

Validation by fidelity — does the model match published aggregates? — is necessary but weak: a model can reproduce the tables its authors fit it to. This project’s standard: a claim counts as validated when it improves prediction of something that later resolves. Five scoring surfaces implement it.

1. **Annually resolving components.** Beneficiary counts by type, average and aggregate benefits, covered earnings and taxable payroll, cost-of-living adjustments, disability incidence, and claiming-age distributions resolve against administrative publications each year. Every published forecast cell carries a resolution rule naming the exact table and vintage that settles it.
2. **Forecasting the forecasters.** Official projections revise every year; predicting the next revision of headline quantities resolves in months rather than decades and is decision-relevant to anyone who acts on the official number. This is partly forecasting an assumptions process — panels advise, committees adopt with a lag — so each cell pre-specifies the naive baseline it must beat (an assumption random walk plus mechanical data update).
3. **Retrodiction with leakage control.** Retrodiction builds the model from data vintages available at a historical date and scores it against realized outcomes. Populace’s versioned data registry pins vintages going forward; pre-registry history is a reconstruction problem — survey redesigns, revised administrative tables, re-released panels — so the protocol grades pre-registry backtests as pseudo-vintage, with a published log of deviations from true vintage, and no registry pins specification leakage: a “2005-vintage” model built today knows 2008 happened, and the protocol says so. Retrodictive calibration under the historical regime does not guarantee calibration under a new one, so backtests complement rather than substitute for live resolution.
4. **Statutory resolution.** Where an output is fixed by law, the rules engine computes it exactly, and enacted policy settles the corresponding conditional cells immediately.
5. **Held-out panel moments.** The protocol scores the population layer against moments it never fit: earnings-mobility matrices, autocorrelation and higher-order moments of earnings changes, cohort age-earnings profiles, and family-transition rates on held-out panel records.

Two governance rules complete the protocol. **Merge on score:** a contribution — a mortality module, a claiming model, an earnings architecture, from any contributor — merges if and only if it improves the population’s score on held-out facts, the rule Populace already applies to its cross-sectional layer. **Publication discipline:** misses publish with the same prominence as hits; superseded methods keep their historical scorecards; and stage gates in the development roadmap are pre-specified score thresholds, not narrative judgments.

Openness supplies most of the refereeing. Resolution rules are pre-registered, scores recompute from public data, and anyone who distrusts a published scorecard can rerun it — or fork the project and publish a rival scorecard. Two pieces sit beyond reproduction. First, restricted-data checks: the most demanding earnings-history moments live in linked administrative records that the public pipeline never touches, so the reader must trust whoever runs the comparison — author or outsider — and a validator without a stake in the result adds evidence where rerunning is impossible. Second, judgment: gate thresholds, disputed resolution rules, and the assumptions library carry discretion that pre-registration narrows but does not remove. The project defines its validation procedures and gate thresholds before the components they gate, and seats an advisory board to review them — and disputes under them — in public. Independent scoring across models is a decision only a third party can make; the design encourages it — published projections from the closed models give any such body a comparison set without model access — but does not depend on it.

6 First validation domain: U.S. Social Security

Social Security is the first domain because eligibility and benefits depend on the highest thirty-five years of indexed earnings, marital and survivorship histories, disability pathways, differential mortality,

and claiming timing — jointly. A layer that scores well here earns reuse in adjacent domains — Supplemental Security Income interactions, long-term care, and retirement adequacy, the last of which is *harder*, not easier: wealth projection challenged even administrative-data models (Favreault and Smith 2016), and a wealth and pension roadmap is future work, not an assumed extension — and in other countries’ pension systems.

The domain also makes the validity framework concrete. The canonical output of existing Social Security models — the 75-year balance and depletion date — is the quantity Section 3 declines to forecast. What the open layer offers instead is the combination the field lacks: tier-1 distributional incidence of reforms with reproducible assumptions; tier-2 near-term components with a public resolution record; and tier-3 sensitivity surfaces that make the assumption-dependence of long-horizon claims the interface rather than the appendix. No existing model, closed or open, publishes that combination. The closed benchmarks do publish validation and cohort tables; what outsiders cannot do is rerun the pipeline, vary its assumptions, or audit the intermediate states behind the published numbers.

7 Related work

Dynamic microsimulation originates with Orcutt et al. (1961); Li and O’Donoghue (2013) survey the field’s alignment and calibration practice. This design borrows the central structural lesson of DYNASIM (Favreault et al. 2015; Urban Institute 2024), MINT (Social Security Administration 2024), and CBOLT (Congressional Budget Office 2018) — an annual state engine with family links and external alignment — while departing on openness, trajectory-level weighting, and scoring; Dekkers and Cumpston (2012) treat the weighting question directly. The Cato model (Chanwong 2026) is the nearest open system.

The pattern repeats internationally — Pensim2 at the United Kingdom’s Department for Work and Pensions, MOSART at Statistics Norway, MIDAS at Belgium’s Federal Planning Bureau (Li and O’Donoghue 2013) — with exceptions. INSEE has published the source of Destinie 2 (Blanchet et al. 2010), the pension model behind France’s official projection exercises (github.com/InseeFr/Destinie-2). SimPaths, from the University of Essex’s Centre for Microsimulation and Policy Analysis, is an open-source life-course model estimated for the United Kingdom and being adapted to several other European countries (Bronka et al. 2025). And open frameworks exist without open populations: LIAM2 — built at the same Federal Planning Bureau that runs MIDAS (Menten et al. 2014) — and OpenM++, an open reimplement of Statistics Canada’s Modgen platform (OpenM++ development team 2026), supply generic simulation engines but ship no calibrated population and no rules stack. On OpenM++, WIFO’s microWELT models welfare transfers comparatively across Austria, Spain, Finland, and the United Kingdom, with a United States variant for labor-force projection (Spielauer et al. 2020) — one open engine carrying a multi-country dynamic model. None of these combines an open codebase with a certified calibrated microdata baseline and a public scoring protocol under which claims resolve. SimPaths is closest on openness; the difference here is inheritance — a certified cross-sectional population, one rules platform across countries, and credibility staked on resolved scores rather than publication.

On the measurement side, administrative-data studies of earnings dynamics discipline the earnings process: lifecycle moments (Guvenen et al. 2021), long-run mobility (Kopczuk et al. 2010), non-stationary volatility (Sabelhaus and Song 2010), and the attenuating link between current and lifetime earnings (Haider and Solon 2006) — with nonlinear panel frameworks (Arellano et al. 2017) the academic cousin of the quantile machinery used here. The quasi-experimental record on claiming responses (Mastrobuoni 2009; Behaghel and Blau 2012) anchors the behavioral scenario library, and the assumption-failure record compiled by successive Technical Panels (Technical Panel on Assumptions and Methods 2023) motivates the domains-of-validity framework.

8 Status and roadmap

The cross-sectional foundation is in production. Populace’s charter specifies the longitudinal kernel rules. This paper is the project’s front door; the supplementary design appendices — operational chapters on earnings-history construction, family and auxiliary benefits, disability and claiming, mortality and projection drift, calibration targets, the Social Security validation program, and a source-based DYNASIM dossier — live in the project repository at github.com/PolicyEngine/populace-dynamics. Development proceeds through stage gates defined as score thresholds — earnings-history credibility first, family and benefit outputs second, forward projection third, productization last — and every gate’s evidence publishes whether it passes or fails.

The layer is open source under the MIT license. The project invites corrections, particularly to the benchmark characterizations, and contributions merge on the same standard as the authors’: improve the held-out score.

References

- Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme. 2017. “Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework.” *Econometrica* 85 (3): 693–734.
- Behaghel, Luc, and David M Blau. 2012. “Framing Social Security Reform: Behavioral Responses to Changes in the Full Retirement Age.” *American Economic Journal: Economic Policy* 4 (4): 41–67.
- Blanchet, Didier, Sophie Buffeteau, Emmanuelle Crenner, and Sylvie Le Minez. 2010. *The New Destinie 2 Microsimulation Model: Main Characteristics and Illustrative Results*. Document de Travail Nos. G2010-13. INSEE.
- Board of Trustees, Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds. 2025. *The 2025 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds*. Social Security Administration. <https://www.ssa.gov/oact/TR/2025/tr2025.pdf>.
- Bronka, Patryk, Justin van de Ven, Daniel Kopasker, S. Vittal Katikireddi, and Matteo Richiardi. 2025. “SimPaths: An Open-Source Microsimulation Model for Life Course Analysis.” *International Journal of Microsimulation* 18 (1): 95–133. <https://doi.org/10.34196/ijm.00318>.
- Chanwong, Krit. 2026. *Social Security Cato Model*. https://github.com/kchanwong/social_security_cato_model.
- Congressional Budget Office. 2018. *An Overview of CBOLT: The Congressional Budget Office Long-Term Model*. Congressional Budget Office. <https://www.cbo.gov/publication/53667>.
- Congressional Budget Office. 2024. *CBO’s 2024 Long-Term Projections for Social Security*. Congressional Budget Office. <https://www.cbo.gov/publication/60392>.
- Dekkers, Gijs, and Richard Cumpston. 2012. “On Weights in Dynamic-Ageing Microsimulation Models.” *International Journal of Microsimulation* 5 (2): 59–65.
- Favreault, Melissa M, Karen E Smith, and Richard W Johnson. 2015. *The Dynamic Simulation of Income Model (DYNASIM): An Overview*. The Urban Institute. <https://www.urban.org/research/publication/dynamic-simulation-income-model-dynasim>.

- Favreault, Melissa, and Karen E Smith. 2016. *The Accuracy of MINT Wealth Projections*. Urban Institute.
- Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2021. “What Do Data on Millions of U.S. Workers Reveal about Lifecycle Earnings Dynamics?” *Econometrica* 89 (5): 2303–39.
- Haider, Steven, and Gary Solon. 2006. “Life Cycle Variation in the Association Between Current and Lifetime Earnings.” *American Economic Review* 96 (4): 1308–20.
- Institute on Taxation and Economic Policy. 2025. *ITEP Tax Microsimulation Model Overview*. <https://itep.org/itep-tax-model>.
- Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. 2010. “Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937.” *The Quarterly Journal of Economics* 125 (1): 91–128.
- Li, Jinjing, and Cathal O’Donoghue. 2013. “Alignment and Calibration of a Dynamic Microsimulation Model.” *Journal of Artificial Societies and Social Simulation* 16 (3): 1–15.
- Look, Spencer U., and Jack VanDerhei. 2024. *Beyond the Retirement Crisis Headlines: Why Employer-Sponsored Plans Are the Key to Retirement Adequacy for Today’s Workers*. Morningstar Center for Retirement & Policy Studies. https://www.morningstar.com/content/cs-assets/v3/assets/blt9415ea4cc4157833/bltd4bb26598046aed4/66a1535de91a178e5c15872a/Introducing_the_Morningstar_Model_of_US_Retirement_Outcomes_-_July_2024_-_final.pdf.
- Mastrobuoni, Giovanni. 2009. “Labor Supply Effects of the Recent Social Security Benefit Cuts: Empirical Estimates Using Cohort Discontinuities.” *Journal of Public Economics* 93 (11-12): 1224–33.
- Meinshausen, Nicolai. 2006. “Quantile Regression Forests.” *Journal of Machine Learning Research* 7: 983–99. <https://jmlr.org/papers/v7/meinshausen06a.html>.
- Menten, Gaëtan de, Gijs Dekkers, Geert Bryon, Philippe Liégeois, and Cathal O’Donoghue. 2014. “LIAM2: A New Open Source Development Tool for Discrete-Time Dynamic Microsimulation Models.” *Journal of Artificial Societies and Social Simulation* 17 (3): 9. <https://doi.org/10.18564/jasss.2574>.
- OpenM++ development team. 2026. *OpenM++: Open Source Microsimulation Platform*. <https://openmpp.org>.
- Orcutt, Guy H, Martin Greenberger, John Korbel, and Alice M Rivlin. 1961. “Simulation of Economic Systems.” *The American Economic Review* 51 (5): 893–907.
- Penn Wharton Budget Model. 2025. *Penn Wharton Budget Model: Microsimulation*. <https://budgetmodel.wharton.upenn.edu/model/microsimulation/>.
- Policy Simulation Library. 2026. *Tax-Calculator*. <https://taxcalc.pslmodels.org/>.
- PolicyEngine. 2026. *PolicyEngine: Open-Source Tax-Benefit Microsimulation*. <https://policyengine.org>.

- Sabelhaus, John, and Jae Song. 2010. "The Great Moderation in Micro Labor Earnings." *Journal of Monetary Economics* 57 (4): 391–403.
- Social Security Administration. 2024. *Projection Methodology: Modeling Income in the Near Term, Version 8 (MINT8)*. <https://www.ssa.gov/policy/docs/projections/methodology.html>.
- Spielauer, Martin, Thomas Horvath, and Marian Fink. 2020. *microWELT: A Dynamic Microsimulation Model for the Study of Welfare Transfer Flows in Ageing Societies from a Comparative Welfare State Perspective*. WIFO Working Papers 609/2020. Austrian Institute of Economic Research (WIFO).
- Tax Foundation. 2025. *The Tax Foundation's Taxes and Growth Model*. <https://taxfoundation.org/research/all/federal/overview-tax-foundations-taxes-growth-model/>.
- Tax Policy Center. 2025. *Microsimulation Model FAQ*. <https://taxpolicycenter.org/resources/tax-model-resources/tpcs-microsimulation-model-faq>.
- Technical Panel on Assumptions and Methods. 2023. *2023 Technical Panel on Assumptions and Methods: Report to the Social Security Advisory Board*. Social Security Advisory Board. <https://www.ssab.gov>.
- The Budget Lab at Yale. 2026. *Tax-Simulator: Microsimulation Model of US Federal Tax System*. <https://github.com/Budget-Lab-Yale/Tax-Simulator>.
- Urban Institute. 2024. *Urban's Dynamic Simulation of Income Model 4 (DYNASIM4)*. <https://www.urban.org/research/publication/urbans-dynamic-simulation-income-model-4>.