The Economic and Fiscal Impact of STEM Immigration in General Equilibrium

Alexander Arnon[1], Duncan Haystead[1], Felix Reichling[1], German Sanchez Sanchez[1], Kent Smetters[1,2,3], and Jesús Villero[1]

[1]Penn Wharton Budget Model
[2]Wharton School of Business, University of Pennsylvania
[3]NBER

April 1, 2026

VERY PRELIMINARY – PLEASE DO NOT CITE.
Comments are welcome.

We study the economic and fiscal effects of STEM immigration using a stochastic, heterogeneous agent overlapping generations model with endogenous innovation, skill-specific native–immigrant complementarity, and a detailed fiscal apparatus—including Social Security, Medicare, Medicaid, and ACA subsidies—in which the government can run deficits and accumulate debt. We provide new empirical estimates of the STEM–TFP elasticity and native–immigrant substitution elasticities and embed them in the model. We analyze a proposal to exempt STEM immigrants from green card caps and find that it raises output by 3.4 percent by 2059, increases average labor earnings by 2.4 percent—with the largest gains accruing to low-educated workers—and reduces federal debt by 4.4 percent, while foreign-born STEM workers are the only group that experiences persistently negative effects.

Keywords: STEM immigration, overlapping generations, general equilibrium, total factor productivity, skill complementarity, fiscal effects

We thank participants in the Global Distribution, Migration, and Shocks session at the 2025 National Tax Association Annual Conference on Taxation for helpful comments. We are particularly grateful to Irina Popova for her insightful discussion of an earlier version.

Contents

1 Introduction

Workers in science, technology, engineering, and mathematics (STEM) occupations are central to the U.S. economy. STEM workers develop the technologies, processes, and products that drive productivity growth, and their contributions extend well beyond the firms that employ them: concentrations of skilled workers generate knowledge spillovers that raise the productivity of other workers in the local economy (Moretti 2004; Peri 2012). College-educated workers in STEM occupations account for approximately 12 percent of total U.S. employment and 29 percent of college-educated employment as of 2023, up from 7 percent and 27 percent, respectively, in 2000.[1] STEM employment nearly doubled over this period, far outpacing total employment growth of 17 percent. Foreign-born workers play a disproportionate and growing role in STEM fields: as of 2023, 23 percent of college-educated STEM workers are foreign-born, up from 18 percent in 2000. This foreign-born share substantially exceeds the corresponding shares in non-STEM college occupations (16 percent) and non-college employment (19 percent). The growth of foreign-born STEM employment over the period (144 percent) has also far outpaced that of native-born STEM employment (79 percent).

This paper develops a large-scale stochastic, heterogeneous-agent overlapping generations (OLG) model in general equilibrium, purpose-built to analyze some of the most pressing questions in immigration policy. The model incorporates several features that are essential for this analysis but are rarely combined in a single framework. On the production side, we adopt a nested constant elasticity of substitution (CES) labor aggregation structure that distinguishes workers by education level, STEM occupation, potential experience, and nativity. This nesting allows us to capture the distinct patterns of substitution and complementarity between native and foreign-born workers across skill groups. We further incorporate endogenous total factor productivity (TFP) driven by STEM employment, building on the approach of Peri, Shih, and Sparber (2015). On the household side, agents face idiosyncratic earnings and health shocks over their life cycle and make decisions

1. The figures are calculated from the American Community Survey (ACS). We define STEM occupations using the SOC Policy Committee crosswalk mapped to 2010 Census occupation codes. Our definition restricts STEM status to workers with at least a bachelor's degree. See Section 4 for details.

over consumption, labor supply, savings, health insurance, and medical treatment. The government operates a detailed fiscal apparatus that includes progressive federal income taxes, payroll taxes, Social Security (OASI), Medicare, Medicaid, Affordable Care Act (ACA) subsidies, and federal debt dynamics.

We apply this framework to analyze a proposal recently under consideration that would exempt STEM immigrants from the annual caps on employment-based permanent residence ("green cards"). Under current law, all employment-based categories are subject to an overall cap of approximately 140,000 visas per year and a per-country limit of 7 percent, generating severe multi-decade backlogs for applicants from high-demand countries. Exempting STEM workers would effectively remove these constraints for a class of immigrants with strong links to innovation and productivity and would change the composition of the U.S. immigrant population in ways that have implications for native and foreign-born wages, aggregate economic performance, the federal budget, and household welfare.

Our paper makes four main contributions. First, we provide new empirical estimates of the substitution elasticities in the nested CES structure, including the elasticity of substitution between native and foreign-born college-educated STEM workers. We estimate this elasticity at 5.33, substantially below the estimate of approximately 18 reported by Gunadi (2019) for all STEM workers. This finding implies considerably stronger complementarity between native and foreign-born STEM workers in roles that require a college education, meaning that admitting an additional college-educated foreign-born STEM worker generates larger productivity gains and smaller displacement effects than admitting workers in other skill categories. Second, we estimate the elasticity of TFP with respect to STEM employment at 0.19, consistent with the range of 0.22 to 0.23 reported for similar concepts by Gunadi (2019) and Bound, Khanna, and Morales (2017). We embed this channel in a general equilibrium framework, which is essential to fully capture how productivity spillovers feed back into wages, capital accumulation, and fiscal outcomes. Third, we are the first to combine the STEM–innovation channel with a full-scale heterogeneous-agent OLG model featuring detailed fiscal policy, allowing us to assess not only wage and output effects but also the fiscal and welfare consequences of STEM immigration policy over the life cycle. Fourth, we apply this framework to evaluate a specific, currently debated policy proposal, quantifying its effects on wages by worker type, macroeconomic aggregates, the federal budget, and household welfare.

We find that the Green Card Policy generates substantial macroeconomic gains over the projection period. Output rises by 3.4 percent relative to the baseline by 2059, driven by the combined expansion of labor inputs (1.7 percent), capital accumulation (3.5 percent), and total factor productivity (0.9 percent). The productivity channel operates through the rising share of STEM workers in total employment, which in turn stimulates capital accumulation as households respond to the permanently higher productivity path.

At the aggregate level, average labor earnings increase by 2.4 percent by 2059, but this headline figure masks considerable heterogeneity across skill and nativity groups. Low-educated workers experience the largest gains—reaching 3.3 percent for domestic and 2.4 percent for foreign workers—as their growing relative scarcity reinforces TFP-driven income effects. High-educated non-STEM workers realize more moderate gains that strengthen over the projection horizon as general equilibrium benefits diffuse through the economy. High-educated STEM foreign workers, by contrast, are the only group that experiences persistently negative effects on labor earnings, with losses ranging from 1.0 to 2.0 percent, reflecting the dominance of the labor substitution effect over TFP-driven gains for this group.

A decomposition of labor income into hours worked and earnings per hour reveals that virtually all of the distributional variation operates through the price margin. Changes in average hours worked are uniformly small across groups and display no systematic trend, consistent with the low Frisch elasticities embedded in the model. The distributional consequences of the policy thus arise almost entirely from wage adjustments rather than labor supply responses.

Consumption patterns broadly mirror the labor income results but reveal important additional dynamics. High-educated domestic STEM workers exhibit a notable divergence between their consumption and earnings trajectories: consumption effects are initially negative before rising to

2.4 percent by 2054, consistent with forward-looking households adjusting savings in anticipation of longer-run gains from capital accumulation. High-educated STEM foreign workers are the only group with persistently negative consumption effects, confirming that the adverse wage effects for this group translate directly into welfare losses that the general equilibrium benefits do not fully offset.

The policy also produces a favorable fiscal impact. Federal revenues rise by 2.7 percent above baseline by 2057, while outlays increase by only 0.4 percent, reflecting the predominantly working-age, high-education composition of the additional immigrants. The resulting improvement in the primary balance reduces debt held by the public by 4.4 percent relative to current law by the end of the projection period.

A decomposition by potential experience reveals that the aggregate effects for each group are not uniformly distributed across cohorts. Among foreign STEM workers, the largest earnings losses are concentrated among younger individuals in their 20s and early 30s, the cohorts most directly affected by the increased inflow. Among domestic STEM workers, younger cohorts benefit most from the strong labor complementarities between foreign and domestic STEM labor, particularly in the medium to long run. For high-educated non-STEM and low-educated workers, the absence of large compositional shifts within these groups implies that TFP gains are the dominant force, generating broadly uniform effects across experience levels.

The remainder of the paper is organized as follows. Section 2 reviews related literature. Section 3 describes the model. Section 4 presents our empirical estimates of the labor aggregation parameters and the TFP–STEM elasticity, as well as the demographic projections underlying our simulations. Section 5 evaluates the model's steady-state performance. Section 6 defines the policy experiment and Section 7 presents the results. Section 8 concludes.

2 Related Literature

Our paper relates to several different strands in the economics of immigration, including analysis of labor market effects, contributions to innovation and productivity, fiscal impact, and the role of general equilibrium effects. We discuss each in turn and describe how our work builds on and extends the prior research.

Immigration and native labor market outcomes. Modern analysis of immigration's labor market effects centers on the degree of substitutability between different types of workers, and between immigrants and U.S.-born workers in particular. Ottaviano and Peri (2012) develop a nested CES production framework and show that immigrants and natives are imperfect substitutes within education–experience cells. Their estimates imply that immigration from 1990 to 2006 had a small positive long-run effect on average native wages (approximately 0.6 percent). Caiumi and Peri (2024) extend this framework through 2022, introducing an improved shift-share instrument to measure exogenous foreign-born labor supply dynamics. They find positive effects of immigration on the wages of U.S.-born workers without a college degree (1.7 to 2.6 percent over 2000–2019) and no significant crowding out of native employment.

A key mechanism behind these outcomes is task specialization: Peri and Sparber (2009) show that immigrants concentrate in manual-task-intensive occupations while natives respond to immigration by shifting toward communication-intensive tasks, generating complementarity rather than direct competition. Peri (2012) finds that immigration is positively associated with state-level TFP growth, consistent with efficient task reallocation and the adoption of unskilled-efficient technologies. A meta-analysis by Havranek et al. (2024) finds a corrected mean elasticity of substitution between skilled and unskilled labor of approximately 4, with a lower bound of 2.

We adopt and extend the nested CES framework of Ottaviano and Peri (2012) and Caiumi and Peri (2024), adding an aggregation level that distinguishes STEM from non-STEM workers among the college-educated and estimating separate native–immigrant substitution elasticities for each skill group. This allows us to identify substantial heterogeneity in the degree of native–immigrant complementarity across skill groups—heterogeneity that is not visible in frameworks that pool all college-educated workers.

High-skilled immigration, STEM workers, and innovation. High-skilled immigrants contribute to U.S. innovation through channels beyond direct labor supply. Hunt and Gauthier-Loiselle (2010) show that immigrants patent at roughly twice the native rate, primarily because they disproportionately hold science and engineering degrees. Bernstein et al. (2022) find that immigrants constitute 16 percent of U.S. inventors but account for 36 percent of aggregate innovation, with two-thirds of this share attributable to positive spillovers on native collaborators. More broadly, Nathan (2013) argues that high-skilled immigration shapes productivity through entrepreneurship, trade, foreign direct investment, and knowledge spillovers in urban agglomerations.

To our knowledge, only two previous papers have directly estimated the link between STEM workers and productivity. Peri, Shih, and Sparber (2015) exploit variation in H-1B-driven STEM employment across U.S. cities and estimate a semi-elasticity of TFP with respect to the STEM employment share of 3.61. Gunadi (2019) transforms this estimate to a log-log elasticity of approximately 0.22 and uses this in conjunction with his own elasticity estimates in a framework similar to ours. Estimating an elasticity of substitution between native and foreign-born STEM workers of approximately 18, he finds that the 2000–2015 foreign STEM labor supply shock increased the average wage of native STEM workers by 4.7 percent, with nearly all of the benefit attributable to productivity spillovers. Similarly, Bongers, Díaz-Roldán, and Torres (2022) use the Peri, Shih, and Sparber (2015) estimate to calibrate productivity spillovers in a two-country DSGE model.

Bound, Khanna, and Morales (2017) provide a second estimate but in a narrower context: they develop a general equilibrium model of the IT labor market calibrated to the H-1B program and estimate a TFP elasticity to computer science employment of approximately 0.23.

The theoretical basis for these effects is the human capital externality literature: Moretti (2004) documents that concentrations of college-educated workers generate spillovers that raise the productivity of all workers, including those without college degrees.

We improve on the approach of Peri, Shih, and Sparber (2015) and provide a new estimate of the elasticity of TFP with respect to STEM employment, building on the shift-share framework common in the immigration literature. We also incorporate insights from the macroeconomic literature on productivity measurement. Following Basu, Fernald, and Kimball (2006) we model variable factor utilization to account for spurious variation in TFP over the business cycle.

Fiscal effects of immigration. The fiscal impact of immigration depends on the age, education, and earnings profile of immigrants over their lifetimes. Summarizing the results of a careful accounting of expected tax payments versus expected government spending costs, Blau and Hunt (2019) report that the long-run fiscal impact is positive at the federal level but negative at the state level due to the costs of educating immigrant children. Clemens (2023) argues that such standard static accounting exercises are incomplete because they miss the role of immigrants in production, which produces tax revenues on capital income as well as the immigrants' own labor. Colas and Sachs (2024) formalize and extend this argument and show that the indirect fiscal benefit of low-skilled immigration through its effect on native wages may outweigh the direct fiscal cost.

We advance this literature by embedding a detailed fiscal infrastructure within a model that also captures the labor market and productivity channels through which STEM immigration affects the economy. This allows us to compute fiscal effects that account for direct taxes and transfers as well as general equilibrium effects on native wages, capital accumulation, and output.

Structural models of immigration. A number of papers use structural general equilibrium models to study the effects of immigration, each emphasizing different channels. Storesletten (2000) develops an OLG model with idiosyncratic earnings risk and a detailed fiscal structure to study whether selective immigration policy can sustain the U.S. fiscal balance as the population ages. He finds that admitting working-age, high-skilled immigrants can substitute for tax increases or benefit cuts, but that the fiscal contribution depends strongly on immigrants' age at arrival and skill level.

Focusing on the innovation channel, Bound, Khanna, and Morales (2017) build a general equilibrium model of the IT labor market in which TFP depends on computer science employment. Calibrated to the H-1B program, they find that expanding skilled immigration raises IT output and TFP but depresses wages of domestic computer scientists, with the net welfare effect depending on the strength of the productivity externality. Bongers, Díaz-Roldán, and Torres (2022) develop a two-country model in which STEM workers generate a TFP externality, finding that attracting STEM workers functions like a positive productivity shock that raises aggregate output and wages.

On the design of immigration policy, Guerreiro, Rebelo, and Teles (2020) characterize the optimal policy when immigrants and natives face the same tax-and-transfer system. They show that the welfare-maximizing policy encourages high-skill immigration and restricts low-skill inflows. Busch et al. (2020) develop an OLG model calibrated to Germany with imperfect native–immigrant substitution to evaluate the 2015–2016 refugee wave. They find that the losses borne by low-skilled natives are small and can be compensated by gains to other groups.

Other work uses dynamic OLG models to study how capital accumulation and skill heterogeneity mediate the effects of immigration over time. Fehr, Jokisch, and Kotlikoff (2004) construct a three-region (U.S., Japan, EU) life-cycle model and find that, although high-skilled immigrants generate larger net fiscal contributions than their low-skilled counterparts, even a substantial expansion of immigration does remarkably little to offset the capital shortages and tax increases projected along the demographic transition—a result that reflects the absence of a productivity channel in their framework. Ben-Gad (2008) incorporates capital–skill complementarity into an OLG model with overlapping dynasties and shows that the immigration surplus accruing to natives depends critically on the skill composition of the inflow and the degree of complementarity between physical capital and skilled labor.s

Our model combines elements from each of these approaches: the STEM–TFP channel of Bound, Khanna, and Morales (2017) and Bongers, Díaz-Roldán, and Torres (2022), the nested CES labor structure of Ottaviano and Peri (2012), and the household heterogeneity and fiscal detail of OLG models in the tradition of Storesletten (2000) and Busch et al. (2020). Unlike most other models in this space, ours traces the full transition dynamics of immigration shocks through all three channels simultaneously—skill-specific labor market complementarity, STEM-driven productivity growth, and a detailed fiscal apparatus that includes Social Security, Medicare, Medicaid, and ACA subsidies. The framework also departs from the balanced-budget rules common in OLG models of immigration by allowing the government to run deficits and accumulate debt, so that the fiscal effects of immigration policy play out through the debt path as well as through contemporaneous taxes and transfers.

3 Model

In this section we provide a general overview of our overlapping generations (OLG) model and a fuller description of key elements, including the labor aggregation structure, wage determination, and endogenous TFP — and the household problem. A complete description of the model underlying OLG framework can be found in the Online Appendix.

3.1 Overview of the Model

Penn Wharton Budget Model's OLG framework is a heterogeneous-agent overlapping generations model that integrates household decisions over consumption, labor supply, and health with detailed fiscal policy dynamics in a general equilibrium framework. It features a large number of households that differ along several dimensions — age, wealth, income, earnings history, health status, and health insurance coverage — alongside a representative firm with constant returns to scale and a government that credibly commits to a fiscal policy, running annual deficits and accumulating debt in accordance with current federal law. Time is discrete, with each model period representing one year.

Households are the core decision-makers. They maximize expected lifetime utility by choosing consumption, labor supply, savings, insurance coverage, and health investment. Utility depends on consumption and leisure, subject to a minimum consumption floor. Working-age households earn wages that vary with age, ability, and health, while retirees receive Social Security (OASI) benefits determined by their earnings history. Health plays a central role in household outcomes and in the path of fiscal policy. Medical expenditure requirements depend on age, health status, and health shocks. Out-of-pocket costs vary by insurance type, with households choosing coverage from among employer-sponsored insurance (ESI), Affordable Care Act (ACA) marketplace plans with income-based subsidies, Medicaid, Medicare, or remaining uninsured.

Production occurs through a representative final goods firm and multiple intermediate sectors. The final goods firm combines intermediate outputs using a constant returns to scale technology. Within each sector, output is produced by two types of businesses — corporate and pass-through — each using capital and labor. Firm decisions determine wages, returns on capital, and the user cost of investment, while equilibrium conditions allocate labor and capital efficiently across sectors. Effective labor input is constructed via a nested constant elasticity of substitution (CES) aggregate that distinguishes workers by education level, STEM occupation, potential experience, and nativity (U.S.-born vs. foreign-born). Total factor productivity has both an exogenous component and an endogenous component that depends on supply of STEM workers, capturing human capital externalities through which STEM workers raise the productivity of all factors.

The government finances public expenditures through tax revenues and debt issuance. It levies progressive individual income taxes, payroll taxes for OASI and Medicare, corporate income taxes, estate taxes, and consumption taxes. Corporate firms face the corporate income tax while pass-through entities are untaxed at the entity level, with net income flowing directly to households to be taxed at the individual level along with other household income . On the spending side, the government provides transfers through Social Security, Medicare, Medicaid, and ACA subsidies, and other transfer programs; purchases output directly; and pays interest on debt. Federal government debt evolves with primary deficits and interest obligations, with securities of varying maturities held by both domestic and foreign investors.

A representative financial intermediary — a mutual fund — aggregates household savings and invests in and government securities. The fund earns a portfolio return composed of dividends from corporate and pass-through capital, interest on government debt, and capital gains. It operates under a zero-profit condition, ensuring that aggregate household returns equal the weighted returns on the assets it holds.

The population evolves over time through age transitions, births, deaths, immigration, and emigration. Baseline paths for future U.S. demographics are simulated using Penn Wharton Budget Model's demographic microsimulation model. Upon death, bequests — net of taxes and transaction costs — are redistributed to survivors based on age and income.

We detail the production structure, the role of STEM workers, and the household's problem below. A full description of the OLG model setup is available in Online Appendix A. Section 4 describes the estimation of key model parameters and projection assumptions.

3.2 Production and Labor

Within each intermediate sector, output is produced using a Cobb–Douglas technology in capital and labor (see Online Appendix A.4 for details). All sectors share the same labor market and face a common TFP externality from STEM employment so labor decisions are effectively characterized by an aggregate production function:

$$Y_t = A_t L_t^{\alpha} K_t^{1-\alpha}, \tag{1}$$

where L_t is the effective labor supply (described in Section 3.2.1), A_t is total factor productivity (described in Section 3.2.2), and K_t is aggregate capital.

3.2.1 Labor Input

Building on the approach of Ottaviano and Peri 2012 and Caiumi and Peri (2024), the labor input L_t is an aggregate of heterogeneous labor types constructed using a sequential four-level CES partitioning. The CES structure is shown in Figure 1.

The top-level nest aggregates college-educated (C) and non-college (NC) effective labor via a

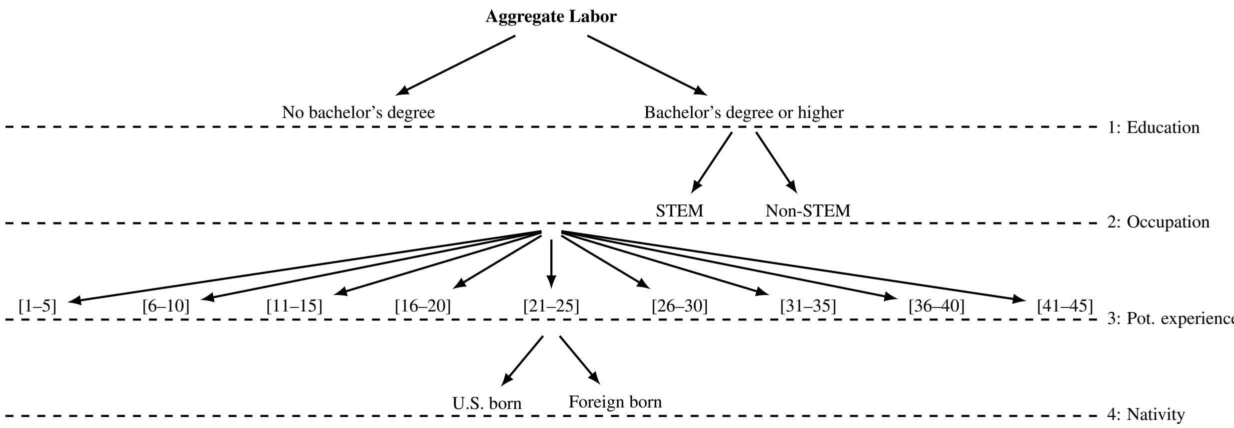


Figure 1: Labor aggregation hierarchy across education, STEM occupation, potential experience, and place of birth.

CES function:[2]

$$L_t = \left[\theta_C \, L_{C,t}^{\frac{\sigma_{COL}-1}{\sigma_{COL}}} + \theta_{NC} \, L_{NC,t}^{\frac{\sigma_{COL}-1}{\sigma_{COL}}}\right]^{\frac{\sigma_{COL}}{\sigma_{COL}-1}} . \tag{2}$$

Here, σ_{COL} denotes the elasticity of substitution between college and non-college workers, while θ_C and θ_{NC} are efficiency parameters capturing the average productivity of each group.

The second nesting level aggregates two types of college-educated workers: those who work in STEM occupations (CS) and those who work in non-STEM occupations (CNS):

$$L_{C,t} = \left[\theta_{CS} \, L_{CS,t}^{\frac{\sigma_{STEM}-1}{\sigma_{STEM}}} + \theta_{CNS} \, L_{CNS,t}^{\frac{\sigma_{STEM}-1}{\sigma_{STEM}}}\right]^{\frac{\sigma_{STEM}}{\sigma_{STEM}-1}} , \tag{3}$$

where σ_{STEM} denotes the elasticity of substitution between STEM and non-STEM college-educated workers and θ_{CS} and θ_{CNS} are relative efficiency parameters.

Together, college degree attainment and STEM occupation define a "skill group" $i \in \{CS, CNS, NC\}$, where CS denotes college-educated STEM workers, CNS college-educated non-STEM workers, and NC non-college workers.[3] Within each skill group, we distinguish nine potential experience groups, indexed by $j \in \{1{-}5,\, 6{-}10,\, \ldots,\, 36{-}40,\, 41{-}45\}$ years of experience. This defines the

2. Throughout this paper, "college-educated" refers to those holding a bachelor's degree or higher.

3. During estimation of the labor parameters, we include a second aggregation level for non-college workers as well to improve the precision of our estimates. However, we adopt a more parsimonious specification in the OLG model by treating non-college labor as a single aggregate. See Appendix B.2 for details on the estimation strategy.

third level of aggregation:

$$L_{i,t} = \left[\sum_{j=1}^{9} \theta_{ij} \, L_{ij,t}^{\frac{\sigma_{EXP_i}-1}{\sigma_{EXP_i}}} \right]^{\frac{\sigma_{EXP_i}}{\sigma_{EXP_i}-1}}, \tag{4}$$

Note that the elasticity of substitution across experience groups σ_{EXP_i} and the efficiency parameters θ_{ij} are specific to skill group i.

The final nest distinguishes workers by nativity, aggregating U.S.-born (US) and foreign-born (F) labor within each skill group i and experience group j:

$$L_{ij,t} = \left[\theta_{ijUS} \, L_{ijUS,t}^{\frac{\sigma_{NAT_{ij}}-1}{\sigma_{NAT_{ij}}}} + \theta_{ijF} \, L_{ijF,t}^{\frac{\sigma_{NAT_{ij}}-1}{\sigma_{NAT_{ij}}}} \right]^{\frac{\sigma_{NAT_{ij}}}{\sigma_{NAT_{ij}}-1}}. \tag{5}$$

Here, $\sigma_{NAT_{ij}}$ denotes the elasticity of substitution between U.S.-born and foreign-born workers within skill group i and experience group j, while θ_{ijk} is the efficiency parameter for nativity $k \in \{US, F\}$.

3.2.2 STEM Workers and TFP Spillovers

Total factor productivity is partly endogenous and depends on the size of the STEM workforce. Similar to Peri, Shih, and Sparber (2015), we specify TFP as a constant-elasticity function of STEM employment:

$$A(N_{CS}, t) = N_{CS}^{\chi} \, \bar{A}(t), \tag{6}$$

where N_{CS} denotes aggregate STEM employment, $\bar{A}(t)$ is an exogenous productivity component, and $\chi \geq 0$ is the elasticity of TFP with respect to STEM employment. Totally differentiating log A with respect to time:

$$\frac{\frac{dA}{dt}}{A} = \frac{\chi}{N_{CS}} \frac{dN_{CS}}{dt} + \frac{1}{\bar{A}} \frac{d\bar{A}}{dt}.$$

This specification implies $\frac{\partial \log A_t}{\partial \log N_{CS}} = \chi$: a one percent increase in STEM employment raises TFP by χ percent, holding all else equal.

Equation (6) captures a *level* effect of STEM workers on productivity, rather than a permanent effect on the growth rate of TFP. The economic mechanism is a human capital externality in the spirit of Lucas (1988): concentrations of STEM workers raise the productivity of all factors through knowledge spillovers—transmitting technical knowledge to non-STEM workers, facilitating the adoption of new technologies, and resolving production frictions that require specialized analytical skills. This formulation has been applied by Peri, Shih, and Sparber (2015) to study the impact of foreign-born STEM workers on productivity in U.S. cities, by Gunadi (2019) to estimate the national wage effects of STEM immigration, and by Bound, Khanna, and Morales (2017) to quantify the contribution of computer scientists to IT-sector TFP.

Because individual firms are atomistically small, each takes aggregate STEM employment N_{CS} as given when choosing its own labor inputs. The TFP externality is therefore not internalized: STEM workers are paid their private marginal product, which is less than their social marginal product by $\chi Y/N_{CS}$. This wedge is particularly important for analyzing immigration, as foreign-born STEM workers may generate productivity gains that accrue broadly to the economy but are not reflected in the immigrants' own wages.

3.2.3 Wage Determination

Given imperfect substitution across all labor groups, profit-maximizing firms operating under perfect competition pay each labor type its marginal product. For each skill group i, experience level j, and nativity k the market wage is given by

$$w_{ijk} = MPL_{ijk} \tag{7}$$

College-educated workers. College-educated workers vary by skill group $i \in \{CS, CNS\}$ (STEM vs. non-STEM), potential experience level $j \in \{1-5,\, 6-10,\, \ldots,\, 36-40,\, 41-45\}$, and nativity $k \in \{US, F\}$. Their marginal product is given by,

$$MPL_{ijk} = \alpha\, \theta_C \theta_i \theta_{ij} \theta_{ijk} \frac{Y_t}{L_t} \times \underbrace{\left(\frac{L_t}{L_{C,t}}\right)^{\frac{1}{\sigma_{COL}}}}_{\text{Substitution across education levels}} \underbrace{\left(\frac{L_{C,t}}{L_{i,t}}\right)^{\frac{1}{\sigma_{STEM}}}}_{\text{Substitution across STEM occupation}} \times \underbrace{\left(\frac{L_{i,t}}{L_{ij,t}}\right)^{\frac{1}{\sigma_{EXP_i}}}}_{\text{Substitution across experience levels}} \underbrace{\left(\frac{L_{ij,t}}{L_{ijk,t}}\right)^{\frac{1}{\sigma_{NAT_{ij}}}}}_{\text{Substitution across nativity}} \tag{8}$$

Non-college workers. Workers with no college degree ($i = NC$) vary only by potential experience level $j \in \{1-5,\ 6-10,\ \ldots,\ 36-40,\ 41-45\}$ and nativity $k \in \{US, F\}$. Their marginal product is given by,

$$MPL_{ijk} = \alpha\, \theta_{NC} \theta_{ij} \theta_{ijk} \frac{Y_t}{L_t} \times \underbrace{\left(\frac{L_t}{L_{i,t}}\right)^{\frac{1}{\sigma_{COL}}}}_{\text{Substitution across education levels}} \times \underbrace{\left(\frac{L_{i,t}}{L_{ij,t}}\right)^{\frac{1}{\sigma_{EXP_i}}}}_{\text{Substitution across experience levels}} \times \underbrace{\left(\frac{L_{ij,t}}{L_{ijk,t}}\right)^{\frac{1}{\sigma_{NAT_{ij}}}}}_{\text{Substitution across nativity}} \tag{9}$$

Two channels of wage spillovers from STEM employment. The expressions above show that an increase in STEM employment N_{CS} affects the wages of every worker type through two distinct channels:

1. *TFP channel.* From (6), an increase in STEM employment raises TFP, which enters the common factor Y_t/L_t in every worker's marginal product. This is an economy-wide productivity gain that benefits all worker types proportionally.[4]

2. *Labor substitution channel.* An increase in N_{CS} raises effective STEM labor L_{CS}, which propagates upward through the CES hierarchy, increasing $L_{C,t}$ and L_t. The resulting shifts

4. Since $Y_t/L_t = A_t(K_t/L_t)^{1-\alpha}$, this term also depends on the capital-labor ratio. In the short run, the increase in L_t from additional STEM workers reduces K_t/L_t, partially offsetting the TFP gain. In general equilibrium, capital adjusts over time to accommodate the larger labor force, so the long-run effect on Y_t/L_t is driven primarily by the TFP gain. See Appendix A.2.2 for a formal treatment of capital adjustment dynamics.

in the ratios of the marginal product expressions affect wages, with the outcome depending on the worker's position in the nesting hierarchy.

For workers without a college degree ($i = NC$), the labor substitution channel operates only through the education-level ratio $(L_t/L_{NC,t})^{1/\sigma_{COL}}$. An increase in college-educated STEM labor raises aggregate labor L_t relative to the non-college component $L_{NC,t}$, so the substitution channel is positive, reflecting complementarity across education levels. Both channels reinforce each other: non-college wages unambiguously rise.

For *college-educated non-STEM workers*, the substitution channel operates through two opposing ratios. The education-level ratio $(L_t/L_{C,t})^{1/\sigma_{COL}}$ falls, because the growth in L_{CS} raises the college aggregate more than it raises total labor supply — college workers become relatively more abundant. At the same time, the higher STEM share within college-educated workers raises the STEM occupation ratio $(L_{C,t}/L_{CNS,t})^{1/\sigma_{STEM}}$, reflecting complementarity between STEM and non-STEM college workers. While the TFP channel remains unambiguously positive, the net substitution effect depends on the relative impact of these two ratios.

For *college-educated STEM workers*, both substitution ratios move against them. The education-level ratio $(L_t/L_{C,t})^{1/\sigma_{COL}}$ falls, as for non-STEM college workers, and the STEM occupation ratio $(L_{C,t}/L_{CS,t})^{1/\sigma_{STEM}}$ also falls as L_{CS} grows. Here the direction of the substitution is unambiguous, as both CES ratios exert downward pressure their wage. The overall net wage effect depends on whether the TFP gain outweighs this combined own-supply effect. When σ_{STEM} is low — meaning STEM and non-STEM college-educated workers are more complementary — or when χ is small, the labor substitution effect may dominate and STEM workers' wages can fall despite productivity gains.

Appendix A.2.1 formalizes these dynamics for a simplified two-type CES labor aggregate and shows directly how wage changes depend on workforce composition, substitution elasticities, and the strength of TFP spillovers.

STEM immigration versus STEM employment. Thus far we have considered the effects of STEM employment generally. We now discuss the effects of STEM *immigration* — increasing the number of foreign-born, college-educated workers in STEM occupations — which is the focus of our analysis.

The lowest level of the CES hierarchy distinguishes U.S.-born and foreign-born workers within each skill-experience cell, with substitution elasticity $\sigma_{NAT_{ij}}$. In response to increased immigration of foreign-born STEM workers, the nativity ratio $(L_{ij,t}/L_{ijk,t})^{1/\sigma_{NAT_{ij}}}$ falls for foreign-born STEM workers, reflecting own-supply pressure, and rises for U.S.-born STEM workers, reflecting complementarity with the additional foreign-born workers.

U.S.-born STEM workers therefore face offsetting substitution pressures in response to STEM immigration. The education-level and STEM occupation ratios move against them as described above, while the nativity ratio moves in their favor. The net effect of labor substitution on U.S.-born STEM workers is therefore less negative than for the STEM group as a whole. By contrast, foreign-born STEM workers face downward wage pressure through all three substitution channels — education, occupation, and nativity.

Unlike the labor substitution channel, the relationship between the STEM workforce, TFP, and wages is the same for both U.S.-born and foreign-born workers. The specification in (6) counts all STEM workers symmetrically, so an immigration-driven increase in N_{CS} generates the same productivity externality as one driven by U.S.-born workers.

3.3 The Household Problem

A household's state at any point in time is summarized by the vector $s \in S$, where $S = J \times A \times B \times Z \times H \times D \times I$ is the individual state space. Each component represents a dimension of heterogeneity: age (J), assets (A), average lifetime earnings (B), labor ability (Z), health (H), medical needs (D), and insurance status (I). Let $\Psi_t^T = \{p_i, \psi_i, X_i\}_{i=t}^T$ denote the sequence of prices, government policies, and household distributions from year t to T, which households and firms take as given.

Let $V(s, \Psi_t^{t+J_D-j})$ be the value function of a household of age j at the beginning of year t. The household solves

$$V(s, \Psi_t^{t+J_D-j}) = \max_{c,l,a',i',\iota} \Big\{ u(c,l) + \beta \mathbb{E}\Big[\Big(\delta_t(j,h) + (1-\delta_t(j,h))\zeta \Big) V(s', \Psi_{t+1}^{t+J_D-j-1}) \Big| s, \Gamma_h \Big] \Big\} \tag{10}$$

subject to constraints on the decision variables,

$$c \geq \underline{c}, \qquad 0 \leq l \leq \Omega_{a,t}(j)\, l_{\max}, \qquad a' \geq a'_{\min}(s), \qquad i' \in I, \qquad \iota \in [0,1],$$

and the laws of motion for the individual state,

$$\begin{aligned}
s' &= (j+1, a', b', z', h', d', i'), \\
a' &= y_t^P + (1+r_t)a + tr_{OASI,t}(j, b, p_{t-j}^t) - \tau_{I,t}(y_t^T, \tilde{r}_t a, tr_{OASI,t}(j, b, p_{t-j}^t)) \\
&\quad - \tau_{P,t}(y_t^T) - \tau_{state}(y_t^T) - \tau_{C,t}(\psi_t^C c) - (1 - \mathbf{1}_{[(i,y,j)=ESI]}\eta^{ESI})\, \Omega_{a,t}(j) p_{i,t} \\
&\quad + \mathbf{1}_{[(i,y,j,a)=ACA]}\pi_{ACA,t}^{PremSub}(y_t^T) - \iota \cdot oop_t(m(j,h,d,i,X), i) + SN_t(s,l) + SSI_t(s,l), \qquad (11) \\
b' &= \mathbf{1}_{\{j<J_R\}} \frac{1}{j-20} \Big[(j-21)\, b\, \frac{w_t}{w_{t-1}} + \min\Big(\frac{y_t^T}{\Omega_{a,t}(j)}, \vartheta_{\max} \Big) \Big] + \mathbf{1}_{\{j \geq J_R\}} b, \qquad (12)
\end{aligned}$$

where $\tilde{r}_t = (r_t^{corp}, r_t^{pass}, r_t^G, r_t^C)$ is the vector of asset returns; $\mathbf{1}_{\{\cdot\}}$ is an indicator function equal to one when its condition holds and zero otherwise; SN_t and SSI_t denote SNAP and SSI transfers, respectively; and $\vartheta_{\max}$ is the OASI maximum taxable earnings cap.

Average historical earnings b' evolve recursively as a weighted average of current (beginning-of-year) wage-indexed earnings $b \cdot w_t / w_{t-1}$ and current OASDI taxable earnings $\min\{y_t^T / \Omega_{a,t}(j),\ \vartheta_{\max}\}$, and remain constant after retirement.[5] We deflate earnings by $\Omega_{a,t}(j)$ so that the earnings history reflects individual rather than household earnings. In Online Appendix Section B.9.1, we describe

5. In the actual AIME calculation, a worker's annual taxable earnings are indexed to reflect the general earnings level in the indexing year (age 60 for most workers); earnings in years after the indexing year are not indexed. See Social Security Administration (2018) for details. We do not model the change in indexing after age 60 in the dynamic OLG model.

how benefits are computed from individual earnings and then rescaled by the average number of adults to approximate total expected household benefits.

y^P denotes paid labor income, computed as gross labor income net of the fully tax-deductible employer-sponsored insurance subsidy for workers enrolled in private health insurance. Gross labor income depends on workers' wages, which, as described in Section 3.2.3, vary by skill group i, potential experience j, and nativity.

We collect the household's decision rules and the law of motion for average historical earnings as $\Gamma_h(s, \Psi_t^{t+J_D-j}) \equiv \{c, l, a', i', \iota, b'\}(s, \Psi_t^{t+J_D-j})$.

4 Estimation

This section describes the estimation of the key model parameters. These include the labor aggregator parameters — elasticities of substitution and efficiency weights — and the elasticity of TFP with respect to STEM employment. We also describe the demographic projections that underlie our policy simulations.

4.1 Data

Our primary data source is the American Community Survey (ACS). We use ACS microdata for 2000 through 2023, excluding 2020 from estimates due to pandemic-related data irregularities. We supplement the ACS with the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) for survey years 2001–2024. All computations use survey person weights to ensure population representativeness. Appendix B provides full details on the construction of both samples.

We restrict our sample to individuals aged 18 and older who are not residing in group quarters and who are not currently enrolled in school. For employment analyses, we include all individuals who reported working at least one week during the reference year. For wage analyses, we further restrict the sample to paid employees with positive earnings, excluding unpaid family workers and the self-employed. Nominal wages are converted to 2024 dollars using the CPI for All Urban Consumers (CPI-U).

Workers are classified into skill groups based on educational attainment and whether they work in a STEM occupation. The three groups used throughout the model are: college-educated STEM (CS), college-educated non-STEM (CNS), and non-college (NC).[6] We classify occupations as STEM using the 2010 Standard Occupational Classification (SOC) Policy Committee definition, which encompasses life and physical sciences, engineering, mathematics, information technology, social sciences, and health occupations, along with related managerial and post-secondary teaching positions.[7]

Potential experience is computed as age less an assumed age of labor market entry that varies by education level: age 17 for workers with no high school degree, 19 for high school graduates, 21 for workers with an associate's degree or some college, 23 for college graduates, and 24 for those with graduate degrees. Workers are grouped into nine five-year potential experience intervals (1–5, 6–10, ..., 41–45). Labor supply is measured in full-time-equivalent (FTE) units, computed as annual hours worked—weekly hours multiplied by weeks worked—divided by 2,000. For years 2008–2018, when the ACS reported weeks worked only in intervals, we impute weeks worked using interval midpoints.

4.2 The Labor Aggregator

The nested CES labor aggregator described in Section 3.2.1 contains two sets of parameters: the elasticities of substitution (σ), which govern the degree of substitutability between worker types at each level of the hierarchy, and the efficiency parameters (θ), which capture the relative productivity of each group. We estimate these using different but complementary methods, designed to fully integrate the empirical dynamics into the OLG model.

6. During estimation of the labor aggregator parameters, we employ a finer partition of non-college workers—less than high school, high school, and some college—to improve the precision of our estimates. The OLG model aggregates these into a single non-college group. See Appendix B.2 for details.

7. Specifically, we follow the crosswalk provided by the 2010 U.S. Standard Occupational Classification (SOC) Policy Committee.

4.2.1 Elasticities of Substitution

We estimate the elasticities of substitution following the sequential bottom-up approach of Ottaviano and Peri (2012) and Caiumi and Peri (2024). Starting at the lowest level of the CES hierarchy—the nativity margin—we estimate the substitution elasticity and then use that estimate to construct CES-weighted labor aggregates for the next level up. This process repeats through the experience, STEM/non-STEM, and college/non-college nests.

At the U.S.-versus-foreign-born level, elasticities are estimated by OLS on relative log wages within education–experience–year cells, with cell-level fixed effects absorbing permanent productivity differences. At higher levels of the hierarchy—where relative effective labor supply is endogenous to wages—we estimate the elasticities by two-stage least squares (2SLS), using log foreign-born hours as instruments for total cell labor supply. The exclusion restriction is that immigration-driven variation in foreign-born labor supply is uncorrelated with domestic demand shocks within skill-experience cells after conditioning on fixed effects.

Appendix B provides the full regression specifications and standard error computations. Our key extension relative to Caiumi and Peri (2024) is the addition of a STEM/non-STEM nest within college-educated workers. We also estimate separate nativity elasticities by skill group, finding substantial heterogeneity in the degree of native–immigrant complementarity across the skill distribution. In principle, these elasticities could also vary by experience group as in (5). However, we find that the elasticity is generally stable across experience groups, so for parsimony we impose a single nativity elasticity for all experience groups within each skill group.

Table 1 reports our elasticity estimates. The results indicate limited substitutability between broad education groups and progressively higher substitutability among workers who share more observable characteristics. The elasticity of substitution between college-educated and non-college workers ($\sigma_{COL} = 1.81$, $SE = 0.57$) is consistent with the benchmark range of 1.5–2.5 estimated by Katz and Murphy (1992) and Ottaviano and Peri (2012), and adopted by Caiumi and Peri (2024) in their simulations. Within college-educated workers, the STEM–non-STEM elasticity ($\sigma_{STEM} = 4.90$, $SE = 3.28$) is roughly three times as large, indicating that while STEM and non-

Table 1: Estimated Elasticity Parameters in the Nested CES Structure

Parameter	Description	Value
Elasticities		
σ_{COL}	College vs. non-college	1.81 (0.57)
σ_{STEM}	College non-STEM vs. college STEM	4.90 (3.28)
$\sigma_{EXP_{NC}}$	Experience groups: Non-college	6.94 (2.42)
$\sigma_{EXP_{CNS}}$	Experience groups: College non-STEM	3.76 (2.12)
$\sigma_{EXP_{CS}}$	Experience groups: College STEM	4.57 (4.19)
$\sigma_{NAT_{NC}}$	U.S.-born vs. foreign-born: Non-college	30.86 (4.58)
$\sigma_{NAT_{CNS}}$	U.S.-born vs. foreign-born: College non-STEM	21.92 (4.50)
$\sigma_{NAT_{CS}}$	U.S.-born vs. foreign-born: College STEM	5.33 (0.21)
Memo: Pooled estimates for by-skill-group elasticities		
σ_{EXP}	Experience groups: All skill groups	6.16 (1.79)
σ_{NAT}	U.S.-born vs. foreign-born: All skill groups	14.61 (2.12)

Notes: Standard errors in parentheses, computed via the delta method. Elasticities estimated from ACS 2000–2023 (excluding 2020). See Appendix B for details.

STEM workers are more interchangeable than workers across education levels, they are far from perfect substitutes. Experience groups are moderately substitutable ($\sigma_{EXP} = 6.16$, $SE = 1.79$), consistent with the range reported in the existing literature.

Along the nativity margin, our pooled estimate of $\sigma_{NAT} = 14.61$ ($SE = 2.12$) is close to the value of approximately 15 reported by Caiumi and Peri (2024) for 2000–2023 and somewhat below the estimate of roughly 20 in Ottaviano and Peri (2012), pointing to somewhat stronger native–immigrant complementarity in more recent data. The nativity elasticity varies sharply by skill group: workers without a college degree are near-perfect substitutes ($\sigma_{NAT_{NC}} = 30.86$, $SE = 4.58$), college-educated non-STEM workers display high but finite substitutability ($\sigma_{NAT_{CNS}} = 21.92$, $SE = 4.50$), and college-educated STEM workers exhibit substantially lower substitutability ($\sigma_{NAT_{CS}} = 5.33$, $SE = 0.21$). This high degree of complementarity between college-educated U.S.-born and foreign-born STEM workers is a central finding of our analysis. Notably, Gunadi (2019) pools STEM workers at all education levels into a single STEM worker type and estimates a substitution elasticity in the range of 13 to 18 between U.S.-born and foreign-born STEM workers. Our estimate of 5.33 treats college-educated STEM workers as a distinct group and instead

Table 2: Summary of efficiency parameters in the nested CES labor aggregation.

Parameter	Description	Normalization	Figure 1 Reference
θ_C	College education	$\theta_C + \theta_{NC} = 1$	Level 1
θ_{NC}	No college		
θ_{CS}	STEM workers (within C)	$\theta_{CS} + \theta_{CNS} = 1$	Level 2
θ_{CNS}	Non-STEM workers (within C)		
For each skill group i:			
θ_{ij}	Workers within experience group j	$\sum_j \theta_{ij} = 1$	Level 3
For each skill group i:			
θ_{iUS}	U.S.-born workers	$\theta_{iUS} + \theta_{iF} = 1$	Level 4
θ_{iF}	Foreign-born workers		

pools all workers with no college degree, regardless of whether they work in a STEM occupation. Our finding of much stronger complementarity suggests that the combination of higher education and STEM work defines a distinct worker type which behaves differently from college-educated workers generally or STEM workers generally.

The low elasticity of substitution between U.S.- and foreign-born STEM workers has direct policy implications: admitting an additional foreign-born STEM worker generates larger complementarity gains and smaller displacement effects for U.S.-born workers than admitting workers in other skill groups.

4.2.2 Efficiency Parameters

Given the estimated elasticities, we pin down the efficiency parameters θ by matching observed average labor earnings across demographic groups, experience levels, and nativity within the full model. For each nest in the CES hierarchy, the θ parameters are chosen to satisfy two objectives:

1. *Targeting.* The model reproduces the corresponding empirical earnings ratios, such as the ratio of non-college to college earnings or of U.S.-born to foreign-born earnings within each skill group.

2. *Normalization.* The efficiency parameters within a given nesting level all sum to one.

Table 2 summarizes these parameters along with their normalizations, consistent with the aggregation structure in Section 3.[8] For reference, the table also reports the corresponding aggregation level as depicted in Figure 1.

Table 3 reports the estimated efficiency parameters, along with the empirical moments targeted in estimation. The targeted moments correspond to average labor earnings of working-age households across different demographic groups and experience levels.[9]

8. In a version of the model with perfect substitution across labor types, the economy-wide average wage per unit of effective labor is normalized to one by calibrating a scalar adjustment to TFP (A in the production function (1)). Introducing imperfect substitutability with normalized efficiency parameters changes the level of the economy-wide average wage per unit of effective labor, and therefore requires recalibrating the scalar attached to A. We jointly calibrate the A adjustment alongside the efficiency parameters, producing a value of 11.85, which ensures that the economy-wide average wage normalization holds as well.

9. The model is expressed in normalized units rather than dollars, so we introduce a conversion factor to map model units into dollars in order to compare average labor earnings at the aggregate level. In the data, average labor earnings of working-age households was \$120,238 in 2024. In the model, the average labor income of all working-age households is 0.9069 in model units. We therefore set the dollar-per-model-unit conversion parameter to

$$\frac{\$120{,}238}{0.9069} = \$130{,}974,$$

expressed in 2024 dollars per model unit. This conversion factor is used throughout the model to translate variables from model units into dollar units.

Table 3: Efficiency parameters and associated empirical targets by demographic group and age.

Parameter	Value	Target moment
Panel A: Aggregate efficiency parameters		
θ_{NC}	0.299	Ratio of non-college to college earnings ($77,549 / $176,402 = 0.440)
θ_{CNS}	0.489	Ratio of non-STEM to STEM earnings ($162,063 / $207,239 = 0.782)
$\theta_{NC_{US}}$	0.568	Ratio of U.S.-born to foreign-born earnings (non-college) ($77,550 / $78,050 = 0.994)
$\theta_{CNS_{US}}$	0.519	Ratio of U.S.-born to foreign-born earnings (college non-STEM) ($159,515 / $177,000 = 0.901)
$\theta_{CS_{US}}$	0.493	Ratio of U.S.-born to foreign-born earnings (college STEM) ($193,929 / $249,129 = 0.778)
Panel B: Non-college by age group		
$\theta_{NC[21\text{–}25]}$	0.109	Earnings in age group 21–25: $41,255
$\theta_{NC[26\text{–}30]}$	0.106	Earnings in age group 26–30: $57,556
$\theta_{NC[31\text{–}35]}$	0.101	Earnings in age group 31–35: $69,345
$\theta_{NC[36\text{–}40]}$	0.104	Earnings in age group 36–40: $81,404
$\theta_{NC[41\text{–}45]}$	0.106	Earnings in age group 41–45: $86,462
$\theta_{NC[46\text{–}50]}$	0.111	Earnings in age group 46–50: $90,266
$\theta_{NC[51\text{–}55]}$	0.117	Earnings in age group 51–55: $92,891
$\theta_{NC[56\text{–}60]}$	0.119	Earnings in age group 56–60: $93,139
$\theta_{NC[61\text{–}65]}$	0.127	Earnings in age group 61–65: $90,621
Panel C: College non-STEM by age group		
$\theta_{CNS[21\text{–}25]}$	0.067	Earnings in age group 21–25: $64,828
$\theta_{CNS[26\text{–}30]}$	0.091	Earnings in age group 26–30: $97,637
$\theta_{CNS[31\text{–}35]}$	0.103	Earnings in age group 31–35: $137,352
$\theta_{CNS[36\text{–}40]}$	0.112	Earnings in age group 36–40: $171,776
$\theta_{CNS[41\text{–}45]}$	0.117	Earnings in age group 41–45: $186,716
$\theta_{CNS[46\text{–}50]}$	0.121	Earnings in age group 46–50: $197,611
$\theta_{CNS[51\text{–}55]}$	0.128	Earnings in age group 51–55: $201,323
$\theta_{CNS[56\text{–}60]}$	0.127	Earnings in age group 56–60: $190,030
$\theta_{CNS[61\text{–}65]}$	0.135	Earnings in age group 61–65: $192,397

Continued on next page

Parameter	Value	Target moment
Panel D: College STEM by age group		
$\theta_{CS[21–25]}$	0.069	Earnings in age group 21–25: \$93,694
$\theta_{CS[26–30]}$	0.094	Earnings in age group 26–30: \$130,452
$\theta_{CS[31–35]}$	0.108	Earnings in age group 31–35: \$183,827
$\theta_{CS[36–40]}$	0.117	Earnings in age group 36–40: \$228,260
$\theta_{CS[41–45]}$	0.119	Earnings in age group 41–45: \$244,085
$\theta_{CS[46–50]}$	0.120	Earnings in age group 46–50: \$252,870
$\theta_{CS[51–55]}$	0.122	Earnings in age group 51–55: \$251,829
$\theta_{CS[56–60]}$	0.122	Earnings in age group 56–60: \$249,445
$\theta_{CS[61–65]}$	0.129	Earnings in age group 61–65: \$252,162

Source: Penn Wharton Budget Model.

4.3 STEM Workers and Total Factor Productivity

We now turn to estimating χ, the elasticity of TFP with respect to STEM employment in the specification $A(N_{CS}, t) = N_{CS}^{\chi}\,\bar{A}(t)$ introduced in Section 3.2.2. This parameter governs the strength of the human capital externality through which STEM workers raise the productivity of all factors. Since TFP is not directly observed, the estimation proceeds in two stages: first, we extract a composition-adjusted TFP series from wage data; second, we estimate its relationship to STEM employment via instrumental variables.

4.3.1 Recovering TFP from Wages

Under the nested CES structure and competitive labor markets, the log marginal product of each worker type is the sum of CES composition terms (which depend on relative labor supplies and the estimated elasticities) plus a common component reflecting TFP and the capital-labor ratio. In principle, the wage residual after removing the composition terms identifies TFP. In practice, three sources of confounding variation must be addressed.

First, *composition effects* arise because changes in the relative supply of different worker types shift relative wages even without any change in technology. In an economy with a growing STEM workforce, a naive average of cell-level wage residuals confounds technology with shifts in labor force composition. The direction and magnitude of the bias depend on the elasticities of substitution and the labor force composition, and vary across worker types. For example, non-STEM workers experience upward-biased residuals (reflecting complementarity with STEM workers), while STEM workers experience downward-biased residuals (reflecting own-supply pressure). Appendix A.2.1 provides a formal derivation of the composition bias in a simplified two-type CES economy.

Second, *sluggish capital adjustment* introduces bias when capital does not adjust instantaneously to changes in TFP or labor supply. The standard approach in the immigration literature—following Ottaviano and Peri (2012) and Caiumi and Peri (2024)—assumes that the economy is on its balanced growth path, so that the capital-labor ratio can be "solved out" of the production function. Over our estimation period (2000–2019), this assumption is unlikely to hold at annual frequency: if capital adjusts slowly to productivity improvements, observed wage growth will understate TFP growth.

Third, *cyclical variation in factor utilization* induces spurious movements in measured TFP over the business cycle. Following Basu, Fernald, and Kimball (2006), we model variable utilization of both labor (effort) and capital (workweek), proxying for unobserved utilization using hours per worker. Since hours per worker are themselves correlated with technology shocks, we instrument for utilization using external shocks—monetary policy, oil prices, and financial shocks from Comin et al. (2025).

We embed all three corrections in a bivariate state-space model in which the state vector consists of TFP growth and capital-labor ratio growth, with the state transition equation allowing for sluggish capital adjustment while imposing balanced-growth restrictions in the long run. The measurement equation treats cell-level wage residuals as noisy, repeated measurements of the same underlying TFP signal, with a covariance structure that reflects the CES nesting hierarchy. This approach pools information across all skill-experience-nativity cells while absorbing composition-driven heterogeneity without requiring explicit cell-by-cell corrections for relative supply shifts. The system parameters and the filtered TFP series are recovered by maximum likelihood estimation via the Kalman filter. Appendix A provides the full specification and estimation details.

The resulting utilization-adjusted TFP estimate grows at an average annual rate of 0.66 percent from 2001 to 2019. As a benchmark, we compare our estimate with the utilization-adjusted TFP series derived by Fernald (2025) from entirely different data sources and – except for the utilization adjustment step – using different methods. His series grows at a very similar average annual rate of 0.73 percent over the same period.

4.3.2 Estimating the STEM–TFP Elasticity

With the composition-adjusted TFP series in hand, we estimate χ by two-stage least squares, regressing TFP growth on growth in STEM employment while instrumenting for the latter. An instrument is needed because of the possibility of reverse causality—high-productivity economies attract STEM workers, particularly through immigration channels that favor skilled workers—and omitted variable bias. In particular, the expansion of the U.S. technology sector during our sample period may drive both STEM employment growth and productivity growth through past innovations, so that a naive regression would overstate the causal effect of STEM workers on TFP.

Inspired by Caiumi and Peri (2024), we construct separate instruments for growth in U.S.-born and foreign-born STEM employment, which we then sum to form the aggregate instrument for STEM employment. The key requirement is that the instruments isolate variation in STEM labor supply that is uncorrelated with contemporaneous productivity shocks.

U.S.-born instrument. We exploit the demographic structure of the labor force to construct a cohort-based instrument. The instrument imputes U.S.-born STEM employment growth over five-year windows by "aging" the existing experience distribution: the employment of the j-th five-year experience group in year $t + 5$ is set equal to the employment of experience group $j - 1$ in year t. To predict the inflow at the bottom of the experience distribution, we fix the STEM share among workers with the corresponding ages at its value in year t and apply it to the total number of workers in that age range in year $t + 5$. The imputed growth is then

$$\Delta_{S,US}(t_k, t_k + 5) = s_{US,\text{exp}_0}\, \Delta_{\text{exp}_0,US}(t_k, t_k + 5) - S_{US,\text{exp}_n}(t_k),$$

where s_{US,exp_0} is the STEM share of the labor force in the age range corresponding to the lowest experience group, $\Delta_{\text{exp}_0,US}(t_k, t_k + 5)$ is the aggregate change in labor force employment in that age range, and $S_{US,\text{exp}_n}(t_k)$ is STEM employment in the highest experience group (who exit the labor force). Variation in this instrument thus arises from cohort sizes and lagged STEM shares, not from contemporaneous productivity conditions.

Foreign-born instrument. For foreign-born STEM workers, we construct a shift-share instrument that isolates variation driven by origin-country population flows. We fix the shares of the population employed in STEM occupations by country and region of origin at their 1990 values—covering the top five sending countries (Mexico, Cuba, China, the Philippines, and Korea) and seven broad regions—and grow each origin group by its subsequent population flows. The imputed growth is

$$\Delta_{S,F}(t_k, t_{k+1}) = \sum_{o} \sum_{\ell \leq k} s_{o,1990} \, \Delta_{o,t_\ell},$$

where $s_{o,1990}$ is the initial STEM employment share of population from origin o and Δ_{o,t_ℓ} is the change in population from origin o between periods t_ℓ and $t_{\ell+1}$. Following Caiumi and Peri (2024), negative flows are set to zero. The identifying variation arises from the changing patterns of origin-country migration flows over time, which are plausibly exogenous to U.S. productivity conditions. Because immigration patterns may be influenced by further lags in growth, the shares are fixed at 1990—well before our estimation period—to ensure the instrument is uncorrelated with contemporaneous shocks.

The total instrument is the sum of the U.S.-born and foreign-born components. Both are computed over five-year reference periods (2000–2005, 2005–2010, 2010–2015, 2015–2019) and linearly interpolated between reference dates to complete the annual series. Appendix A.3 provides additional implementation details.

Table 4: Estimated Elasticity of TFP to STEM Employment

Elasticity (χ)	0.188
Standard Error	0.059
t-statistic	3.167
p-value	0.005

4.3.3 Results

Table 4 reports our estimation results. We find an elasticity of TFP with respect to STEM employment of 0.19, comparable to the estimates of 0.22 to 0.23 obtained by Gunadi (2019) and Bound, Khanna, and Morales (2017).

4.4 Demographic Projections

The population paths that enter the OLG model are generated by Penn Wharton Budget Model's demographic and economic microsimulation (*demsim*) model, a discrete-time microsimulation that projects the U.S. population forward at the individual level. Each simulated person is characterized by age, sex, race, educational attainment, country of birth, immigration status (lawful permanent resident, temporary work visa, unauthorized, etc.), STEM occupation, and other demographic and economic attributes. The model advances the population year by year through births, deaths, immigration, and emigration, updating individual characteristics at each step. The resulting demographic projections are broadly consistent with those produced by the Congressional Budget Office and the Social Security Administration. We provide a brief overview of key components of the model below; a full description will be provided in Penn Wharton Budget Model (Forthcoming).

Initial population. The initial population is constructed by sampling from ACS microdata. Households are sampled with probability proportional to their survey weights, preserving the joint distribution of individual and household characteristics, including family structure such as spousal and parent-child links. Survey weights are then recalibrated so that the resulting synthetic population assigns equal weight to all persons (i.e. a weight of one) while matching control population totals on a set of demographic dimensions (generally, the same dimensions used as controls for ACS survey weights).

Mortality. Mortality rates are estimated from CDC National Center for Health Statistics (NCHS) mortality data. The estimated rates vary by age, sex, race, educational attainment, and marital status. Hazard rates are converted to annual survival probabilities, which govern the stochastic removal of individuals from the population at the end of each simulation year. Survival is further conditioned on health status: the age-by-year survival rates estimated from NCHS data are combined with health-state-specific mortality differentials estimated from the Medical Expenditure Panel Survey (MEPS), as described in Online Appendix A.2.2. Mortality rates are projected forward based on maximum autocorrelation factors (MAF) applied to a tensor decomposition of mortality rates by year, age, and demographic group. This approach is essentially a multi-factor, multi-dimensional generalization of the Lee and Carter (1992) method that selects for the most persistent, predictable signals in the historical data.

Fertility. Fertility rates are estimated from CDC NCHS natality data. Each year, women aged 15 to 49 are assigned a probability of giving birth that depends on their age, race, educational attainment, marital status, and number of prior children. Births arrive stochastically based on these rates. Newborns inherit demographic characteristics from their parents, and the model updates parent-child links and household composition to reflect the new persons. Fertility rates are projected forward based on the same method as mortality rates, applying MAF to a tensor decomposition of fertility rates by year, age, and demographic group.

Immigration. New immigrants entering the simulation each year are sampled from a pool of recently-arrived foreign-born individuals observed in the ACS. The pool is drawn from actual survey respondents who immigrated within a short window of the date they were surveyed, so the sampled immigrants preserve the joint distribution of demographic characteristics observed among recent arrivals, including age at immigration, educational attainment, country of origin, STEM occupation, and family structure. This approach ensures that the composition of simulated new immigrants reflects the observable heterogeneity in the actual new arrival population without parametric assumptions on the joint distribution of characteristics.

Immigrants transition across legal status categories over time and may emigrate. For example, a lawful permanent resident may naturalize and become a citizen; a refugee arrival may adjust status to lawful permanent residence; a temporary work visa holder may depart when their visa reaches its maximum duration, or overstay and become part of the unauthorized immigrant population; and so on. These transitions are partly stochastic and partly governed by statutory and administrative constraints.

The flow of future new immigrant arrivals is projected based on the provisions of current law, global population projections, and historical patterns.

STEM occupation. Working-age individuals in the simulation are assigned to either a STEM or non-STEM occupation. The assignment is based on a logit model estimated on longitudinally linked CPS ASEC data. The model conditions on age, sex, race, educational attainment (including degree type), immigrant status, and the individual's occupation in the prior year, allowing for persistence in occupational choice. Immigrants entering the simulation inherit their STEM occupation status from the donor observation in the ACS pool from which they were sampled.

5 Model performance

This section evaluates how well the model's steady state replicates key moments in the data and examines additional equilibrium outcomes of interest. The steady state corresponds to the static equilibrium of the economy, abstracting from transitional dynamics. All results are calibrated to match the economic environment of 2024.

We proceed in three steps. First, we assess the model's fit along the moments targeted during calibration. Second, we examine untargeted moments to evaluate the model's out-of-sample performance. Third, we decompose labor earnings into hours worked and earnings per hour, and report consumption allocations across groups.

Throughout this section, we denote skill groups by $i \in \{CS, CNS, NC\}$ (college STEM, college non-STEM, and non-college), nativity status by $k \in \{US, F\}$ (U.S.-born and foreign-born), and potential experience groups by $j \in \{1-5,\ 6-10,\ \ldots,\ 36-40,\ 41-45\}$, as defined in Section 3.

Targeted Moments We first assess the model's ability to replicate the empirical moments targeted during calibration, as described in Section 4.2.2.

Table 5 reports model-implied average labor income by skill group. The model successfully captures the substantial earnings differences across education levels, as well as within college-educated groups by occupation type. The largest deviation from the data is for college STEM workers, where the model overpredicts average earnings by approximately 1.4%.

Figure 2 reports model-implied average labor income by skill group and potential experience group. The model replicates not only the cross-sectional dispersion in earnings across skill groups, but also the hump-shaped life-cycle profile of earnings across experience groups.

Finally, Table 6 reports average labor earnings by skill group and nativity status. These moments are directly governed by the parameters θ_{iUS} and θ_{iF}, which pin down the relative earnings of U.S.-born and foreign-born workers within each skill group. The model closely matches the data along this dimension, with all deviations below 1.5%.

Table 5: Average Labor Income by Skill Group

	Model	**Data**	**% Dev.**
All Workers	120,238	120,238	0.0
By Education			
Non-college (NC)	78,690	77,640	1.4
College (C)	178,581	176,402	1.2
Conditional on College			
Non-STEM (CNS)	163,597	162,063	0.9
STEM (CS)	210,070	207,239	1.4

Notes: Dollars are in 2024 USD.

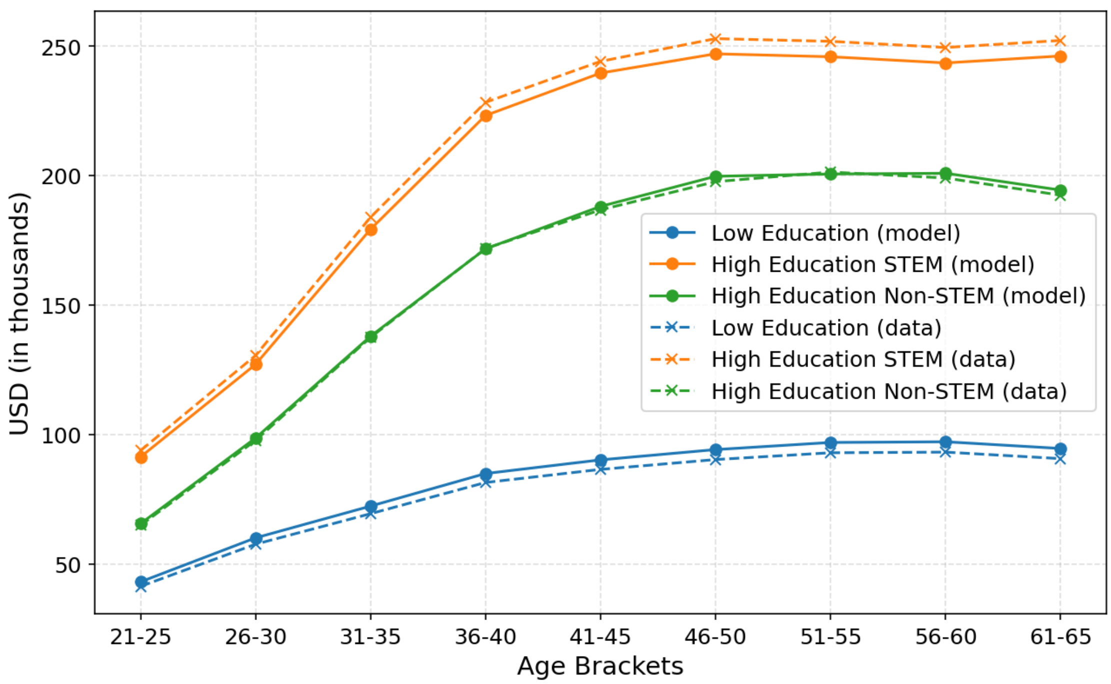


Figure 2: Average Labor Income by Skill Group and Age Bracket

Notes: Dollars are in 2024 USD.

Untargeted Moments Although the calibration targets average labor earnings by skill group and experience brackets (Figure 2), the model also generates implications for dimensions not directly targeted. We now evaluate the model's ability to match the joint distribution of earnings across education, experience, and nativity—a dimension the calibration does not directly discipline.

Figures 3–5 report model-implied and empirical average labor income separately for each skill

Table 6: Average Labor Income by Skill Group and Nativity

Group	Model	Data	% Dev.
U.S.-born			
Non-college	78,586	77,549	1.3
College Non-STEM	158,965	159,514	−0.3
College STEM	192,016	193,929	−1.0
Foreign-born			
Non-college	79,084	78,050	1.3
College Non-STEM	178,406	176,999	0.8
College STEM	247,002	249,128	−0.9

Notes: Dollars are in 2024 USD.

group: non-college (NC), college non-STEM (CNS), and college STEM (CS) households, disaggregated by nativity and experience.

Despite not targeting these joint moments directly, the model fits them well. The key mechanism is the set of parameters θ_{iUS} and θ_{iF}, which act as *level shifters*: they scale the calibrated life-cycle earnings profiles (Figure 2) up or down to match the average earnings gap between U.S.-born and foreign-born workers within each skill group. Because the life-cycle shape is common within a skill group, the nativity gap is approximately proportional across experience brackets, consistent with the patterns observed in the data.

5.1 Decomposing Earnings: Hours and Hourly Wages

Total labor earnings reflect two margins: hours worked and earnings per hour. Decomposing earnings along these margins is informative for understanding the sources of earnings inequality across groups. In particular, it reveals whether high-earning groups command higher wages, supply more labor, or both—and whether the same forces operate symmetrically across nativity status.

Table 7 reports model-implied average hours worked, earnings per hour, and consumption by skill group, all normalized relative to the all-workers average. The earnings advantage of college-educated workers reflects both margins: they work approximately 11% more hours and earn 36% more per hour than the average worker. Within the college group, STEM workers supply more

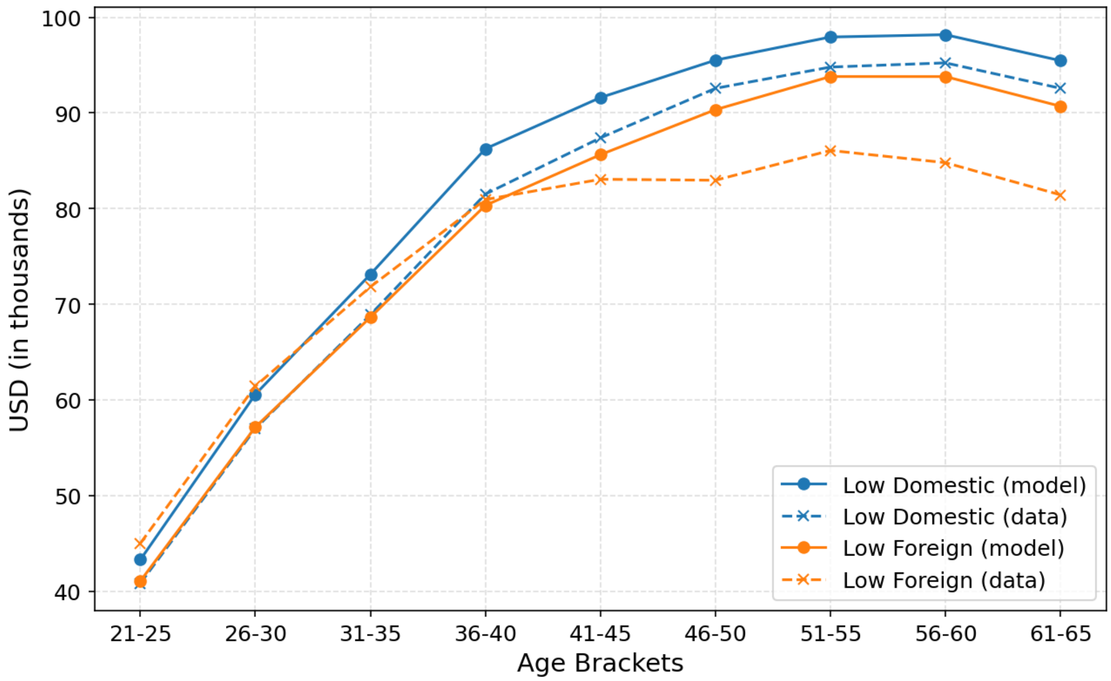


Figure 3: Average Labor Income by Nativity and Age Bracket: Non-College Households

Notes: Dollars are in 2024 USD.

hours and earn a higher hourly wage than non-STEM workers. Consumption closely tracks total earnings across skill groups.

Table 8 further disaggregates these outcomes by nativity status. A notable pattern emerges: foreign-born workers supply substantially more hours than their U.S.-born counterparts across all skill groups. For non-college workers, this gap is particularly stark—foreign-born non-college workers supply 23% more hours than U.S.-born non-college workers (1.08 vs. 0.88)—yet earn considerably less per hour (0.64 vs. 0.77). This suggests that foreign-born non-college workers compensate for lower hourly wages by working longer hours. Among college-educated groups, foreign-born workers both work more hours and, in the case of STEM workers, earn a substantially higher hourly wage (1.66 vs. 1.45), consistent with positive selection into high-skill immigration.

Figure 6 displays the life-cycle profile of hours worked by skill group, nativity, and experience bracket. The figure reveals that the hours gap between foreign-born and U.S.-born workers is persistent across the life cycle, rather than concentrated at particular career stages.

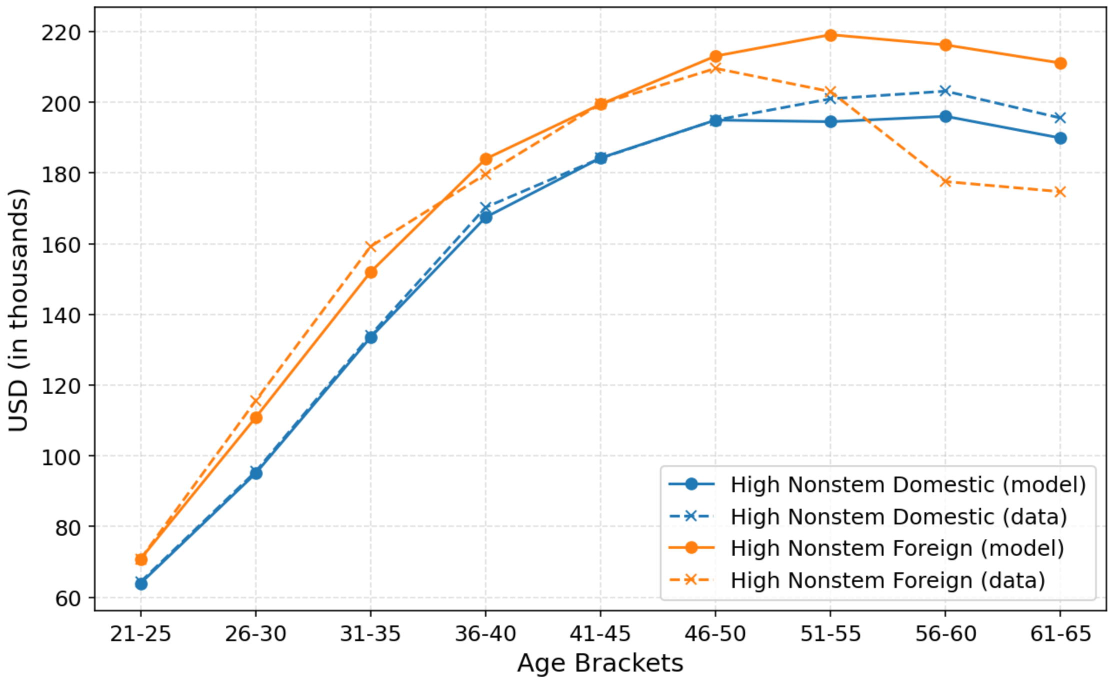


Figure 4: Average Labor Income by Nativity and Age Bracket: College Non-STEM Households

Notes: Dollars are in 2024 USD.

5.2 Earnings per Hour Worked

Figure 7 reports the life-cycle profile of earnings per hour worked. Hourly earnings exhibit a steeper life-cycle gradient for college-educated workers than for non-college workers, consistent with higher returns to experience in skill-intensive occupations. Among college-educated workers, foreign-born STEM workers command the highest hourly earnings at all experience levels, while the nativity gap in hourly wages is smaller and less systematic for non-STEM workers.

5.3 Consumption

Figure 8 reports the life-cycle profile of consumption. Consumption inequality across groups broadly mirrors earnings inequality, but is slightly compressed: the consumption gap between the highest-earning group (foreign-born college STEM) and the lowest-earning group (foreign-born non-college) is narrower than the corresponding earnings gap. This compression reflects the role of savings and, where applicable, transfers in partially smoothing consumption relative to income

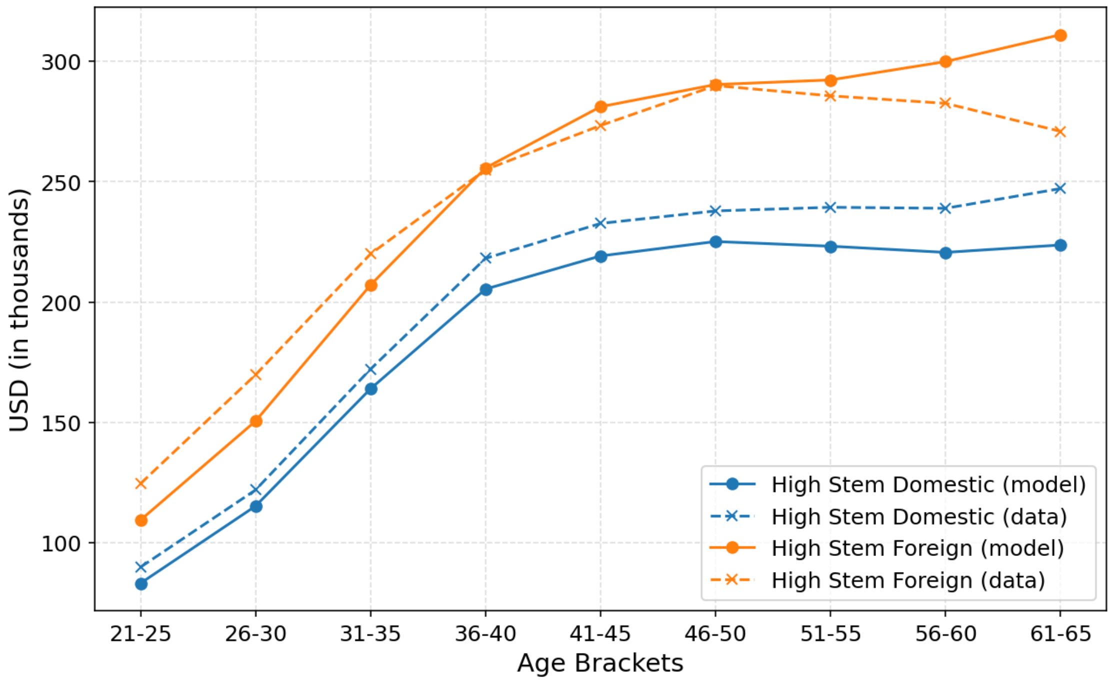


Figure 5: Average Labor Income by Nativity and Age Bracket: College STEM Households

Notes: Dollars are in 2024 USD.

over the life cycle.

Table 7: Model-Implied Average Hours Worked, Earnings per Hour, and Consumption by Skill Group

	Hours Worked	**Earnings per Hour**	**Consumption**
All Workers	1	1	1
By Education			
Non-college (NC)	0.92	0.74	0.73
College (C)	1.11	1.36	1.38
Conditional on College			
Non-STEM (CNS)	1.09	1.29	1.31
STEM (CS)	1.17	1.52	1.53

Notes: All values are expressed relative to the all-workers average, which is normalized to 1. Hours worked denotes average annual hours per worker. Earnings per hour is computed as total labor income divided by total hours worked within each group. Consumption is average per-capita consumption expenditure.

Table 8: Model-Implied Average Hours Worked, Earnings per Hour, and Consumption by Nativity and Skill Group

	Hours Worked	**Earnings per Hour**	**Consumption**
All Workers	1	1	1
U.S.-born			
Non-college (NC)	0.88	0.77	0.75
College Non-STEM (CNS)	1.05	1.29	1.33
College STEM (CS)	1.12	1.45	1.50
Foreign-born			
Non-college (NC)	1.08	0.64	0.66
College Non-STEM (CNS)	1.22	1.26	1.24
College STEM (CS)	1.27	1.66	1.59

Notes: All values are expressed relative to the all-workers average, which is normalized to 1. Hours worked denotes average annual hours per worker. Earnings per hour is computed as total labor income divided by total hours worked within each group. Consumption is average per-capita consumption expenditure.

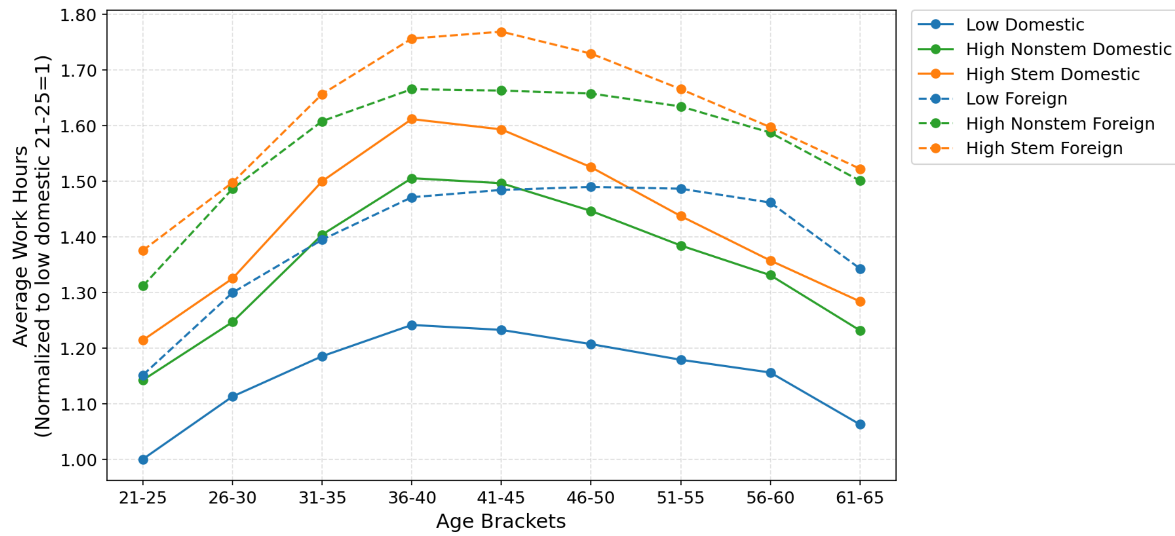


Figure 6: Average Hours Worked by Skill Group, Nativity, and Age Bracket

Notes: Average hours worked denotes annual hours per worker within each skill-nativity-age group. All values are normalized relative to U.S.-born non-college (NC) workers aged 21–25, which is set to 1. Age brackets are five-year intervals. U.S.-born and foreign-born workers are disaggregated by skill: NC = non-college, CNS = college non-STEM, CS = college STEM.

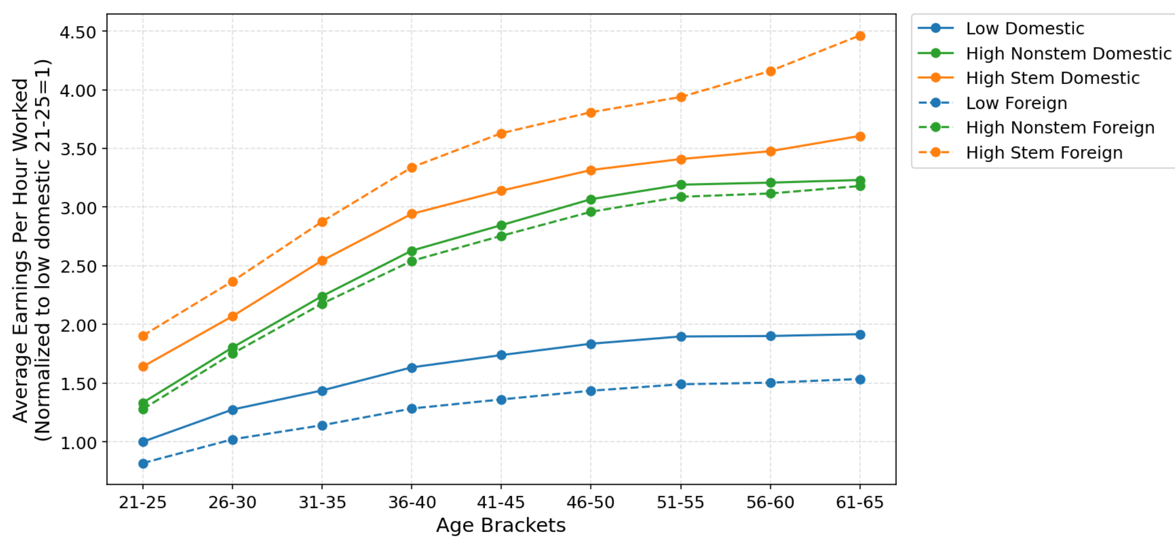


Figure 7: Average Earnings per Hour Worked by Skill Group, Nativity, and Age Bracket

Notes: Average earnings per hour is computed as total labor income divided by total hours worked within each skill-nativity-age group. All values are normalized relative to U.S.-born non-college (NC) workers aged 21–25, which is set to 1. Age brackets are five-year intervals. U.S.-born and foreign-born workers are disaggregated by skill: NC = non-college, CNS = college non-STEM, CS = college STEM.

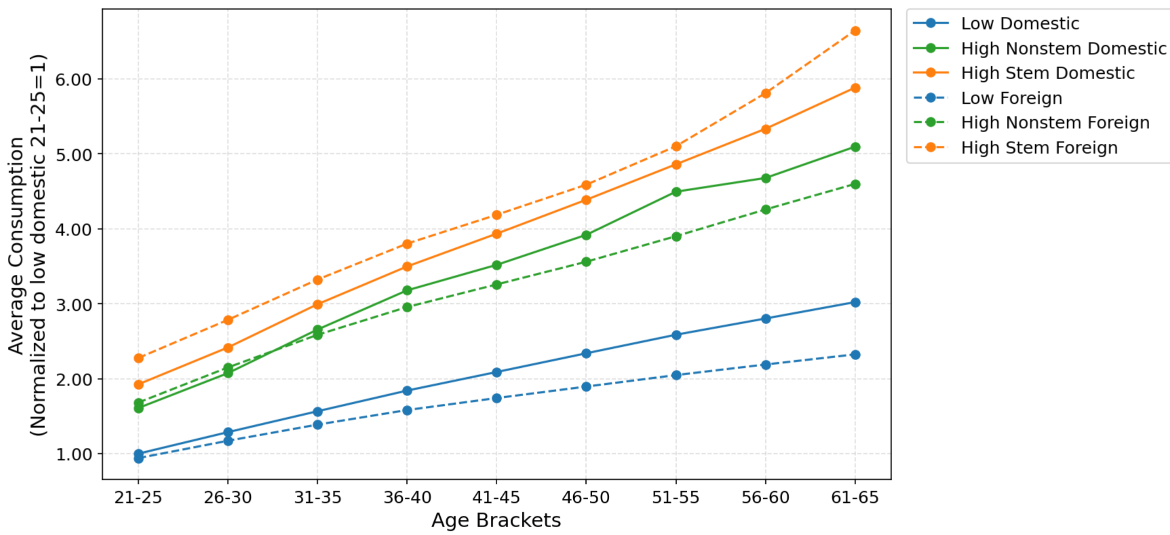


Figure 8: Average Consumption by Skill Group, Nativity, and Age Bracket

Notes: Average consumption is per-capita consumption expenditure within each skill-nativity-age group. All values are normalized relative to U.S.-born non-college (NC) workers aged 21–25, which is set to 1. Age brackets are five-year intervals. U.S.-born and foreign-born workers are disaggregated by skill: NC = non-college, CNS = college non-STEM, CS = college STEM.

6 Policy Experiment

We analyze a proposal that has been actively debated in Congress: exempting STEM immigrants from the statutory caps on U.S. permanent residency ("green cards"). The policy takes effect in 2027 in our model. We trace its effects on wages across skill and nativity groups, macroeconomic aggregates, the federal budget, and household welfare.

6.1 Institutional Background

Current U.S. immigration law divides paths to permanent residency into capped and uncapped categories. Immediate relatives of U.S. citizens enter through the primary uncapped category. All Employment-Based (EB) categories—the principal route to permanent immigration for skilled workers—are subject to an annual cap of approximately 140,000 visas.[10]

The system is further constrained by a per-country limit that restricts any single country to 7 percent of the total available visas each year. Because demand for EB visas far exceeds supply from high-sending countries, this rule generates multi-decade backlogs—particularly for applicants from India and China. Within each EB category and country of origin, applicants are processed in order of their "priority date," the date on which their employer's initial petition was filed.

6.2 The Green Card Policy

Our policy experiment exempts STEM workers from both the annual EB visa cap and the per-country limit, effectively treating STEM green cards as an uncapped category. The total number of non-STEM immigrant visas remains unchanged. This proposal follows section 80303 of the America COMPETES Act of 2022 and the Keep STEM Talent Act of 2023, neither of which was enacted. Under the proposal, applicants holding a master's or doctoral degree in a STEM field

10. The statutory annual minimum for employment-based visas is 140,000. In practice, this number often rises because any unused family-sponsored visa numbers from the preceding fiscal year "spill over" into the employment-based pool.

from a qualified institution would be eligible for permanent residency without being subject to the numerical limitations of the EB system, provided they otherwise satisfy the requirements of an existing EB preference category. The exemption would extend to the spouses and minor children of qualifying immigrants.

We implement the policy in the microsimulation model beginning in 2027, modifying the baseline immigration projections described in Section 4.4 by increasing the flow of new lawful permanent residents with advanced STEM degrees.[11] Because the exemption eliminates the per-country ceiling for STEM immigrants, it initially accelerates the admission of applicants from backlogged countries who have already been approved but are waiting for visa numbers to become available. The population effects of the policy grow over time as successive cohorts of additional immigrants arrive, remain in the country, and have U.S.-born children.

6.3 Demographic Effects

Figure 9 shows the path of the working-age population (ages 21–65) under the baseline and the policy. The gap between the two paths widens steadily over the projection horizon, reflecting the cumulative inflow of STEM immigrants and their dependents.

The most direct impact is on the share of college-educated STEM workers. Figure 10 shows that the STEM share of the working-age population rises substantially under the policy, driven almost entirely by additional foreign-born workers. This increase in STEM employment is the primary channel through which the policy raises total factor productivity, as described in Section 3.2.2. At the same time, it increases the relative supply of STEM labor within the CES aggregation hierarchy, generating the substitution effects on wages analyzed in Section 7.

The effects on other skill groups are more modest. The college-educated non-STEM population is largely unaffected (Figure 11), since the policy targets STEM immigrants specifically; any increase in non-STEM college workers arises indirectly through accompanying family mem-

11. For additional detail and an earlier analysis of this proposal using only the microsimulation model, see https://budgetmodel.wharton.upenn.edu/p/2024-01-18-budgetary-effects-of-granting-green-cards-to-immigrants-with-advanced-stem-degrees.

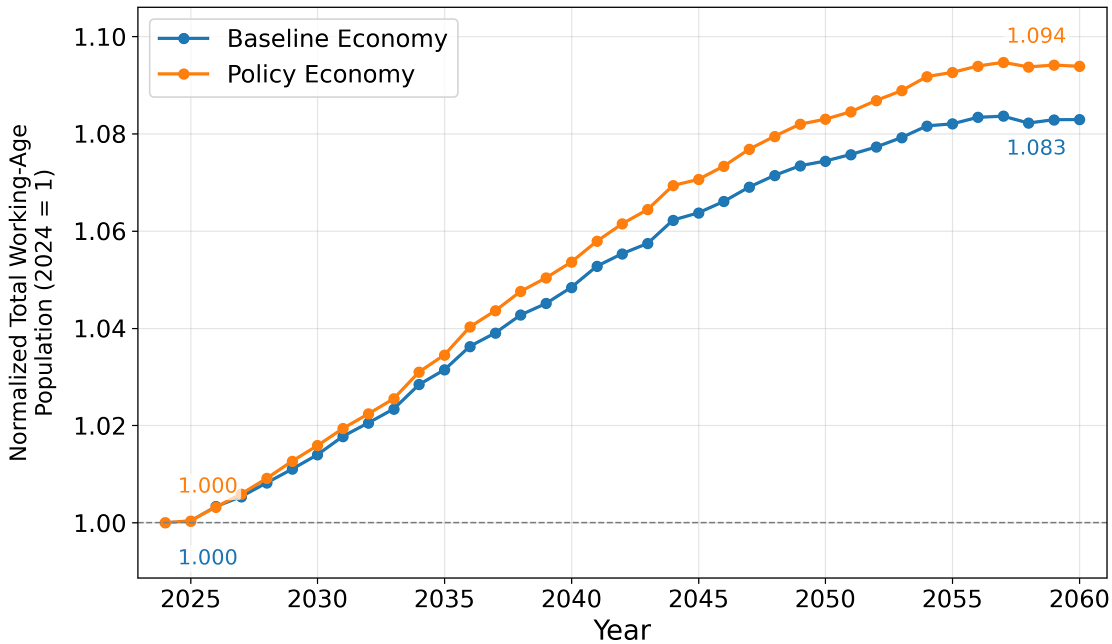


Figure 9: Working-Age Population in the Baseline and Green Card Policy

Notes: The working-age population is defined as individuals aged 21 to 65. The total population is normalized to 1 in the initial year, which coincides with the initial steady state.

bers. The share of non-college workers in the working-age population declines mechanically (Figure 12)—not because their absolute numbers fall, but because the total working-age population grows faster than this group. The compositional shift in the workforce therefore operates predominantly through the STEM channel, and the resulting changes in relative labor supply, combined with the STEM–TFP externality, are the forces that drive the economic and fiscal effects documented in the next section.

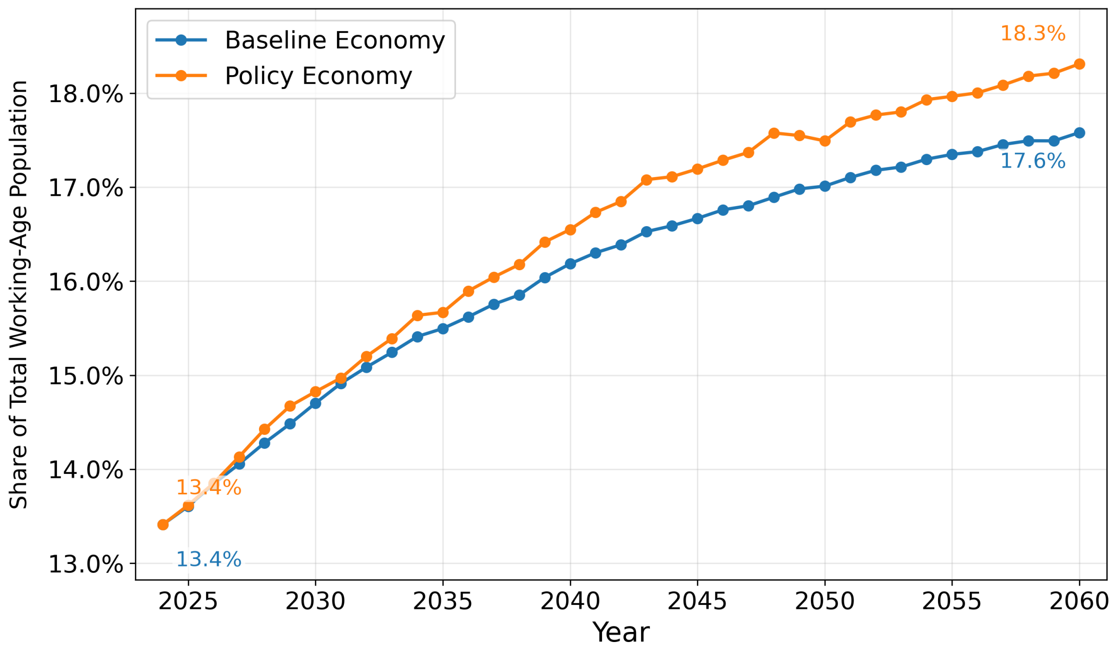


Figure 10: College-Educated STEM Working-Age Population in the Baseline and Green Card Policy

Notes: The working-age population is defined as individuals aged 21 to 65.

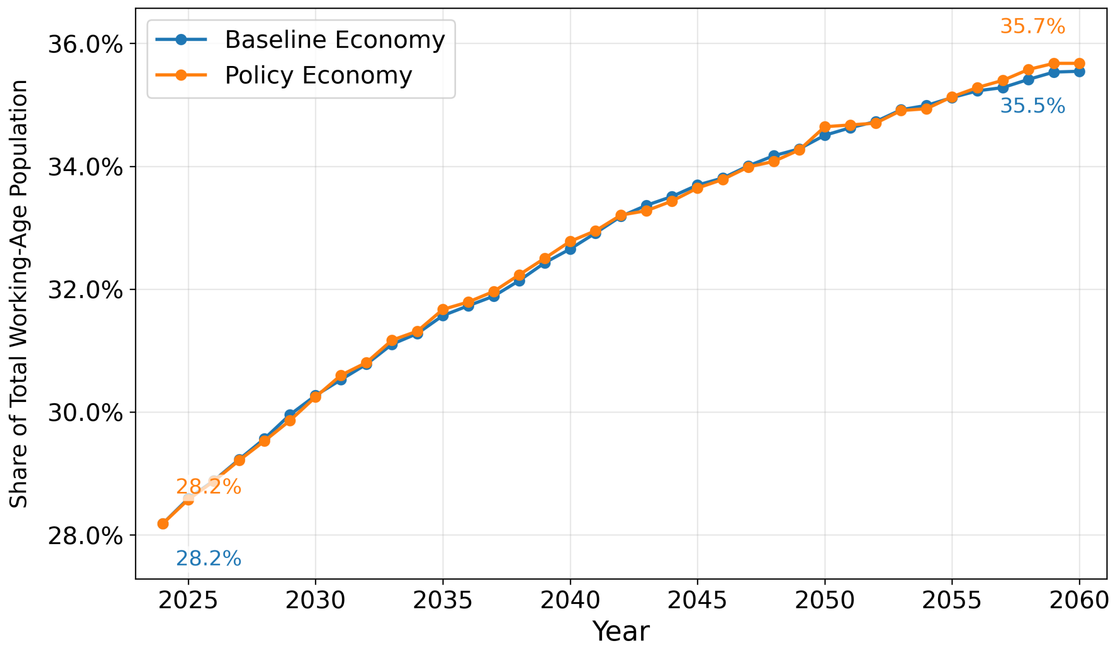


Figure 11: College-Educated Non-STEM Working-Age Population in the Baseline and Green Card Policy

Notes: The working-age population is defined as individuals aged 21 to 65.

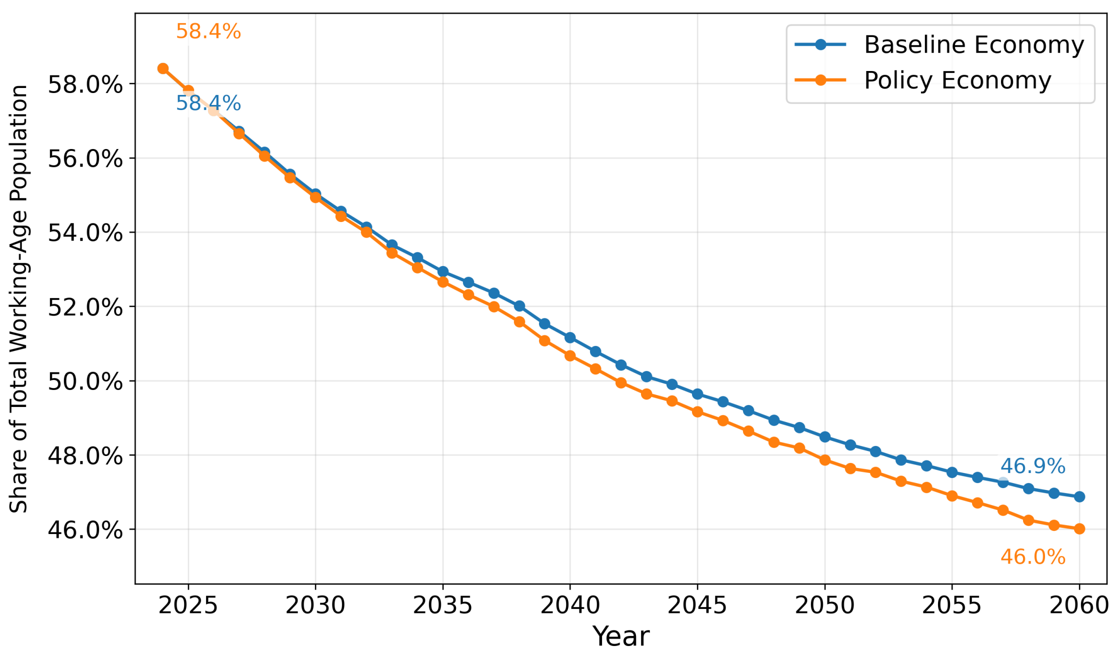


Figure 12: Non-College Working-Age Population in the Baseline and Green Card Policy

Notes: The working-age population is defined as individuals aged 21 to 65.

7 Effects of Policy Changes

7.1 Effects on the economy

Tables 9 and 10 present the macroeconomic effects of the Green Card Policy on production-side and household-side aggregates, respectively. On the production side, the most prominent result is the large and persistent increase in the capital stock, which rises from negligible levels in 2029 to 3.5 percent by 2059. The aggregate labor input also increases substantially, reaching 1.7 percent by 2059, reflecting the net effect of a larger working-age population and the small hours-worked adjustments documented below. Total factor productivity rises more modestly, from 0.3 percent in 2029 to 0.9 percent by 2059. Although TFP gains are smaller in magnitude than the factor input responses, the productivity channel remains an important driver of output growth: in the model, TFP responds to the share of STEM workers in total employment, so the inflow of high-education STEM immigrants under the Green Card Policy directly raises economy-wide productivity. Output grows steadily over the projection period, reaching 3.4 percent by 2059, driven by the combined expansion of labor, capital, and TFP. The increasing contribution of capital over the projection horizon is consistent with the transitional dynamics of the model: households gradually accumulate assets in response to the permanently higher productivity path, and capital accumulation accelerates as higher output raises saving and investment.

The household-side aggregates in Table 10 confirm these patterns. Private consumption grows in step with output, reaching 2.6 percent by 2059, while aggregate work hours rise by a more modest 1.2 percent. Private wealth accumulation is initially negligible but builds to 1.2 percent by 2059, reflecting the gradual transition toward a higher steady-state capital stock. The aggregate wage — measured as average compensation per efficiency unit of labor — increases by 2.2 percent by 2059, driven by TFP gains that raise the marginal product of labor on average, even as compositional effects reduce wages for specific groups. Taken together, these results indicate that the Green Card Policy generates substantial macroeconomic gains, with output rising by over 3 percent within three decades. The gains are driven by a combination of labor force expansion,

Table 9: Production Function Aggregate Variables

Year	Output	TFP	Labor	Capital
2029	0.5	0.3	0.2	0.0
2034	0.6	0.4	0.3	0.1
2039	1.1	0.5	0.7	0.4
2044	1.8	0.7	1.2	1.0
2049	2.4	0.8	1.5	1.8
2054	2.9	0.8	1.6	2.6
2059	3.4	0.9	1.7	3.5

Notes: Values represent the percent difference under the Green Card Policy relative to the baseline for each respective variable.

Table 10: Household Aggregate Variables

Year	Private Consumption	Work Hours	Private Wealth	Wage
2029	0.3	0.1	0.0	0.4
2034	0.5	0.1	0.0	0.5
2039	0.8	0.4	0.1	0.7
2044	1.1	0.8	0.3	0.9
2049	1.5	0.9	0.6	1.3
2054	2.0	1.1	0.9	1.6
2059	2.6	1.2	1.2	2.2

Notes: Values represent the percent difference under the Green Card Policy relative to the baseline for each respective variable.

induced capital accumulation, and the productivity externality associated with STEM labor, with factor accumulation playing the quantitatively largest role.

7.2 Effects on the federal budget

Table 11 reports the effects of the Green Card Policy on key federal budget variables. The policy produces a favorable fiscal impact that strengthens over time, driven by the combination of a larger and more productive workforce generating higher tax revenues while government outlays increase only modestly.

On the revenue side, federal revenues rise from 0.3 percent above baseline in 2032 to 2.7

Table 11: Deficit Effects of the Green Card Policy

	2032	**2037**	**2042**	**2047**	**2052**	**2057**
Outlays	0.1	0.1	0.2	0.2	0.3	0.4
Revenues	0.3	0.9	1.3	2.0	2.3	2.7
Primary deficit	-1.5	-7.0	-13.6	-32.1	-46.0	-38.6
Debt held by the public	-0.2	-0.5	-1.1	-2.0	-3.2	-4.4

Notes: Values represent the percent difference under the Green Card Policy relative to the baseline for each respective variable.

percent by 2057. This growth reflects the expanding tax base associated with higher aggregate labor income, consumption, and capital income documented in the preceding subsection. The revenue effect accelerates notably after 2042, as the cumulative increase in the STEM workforce and the associated TFP gains translate into substantially higher taxable income across the economy.

On the expenditure side, federal outlays increase only marginally — from 0.1 percent in 2032 to 0.4 percent by 2057. The modest outlay response reflects the fact that the additional immigrants under the Green Card Policy are predominantly working-age, high-education individuals who draw relatively little from means-tested transfer programs.

Because revenues grow substantially faster than outlays, the primary deficit declines markedly relative to the baseline. By 2052, the primary deficit is 46.0 percent lower than under current law, though the effect moderates somewhat to 38.6 percent by 2057. The cumulative fiscal improvement results in a progressively lower stock of debt held by the public, which falls 4.4 percent below the baseline by 2057. This debt reduction reflects the compounding effect of smaller annual deficits over the projection period, reinforced by lower interest expenditures on a smaller debt stock.

7.3 Effects on labor earnings by worker type

Table 12 shows that, although the Green Card Policy raises average labor earnings at the aggregate level, these gains mask substantial heterogeneity across skill and nativity groups. In aggregate, the Green Card Policy consistently produces higher average labor income compared to the current-law baseline, with the effect growing over time and reaching 2.4 percent by 2059. However, this

Table 12: Average Labor Income, by Skill and Nativity

	Low Education		High Education Non-STEM		High Education STEM		All Economy
Year	**Domestic**	**Foreign**	**Domestic**	**Foreign**	**Domestic**	**Foreign**	
2029	0.2	0.4	0.2	0.5	0.5	-1.9	0.3
2034	0.5	0.5	-0.1	-0.3	0.0	-1.0	0.3
2039	0.7	1.6	0.0	-1.0	0.0	-1.4	0.6
2044	1.2	2.0	0.6	0.2	0.7	-2.0	1.1
2049	1.7	2.3	1.2	0.3	0.8	-1.7	1.6
2054	2.3	2.1	1.1	0.9	1.1	-1.4	1.9
2059	3.3	2.4	1.0	0.8	1.4	-1.5	2.4

Notes: Values represent the percent difference in average labor income under the Green Card Policy relative to the baseline.

aggregate improvement is unevenly distributed. Low-education workers experience sizable gains — reaching 3.3 percent for domestic workers by 2059 — whereas high-education STEM foreign workers face persistently negative effects on their average labor earnings.

The aggregate increase is primarily driven by gains in total factor productivity (TFP) associated with the growing STEM workforce, together with the resulting rise in capital accumulation. The observed heterogeneity reflects substitution patterns across labor types: as the supply of STEM workers expands, their wages tend to decline. In the model, the labor substitution effect dominates the positive income effect from higher TFP, leading to declines in average earnings for high-education STEM workers. For low-education workers, by contrast, their growing relative scarcity reinforces the TFP-driven income gains, generating a strong increase in average labor earnings.

Figure 15 illustrates the underlying demographic mechanism, showing a substantial increase in the STEM working-age population driven predominantly by foreign-born workers.

7.4 Decomposing labor income: hours worked versus earnings per hour

To better understand the distributional results reported in Table 12, we decompose the policy effects into changes in average hours worked and average earnings per hour across demographic groups. Table 13 reports the effects on earnings per hour, while Table 14 presents the corresponding results

Table 13: Average Earnings Per Hour, by Skill and Nativity

	Low Education		High Education Non-STEM		High Education STEM		All Economy
Year	**Domestic**	**Foreign**	**Domestic**	**Foreign**	**Domestic**	**Foreign**	
2029	0.4	0.6	0.4	0.0	0.0	-1.0	0.4
2034	0.7	0.9	0.1	-0.1	-0.0	-1.0	0.5
2039	1.1	1.5	0.1	-0.5	0.2	-1.6	0.6
2044	1.3	1.8	0.4	-0.0	0.4	-1.7	0.9
2049	1.9	2.4	0.7	0.3	1.1	-1.6	1.4
2054	2.4	2.3	1.0	0.8	1.6	-1.6	1.7
2059	3.3	3.1	1.3	0.8	1.4	-1.1	2.3

Notes: Values represent the percent difference in average earnings per hour under the Green Card Policy relative to the baseline.

for hours worked.

A key finding is that most of the variation over time and across demographic groups is driven by changes in average earnings per hour, rather than hours worked. The earnings-per-hour results closely mirror the labor income patterns in Table 12. Low-education workers — both domestic and foreign — experience steadily rising hourly earnings, reaching approximately 3.3 and 3.1 percent, respectively, by 2059. High-education STEM foreign workers, by contrast, face persistent hourly earnings losses on the order of 1.0 to 1.7 percent throughout the projection period, consistent with the increased supply of STEM labor reducing their marginal product. High-education non-STEM workers occupy an intermediate position: their hourly earnings effects are initially near zero but turn modestly positive by mid-century as the general equilibrium benefits of higher TFP and capital accumulation diffuse across the economy.

By contrast, the hours-worked responses in Table 14 are uniformly small, rarely exceeding half a percentage point in absolute value, and display no clear trend over time. This muted labor supply response is consistent with the relatively low Frisch elasticities embedded in the model's preference specification. The small magnitude of the hours channel implies that the distributional consequences of the Green Card Policy operate almost entirely through price (wage) effects rather than quantity (hours) adjustments.

Table 14: Average Hours Worked, by Skill and Nativity

	Low Education		High Education Non-STEM		High Education STEM		All Economy
Year	**Domestic**	**Foreign**	**Domestic**	**Foreign**	**Domestic**	**Foreign**	
2029	-0.3	-0.3	-0.2	0.4	1.0	-1.4	-0.1
2034	-0.2	-0.5	-0.3	-0.4	0.5	-0.2	-0.1
2039	-0.4	0.0	-0.3	-0.7	0.3	0.1	-0.1
2044	-0.2	0.3	0.0	-0.2	0.5	-0.5	0.2
2049	-0.2	0.0	0.2	-0.4	-0.1	-0.2	0.1
2054	-0.0	-0.2	-0.0	-0.4	-0.3	0.2	0.2
2059	0.1	-0.6	-0.3	-0.3	0.3	-0.6	0.2

Notes: Values represent the percent difference in average hours worked under the Green Card Policy relative to the baseline.

7.5 Effects on household consumption

Table 15 reports the effects of the Green Card Policy on average private consumption by skill and nativity group. Consumption is a more comprehensive measure of household welfare than labor income alone, as it reflects not only wage changes but also adjustments in savings behavior, transfers, and the returns to accumulated wealth.

The consumption results largely reinforce the distributional patterns observed for labor earnings but reveal some notable differences. Low-education households — both domestic and foreign — experience consumption gains that grow steadily over time, reaching approximately 3.2 and 2.8 percent, respectively, by 2059. These gains are broadly comparable to the corresponding labor income effects reported in Table 12, suggesting that higher wages are the primary channel through which consumption increases for this group.

High-education STEM domestic workers exhibit a notable divergence between their consumption and labor income trajectories. While their average labor income effects remain modestly positive throughout, their consumption effects are initially slightly negative (-0.5 percent in 2029) before rising to 2.4 percent by 2054. This pattern is consistent with forward-looking households adjusting their savings behavior in anticipation of longer-run gains from capital accumulation and rising TFP, which gradually translate into higher wealth and consumption.

Table 15: Average Consumption, by Skill and Nativity

	Low Education		High Education Non-STEM		High Education STEM		All Economy
Year	Domestic	Foreign	Domestic	Foreign	Domestic	Foreign	
2029	0.4	0.6	0.3	-0.5	-0.5	-1.1	0.2
2034	0.7	1.1	0.3	-0.8	0.0	-1.5	0.4
2039	1.0	1.7	0.3	-1.4	0.3	-2.4	0.5
2044	1.4	2.2	0.5	-1.4	0.7	-2.6	0.7
2049	1.9	2.6	0.6	-0.7	1.5	-2.5	1.2
2054	2.4	2.4	1.2	-0.2	2.4	-2.3	1.7
2059	3.2	2.8	1.3	-0.2	1.9	-1.6	2.1

Notes: Values represent the percent difference in average consumption under the Green Card Policy relative to the baseline.

High-education STEM foreign workers are the only group that experiences persistently negative consumption effects, ranging from -1.1 to -2.6 percent. For this group, the adverse wage effects documented in Tables 12 and 13 translate directly into lower consumption, and the general equilibrium benefits are insufficient to fully offset the substitution effect on their earnings. High-education non-STEM foreign workers also experience consumption losses during the first two decades of the projection, though these effects attenuate toward zero by 2059.

At the aggregate level, average consumption gain reaches 2.1 percent by 2059, indicating that the policy generates broad-based welfare improvements for the economy as a whole despite the losses concentrated among foreign-born STEM workers.

7.6 Effects on labor income by worker type and potential experience

7.6.1 Average labor income by experience level for high-educated STEM workers

Table 16 reports the effect of the Green Card Policy on average labor income for high-educated domestic STEM workers, and Table 17 reports the corresponding effect for high-educated foreign STEM workers.

As established in Table 12, foreign STEM workers experience the largest decline in labor income among all groups. The results in this section decompose that aggregate effect by potential

Table 16: Average Labor Income, High STEM Domestic, by Potential Experience

Year	**21-25**	**26-30**	**31-35**	**36-40**	**41-45**	**46-50**	**51-55**	**56-60**	**61-65**
2029	0.0	0.7	-0.2	1.2	0.4	0.5	0.8	0.2	0.9
2034	1.3	0.7	-1.1	0.2	0.7	-1.0	0.8	1.0	-0.1
2039	1.9	1.8	-0.8	-0.0	0.0	0.5	-0.7	1.0	-0.4
2044	4.2	1.1	0.6	1.6	0.6	0.3	0.2	0.2	0.3
2049	3.3	2.2	1.7	1.6	1.0	0.2	0.6	-0.3	-1.4
2054	2.8	2.8	2.2	1.6	0.5	0.9	-0.1	0.9	-0.9
2059	5.8	1.7	1.8	2.3	1.8	0.9	1.1	0.6	2.3

Notes: Values represent the percent difference in average labor income under the Green Card Policy relative to the baseline.

experience, identifying the specific cohorts that drive the outcome.

Because the policy predominantly increases the entry of younger foreign STEM workers, Table 17 shows that the largest losses are concentrated among individuals in their 20s and early 30s. Losses also extend to workers in their 30s and 40s in the years following implementation, though these effects are attenuated by the offsetting gains from higher aggregate TFP.

High-educated domestic STEM workers, by contrast, exhibit an overall increase in average labor earnings (Table 12), with gains concentrated among younger cohorts. Given the relatively strong labor complementarities between foreign and domestic STEM workers ($\sigma_{NAT_{CS}} = 5.33$), younger domestic STEM workers benefit most from the inflow of foreign STEM labor—particularly in the medium to long run, as a substantial share of foreign STEM workers becomes established in the U.S. labor market. Older cohorts benefit less from these complementarities but still experience moderate gains through the aggregate increase in TFP.

7.6.2 Average labor income by experience level for high-educated non-STEM workers

Table 18 reports the effect of the Green Card Policy on average labor income for high-educated domestic non-STEM workers, and Table 19 reports the corresponding effect for high-educated foreign non-STEM workers.

High-educated non-STEM workers are affected through two main channels: labor complemen-

Table 17: Average Labor Income, High STEM Foreign, by Potential Experience

Year	**21-25**	**26-30**	**31-35**	**36-40**	**41-45**	**46-50**	**51-55**	**56-60**	**61-65**
2029	-6.3	-4.1	-1.1	-0.9	-1.1	-0.9	-0.7	-1.3	0.0
2034	1.3	-2.6	-1.3	-0.6	0.5	-0.7	-0.7	0.7	-0.6
2039	-6.6	-2.7	-2.0	-0.5	-1.2	-0.6	-0.9	1.1	-0.4
2044	-6.8	-4.4	-3.0	-2.6	-0.9	-1.5	-0.2	-1.6	0.5
2049	-4.8	-3.5	-3.5	-1.7	-1.8	-0.5	-1.0	-0.5	-0.2
2054	-5.5	-3.6	-2.2	-1.4	-0.5	-1.8	-1.2	-0.4	0.5
2059	-4.7	-3.7	-1.7	-1.7	-0.8	-1.0	-1.7	-1.6	-0.3

Notes: Values represent the percent difference in average labor income under the Green Card Policy relative to the baseline.

tarities with high-educated STEM workers ($\sigma_{STEM} = 4.90$) and complementarities between high- and low-educated workers ($\sigma_{COL} = 1.81$). The substantial policy-induced increase in the high-educated STEM population generates two opposing forces. First, non-STEM workers become relatively scarcer within the high-education group, which tends to raise their wages. Second, the expansion of the high-education group as a whole increases its relative abundance compared to low-educated workers, exerting downward pressure on wages.

These two channels operate symmetrically across domestic and foreign high-educated non-STEM workers. Consequently, the net effect on labor income largely tracks the time path of TFP, with limited impacts in the early years and progressively larger gains over time.

The population shares of high-educated non-STEM workers, shown in Figure 14, provide additional context. Although the population of foreign non-STEM workers increases over the period, the magnitude of this change is insufficient to generate substantial heterogeneity in labor income effects across experience groups.

7.6.3 Average labor income by experience level for low-educated workers

Table 20 reports the effect of the Green Card Policy on average labor income for low-educated domestic workers, and Table 21 reports the corresponding effect for low-educated foreign workers.

The mechanism for low-educated workers closely parallels that described for high-educated

Table 18: Average Labor Income, High Non-STEM Domestic, by Potential Experience

Year	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65
2029	1.0	-0.0	0.1	-0.2	0.8	-0.7	-0.1	0.6	0.0
2034	-0.2	0.3	-0.1	-0.5	-0.4	1.4	-1.0	-0.6	0.1
2039	-1.9	-0.6	0.7	-0.3	-0.4	-0.1	1.7	-0.7	0.4
2044	0.7	0.3	0.5	0.8	-0.2	0.7	1.0	1.3	0.2
2049	0.4	0.4	1.2	0.4	0.7	0.8	1.4	1.9	2.0
2054	1.4	1.2	0.7	1.0	0.6	1.1	1.5	1.7	1.0
2059	-0.5	1.1	1.5	1.2	0.4	0.7	1.2	1.5	1.9

Notes: Values represent the percent difference in average labor income under the Green Card Policy relative to the baseline.

Table 19: Average Labor Income, High Non-STEM Foreign, by Potential Experience

Year	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65
2029	3.7	1.4	0.3	0.4	0.4	1.0	0.4	0.3	-0.4
2034	-1.1	0.2	-0.0	-0.8	0.3	-0.1	0.9	0.3	-0.5
2039	-0.5	-0.9	-1.9	-1.5	-1.2	0.3	-0.5	-0.4	0.7
2044	-0.6	0.4	0.2	1.0	0.5	0.1	1.0	-0.7	0.9
2049	1.7	0.1	-0.3	1.4	1.0	-0.5	-0.2	1.2	-0.1
2054	-2.7	0.6	0.0	1.9	1.6	0.8	0.7	0.7	0.6
2059	2.1	0.0	-0.2	1.7	1.6	2.0	1.5	1.2	0.5

Notes: Values represent the percent difference in average labor income under the Green Card Policy relative to the baseline.

non-STEM workers. Regardless of nativity, low-educated workers benefit from the policy because the inflow of high-educated STEM workers increases the relative supply of high-educated labor, rendering low-educated workers comparatively scarcer and placing upward pressure on their wages.

The population shares of low-educated workers, shown in Figure 13, further inform this result. Although the share of low-educated foreign workers declines modestly over time, the magnitude is too small to generate meaningful heterogeneity in labor income effects across experience groups. The aggregate increase in TFP therefore emerges as the dominant force, producing broad-based gains in labor income across nearly all experience levels among low-educated workers over the period under study.

Table 20: Average Labor Income, Low Domestic, by Potential Experience

Year	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65
2029	-0.4	0.4	0.5	0.3	0.1	0.3	0.4	0.3	0.4
2034	-0.1	0.4	0.7	0.6	0.5	0.6	0.5	0.5	0.6
2039	0.3	0.2	0.5	0.8	0.8	0.7	0.9	1.0	0.9
2044	0.9	1.0	0.9	1.2	1.2	1.4	1.2	1.7	1.6
2049	1.4	1.3	1.3	1.6	2.0	2.0	2.2	1.9	2.3
2054	1.9	2.0	2.0	2.4	2.4	2.4	2.4	2.4	2.2
2059	3.0	3.1	2.7	2.7	3.3	3.1	3.2	3.4	3.6

Notes: Values represent the percent difference in average labor income under the Green Card Policy relative to the baseline.

8 Conclusion

This paper develops a stochastic, heterogeneous agent overlapping generations model to study the economic and fiscal effects of STEM immigration. The framework integrates three channels that have previously been studied in isolation: skill-specific labor market complementarities between native and immigrant workers, a productivity channel through which STEM workers raise total factor productivity, and a detailed fiscal apparatus encompassing Social Security, Medicare, Medicaid, and ACA subsidies, with endogenous deficits and debt accumulation. We provide new empirical estimates of both the STEM–TFP elasticity and native–immigrant substitution elasticities across skill groups and embed them in the model.

We apply the framework to evaluate a proposal to exempt STEM immigrants from green card caps. The policy generates substantial and broad-based economic gains: output rises 3.4 percent relative to current law by 2059, driven by the expansion of labor inputs, capital accumulation, and TFP growth. The productivity channel, while modest in magnitude relative to factor accumulation, plays a central role by placing the economy on a permanently higher productivity path that in turn stimulates capital deepening.

Table 21: Average Labor Income, Low Foreign, by Potential Experience

Year	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65
2029	-0.8	-0.3	0.1	0.3	0.5	0.5	0.8	0.7	-0.3
2034	-0.5	-0.2	-0.2	0.2	0.4	0.3	0.3	0.7	1.2
2039	0.7	0.4	0.6	0.6	1.2	1.6	1.5	1.1	1.5
2044	2.2	3.2	2.2	1.8	1.3	1.7	1.7	1.6	1.2
2049	2.6	2.4	2.7	2.6	2.2	1.4	1.8	2.2	1.9
2054	1.7	2.3	2.9	3.5	3.1	2.2	1.4	2.0	2.6
2059	2.7	2.9	2.9	3.0	3.8	3.0	2.2	2.2	2.8

Notes: Values represent the percent difference in average labor income under the Green Card Policy relative to the baseline.

The distributional effects are heterogeneous but largely favorable. Low-educated workers—both domestic and foreign-born—experience the largest gains, as their growing relative scarcity reinforces the TFP-driven increase in their marginal product. Domestic STEM workers also benefit, particularly younger cohorts who gain from strong complementarities with foreign STEM labor. Foreign-born STEM workers are the only group that experiences persistently negative wage and consumption effects, as increased labor supply within their skill cell dominates the economy-wide productivity gains. The policy also produces a significant fiscal dividend: federal revenues rise well above outlays, reducing debt held by the public by 4.4 percent by 2057.

Several directions for future research emerge naturally from our analysis. First, our model treats the STEM–TFP relationship as a reduced-form externality; embedding a microfounded innovation sector would allow for richer feedback between immigration, R&D investment, and growth. Second, we abstract from endogenous migration decisions; incorporating self-selection on the part of potential immigrants would permit the analysis of policies that affect not only the quantity but also the quality of the immigrant inflow. Third, our fiscal apparatus models the federal budget in detail but treats state and local governments in reduced form; given that the fiscal impact of immigration differs across levels of government, a richer intergovernmental structure could yield additional insights. Finally, extending the model to incorporate occupational choice and human capital accumulation by native workers would allow us to study whether and how domestic workers respond to STEM immigration by reallocating across occupations or investing in complementary skills.

References

Basu, Susanto, John G. Fernald, and Miles S. Kimball. 2006. "Are Technology Improvements Contractionary?" [In en]. *American Economic Review* 96, no. 5 (December): 1418–1448. https://doi.org/10.1257/aer.96.5.1418.

Ben-Gad, Michael. 2008. "Capital–skill complementarity and the immigration surplus" [in en]. *Review of Economic Dynamics* 11, no. 2 (April): 335–365. https://doi.org/10.1016/j.red.2007.08.001.

Bernstein, Shai, Rebecca Diamond, Abhisit Jiranaphawiboon, Timothy McQuade, and Beatriz Pousada. 2022. *The Contribution of High-Skilled Immigrants to Innovation in the United States.* Working Paper, December. https://doi.org/10.3386/w30797.

Blau, Francine D., and Jennifer Hunt. 2019. "The economic and fiscal consequences of immigration: highlights from the National Academies report" [in en]. *Business Economics* 54, no. 3 (July): 173–176. https://doi.org/10.1057/s11369-019-00132-9.

Bongers, Anelí, Carmen Díaz-Roldán, and José L. Torres. 2022. "Highly Skilled International Migration, STEM Workers, and Innovation" [in English]. Num Pages: 73-89 Place: Berlin, Germany, *Economics* 16 (1): 73–89. https://doi.org/10.1515/econ-2022-0022.

Bound, John, Gaurav Khanna, and Nicolas Morales. 2017. "Understanding the Economic Impact of the H-1B Program on the United States." In *High-Skilled Migration to the United States and Its Economic Consequences,* 109–175. University of Chicago Press, May.

Busch, Christopher, Dirk Krueger, Alexander Ludwig, Irina Popova, and Zainab Iftikhar. 2020. "Should Germany have built a new wall? Macroeconomic lessons from the 2015-18 refugee wave." *Journal of Monetary Economics,* SI:NOV2019 CRN CONFERENCE, 113 (August): 28–55. https://doi.org/10.1016/j.jmoneco.2020.04.004.

Caiumi, Alessandro, and Giovanni Peri. 2024. *Immigration's Effect on US Wages and Employment Redux.* Working Paper, April. https://doi.org/10.3386/w32389.

Clemens, Michael A. 2023. *The Fiscal Effect of Immigration: Reducing Bias in Influential Estimates.* Technical report. Center for Global Development.

Colas, Mark, and Dominik Sachs. 2024. "The Indirect Fiscal Benefits of Low-Skilled Immigration" [in en]. *American Economic Journal: Economic Policy* 16, no. 2 (May): 515–550. https://doi.org/10.1257/pol.20220176.

Comin, Diego, Javiar Quintana, Tom Schmitz, and Antonella Trigari. 2025. "Revisiting Productivity Dynamics in Europe: A New Measure of Utilization-Adjusted TFP Growth-Supplementary Material." Comin.etal:2025, *Journal of the European Economic Association* 23, no. 4 (April): 1598–1633.

Fehr, Hans, Sabine Jokisch, and Laurence J. Kotlikoff. 2004. "The Role of Immigration in Dealing with the Developed World's Demographic Transition." *FinanzArchiv / Public Finance Analysis* 60 (3): 296–324.

Fernald, John. 2025. *TFP Data.* https://www.johnfernald.net/TFP, February.

Guerreiro, Joao, Sergio Rebelo, and Pedro Teles. 2020. "What is the optimal immigration policy? Migration, jobs, and welfare." *Journal of Monetary Economics,* SI:NOV2019 CRN CONFERENCE, 113 (August): 61–87. https://doi.org/10.1016/j.jmoneco.2020.04.009.

Gunadi, Christian. 2019. "An inquiry on the impact of highly-skilled STEM immigration on the U.S. economy." *Labour Economics* 61 (December): 101751. https://doi.org/10.1016/j.labeco.2019.101751.

Havranek, Tomas, Zuzana Irsova, Lubica Laslopova, and Olesia Zeynalova. 2024. "Publication and Attenuation Biases in Measuring Skill Substitution." *The Review of Economics and Statistics* 106, no. 5 (September): 1187–1200. https://doi.org/10.1162/rest_a_01227.

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. 2010. "How Much Does Immigration Boost Innovation?" [In en]. *American Economic Journal: Macroeconomics* 2, no. 2 (April): 31–56. https://doi.org/10.1257/mac.2.2.31.

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics* 107 (1): 35–78. https://doi.org/10.2307/2118323.

Lee, Ronald D., and Lawrence R. Carter. 1992. "Modeling and Forecasting U.S. Mortality." Lee.Carter:1992 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.1992.10475265, *Journal of the American Statistical Association* 87, no. 419 (September): 659–671. https://doi.org/10.1080/01621459.1992.10475265.

Lucas, Robert E. 1988. "On the mechanics of economic development." *Journal of Monetary Economics* 22, no. 1 (July): 3–42. https://doi.org/10.1016/0304-3932(88)90168-7.

Moretti, Enrico. 2004. "Chapter 51 - Human Capital Externalities in Cities." In *Handbook of Regional and Urban Economics,* 4:2243–2291. https://doi.org/https://doi.org/10.1016/S1574-0080(04)80008-7.

Nathan, Max. 2013. "The Wider Economic Impacts of High-Skilled Migrants: A Survey of the Literature" [in en]. *SSRN Electronic Journal,* https://doi.org/10.2139/ssrn.2342525.

Ottaviano, Gianmarco I. P., and Giovanni Peri. 2012. "Rethinking the Effect of Immigration on Wages." *Journal of the European Economic Association* 10, no. 1 (February): 152–197. https://doi.org/10.1111/j.1542-4774.2011.01052.x.

Penn Wharton Budget Model. *Overview of Penn Wharton Budget Model's Demographic and Economic Microsimulation (demsim).*

Peri, Giovanni. 2012. "The Effect Of Immigration On Productivity: Evidence From U.S. States." *The Review of Economics and Statistics* 94, no. 1 (February): 348–358. https://doi.org/10.1162/REST_a_00137.

Peri, Giovanni, Kevin Shih, and Chad Sparber. 2015. "STEM Workers, H-1B Visas, and Productivity in US Cities" [in en]. *Journal of Labor Economics* 33 (S1): S225–S255. https://doi.org/10.1086/679061.

Peri, Giovanni, and Chad Sparber. 2009. "Task Specialization, Immigration, and Wages" [in en]. *American Economic Journal: Applied Economics* 1, no. 3 (July): 135–169. https://doi.org/10.1257/app.1.3.135.

Social Security Administration. 2018. *Annual Statistical Supplement to the Social Security Bulletin, 2017.* Technical report. March.

Storesletten, Kjetil. 2000. "Sustaining Fiscal Policy through Immigration." *Journal of Political Economy* 108 (2): 300–323. https://doi.org/10.1086/262120.

Appendix

A Composition-Adjusted TFP: Identification and Estimation

Let L_t be the labor input to the Cobb-Douglas production function

$$Y_t = A_t L_t^{\alpha} K_t^{1-\alpha}, \tag{A.1}$$

where L_t a composite of an arbitrary nested CES model. Adopting the setup of Ottaviano and Peri (2012), consider each worker as having $N+1$ characteristics, $c_0, ..., c_N$, where characteristic 0 is common to all workers. At the level of characteristic 1, the labor force is partitioned into M_1 groups which differ with respect to characteristic 1; the same follows down to characteristic N. The elasticity and productivity parameters for workers with arbitrary k-th characteristic are allowed to depend on the values of characteristics $k-1, ..., 1$; let the vector $v_c = (c_1, ..., c_N)$ denote a representative worker with the corresponding characteristics, which we refer to as a cell. In a competitive market, the log-marginal product of a worker with characteristics $(c_1, ..., c_N)$ for $N > 1$ is (suppressing the time subscript t for convenience)

$$\begin{aligned}\log\left(w_{c_1,...,c_N}\right) = &\frac{1}{\sigma_0}\log\left(L\right) + \sum_{k=1}^{N}\log\left(\theta_{c_1,...,c_k}\right)\\ &+ \sum_{k=1}^{N-1}\left(\frac{1}{\sigma_{c_1,...,c_{k+1}}} - \frac{1}{\sigma_{c_1,...,c_k}}\right)\log(L_{c_1,...c_k}) - \frac{1}{\sigma_{c_1,...,c_N}}\log(L_{c_1,...,c_N})\\ &+ \log\left(\alpha A \kappa^{1-\alpha}\right)\end{aligned} \tag{A.2}$$

where $\kappa = K/L$ is the capital-labor ratio and $w_{c_1,...,c_N}$ is average wage of a worker with those characteristics. We assume the labor supply of each cell is pre-determined before each period t, so is exogenous to contemporaneous productivity shocks.

A.1 Model of Endogenous Productivity Growth via Spillovers from STEM Workers

For notational convenience, we write $S \equiv N_{CS}$ for aggregate STEM employment throughout this appendix. Restating (6), TFP takes the following form:

$$A(S,t) = S^{\chi}\bar{A}(t) \tag{A.3}$$

where $\bar{A}(t)$ is an exogenous component of productivity and S^{χ} is an endogenous component. S^{χ} captures the spillover effect of the number of STEM workers, S, on aggregate productivity. Totally differentiating log A with respect to time

$$\frac{\frac{dA}{dt}}{A} = \frac{\chi}{S}\frac{dS}{dt} + \frac{1}{\bar{A}}\frac{d\bar{A}}{dt}.$$

From the above specification, $\frac{\partial \log A_t}{\partial \log S} = \chi$ is the elasticity of TFP to the number of STEM workers.

This formulation of the productivity spillovers has been applied to study the national impact of high skilled STEM immigration on the wages of U.S.-born workers by Gunadi (2019), the impact of foreign-born STEM workers on productivity in U.S. cities by Peri, K. Y. Shih, and Sparber (2014), and the increase in TFP in the IT sector contributed by the number of computer scientists by Bound, Khanna, and Morales (2017).

To clarify the interpretation of (A.3), note that the above specification does not suggest an 'innovations' productivity boost, wherein STEM workers contribute new ideas and technologies which permanently boost the growth rate of TFP. We refer to formulation (A.3) as the level formulation, to distinguish it from the aforementioned growth rate or 'innovations effect'. The level formulation can be interpreted as encompassing two effects: an indirect knowledge spillover effect and a direct idiosyncratic improvement in production efficiency. In a technologically advanced economy, many tasks regardless of occupation require technical knowledge. The presence of STEM workers can improve the productivity of non-STEM workers via the transmission of information about the use of new or existing technologies and consequent improvements the aggregate knowledge base–this is the knowledge or learning spillover effect. Furthermore, STEM workers can directly boost productivity by resolving idiosyncratic problems in production. Note that while our formulation ignores the aforementioned innovations effect, it does not preclude 'one-off' improvements in production efficiency attributable to the size of the STEM workforce. In particular, consider idiosyncratic problems which may arise in production. If STEM workers have unique knowledge or analytical skills, their presence can facilitate the solution of such unanticipated problems. Thus STEM workers can be seen as reducing these frictions. Note that STEM workers may not be remunerated for such contributions, due to their unanticipated nature–in which case this effect is indeed an externality.

We note that the productivity effect of STEM workers is likely a combination between the level and innovations effects. Thus our resulting estimate may overstate the 'level' effect of STEM workers.

A.2 TFP Identification

We begin by reviewing the approach standard in the literature on relative labor demand and supply.

When nested in the Solow growth model, the production function (A.1) implies that in the long run the economy follows a balanced growth path, along which the real interest rate r and the aggregate capital–output ratio are both constant while the capital–labor ratio κ grows at a constant rate proportional to the growth rate of TFP. In particular with constant, exogenous return to capital r, then the Cobb-Douglas aggregate production function (A.1) implies

$$\begin{aligned} \frac{\partial Y_t}{\partial K_t} &= (1-\alpha)A_t L_t^{\alpha} K_t^{-\alpha} = r \\ \implies Y_t/K_t &= A_t L_t^{\alpha} K_t^{-\alpha} = \frac{r}{1-\alpha} \\ \text{and } K_t/L_t &= \left(\frac{(1-\alpha)A_t}{r}\right)^{1/\alpha} \end{aligned}$$

This implies that capital K adjusts to match the supply of labor L, and that κ_t is driven by the rate of growth of TFP. Under these conditions, capital can be 'solved out' of the production function, with output Y_t expressed as a product of L_t and augmented TFP. Define augmented TFP as $\tilde{A}_t \coloneqq \left(\frac{1-\alpha}{r}\right)^{\frac{1-\alpha}{\alpha}} A_t^{\frac{1}{\alpha}}$. Whence (A.1) becomes

$$Y_t = \tilde{A}_t L_t. \tag{A.4}$$

This the assumption made in Caiumi and Peri (2024) and Ottaviano and Peri (2012).

Note that Gunadi (2019) and Peri, K. Shih, and Sparber (2015) posit a production function (with K solved out) of the form $Y = (A(S)((1-\theta)L^\rho + \theta H^\rho))^{1/\rho}$, where H and L are high and low-skilled labor, respectively. This does not influence the elasticity of A with respect to STEM workers, as the inverse substitution parameter $1/\rho$ is a scalar multiple of A in the log-marginal product. Peri, K. Shih, and Sparber (2015) derive an elasticity of A with respect to the intensity of STEM workers (i.e., proportion in employment) via cross-equation restrictions, and not directly from the wage residual. Gunadi (2019) translates this into a elasticity with respect to the number of STEM workers. This implies our estimate and the elasticity estimate in Gunadi (2019) are in principle comparable, though the wage spillover is mediated by $1/\rho$.

From equilibrium condition for capital and (A.2) it follows that, ceteris paribus, the change in $\log \tilde{A}_t$ suffices to identify TFP growth: this is the residual of the marginal product (the part unaccounted for by the nested CES model). Denote this $\epsilon_{v_c,t}$. Thus, the year-over-year difference in the residual gives

$$\Delta \log \epsilon_{v_c,t} \coloneqq \log(\epsilon_{v_c,t}) - \log(\epsilon_{v_c,t-1}) = \frac{1}{\alpha} \log(A_t/A_{t-1})$$

Thus TFP growth is identified up to labor's share, α by $\Delta \log \epsilon_{v_c,t}$. But this assumption is unrealistic in short run. Two of these reasons are immediate from the preceding discussion. First, $\Delta \log \epsilon_{v_c,t}$ depends on relative labor supply changes across cells. Second, the capital-labor ratio κ depends on aggregate labor supply shifts. Thus the standard approach fails in our setting for two reasons: functional form of the production function and short estimation period. To extract unbiased estimates of TFP growth from the wage residuals, we must account for such confounding factors.

A.2.1 Confounding Labor Supply Effects

Confounding labor supply effects arise from the CES labor aggregator. The relevant error is attributable to intertemporal exogenous shifts in labor supply within worker cells. The confounding effect can be summarized as follows: in the presence of a productivity spillover from STEM labor supply and with imperfect substitutability between STEM and non-STEM workers, intertemporal cell-wise residual change is a function of both the productivity spillover and changes in labor force composition. In the case of a growing STEM workforce, non-STEM workers will experience wage changes resulting from both complementarity and the productivity spillover (if it is non-trivial); the effect on STEM workers will depend on the magnitude of the spillover and the degree of substitutability. In particular, the wage residuals will systematically depend on changes correlated with TFP growth, so that a naive average of cross-sectional residual change will be be biased estimator of TFP growth.

We provide a detailed example. In what follows, our discussion generalizes that of Moretti (2004), who considered an analogous problem in the context of human capital externalities from education with a Cobb-Douglas labor aggregate. Consider an economy with aggregate constant returns to scale production function as in (A.4), where labor L is a composite given by a two-factor CES model for labor, composed of STEM workers ($i = S$) and non-STEM workers ($i = NS$), which are imperfect substitutes with constant elasticity of substitution $\sigma \in \mathbb{R}^{+}$:

$$L_t = \left[\theta_S {L_{S,t}}^{\frac{\sigma-1}{\sigma}} + \theta_{NS} {L_{NS,t}}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$

Then the average wage of worker type $i \in \{S, NS\}$ at time t_0

$$\log(w_{i,t_0}) = \frac{1}{\sigma}\log(L_{t_0}) + \log(\theta_i) - \frac{1}{\sigma}\log(L_{i,t_0}) + \log(\tilde{A}_{t_0})$$

Suppose there is a change in STEM labor supply L_S, with non-STEM labor L_{NS} fixed. Assume $\chi \geq 0$ (i.e., either there is no spillover effect or it is positive). Then the change in log-marginal product experienced by a worker of type i is

$$\frac{d}{dt}\log(w_{i,t_0}) = \frac{1}{\sigma}\frac{L_S^{-1/\sigma}\frac{dL_S}{dt}}{L_{t_0}^{\frac{\sigma-1}{\sigma}}} - \frac{1}{\sigma}\frac{\frac{dL_S}{dt}}{L_i} + \frac{d\tilde{A}_{t_0}}{dL_S}\frac{dL_S}{dt}$$

Applying (A.3), the change in wage for STEM workers is thus

$$\frac{d}{dt}\log(w_{S,t_0}) = \frac{1}{\sigma}\frac{\frac{dL_S}{dt}}{L_{t_0}^{\frac{\sigma-1}{\sigma}}L_S^{1/\sigma}} - \frac{1}{\sigma}\frac{\frac{dL_S}{dt}}{L_S} + \frac{\chi}{S}\frac{dS}{dt}$$

and for non-STEM workers

$$\frac{d}{dt}\log(w_{NS,t_0}) = \frac{1}{\sigma}\frac{\frac{dL_S}{dt}}{L_{t_0}^{\frac{\sigma-1}{\sigma}}L_S^{1/\sigma}} + \frac{\chi}{S}\frac{dS}{dt}.$$

Thus the change in log-marginal product is depends on the labor force composition at t_0, the elasticity of substitution, and the productivity spillover. For a non-STEM worker, an increase in STEM labor supply necessarily raises their marginal product via relative supply effects (complementarity). If $\chi > 0$, the wage of a non-STEM worker will gain an additional benefit from the productivity spillover. The wage increase experienced by a STEM worker will be reduced according to the strength of complementarities between STEM and non-STEM workers. Thus, except of the case of perfect substitutability ($\sigma = \infty$), the *observed* productivity effect will not be uniform across worker characteristics. In particular, it will be biased upward for non-STEM workers and biased downward for STEM workers. The magnitude of the bias depends not only on the elasticity of substitution, but also labor force composition at the previous period, i.e., the bias is time-varying.

Under the nested CES structure, shifts in relative labor composition across multiple characteristics imply the residual change for a worker of characteristic v_c depends on the relative change in the labor supply of that cell, the changes in relative labor supply for each level in the nesting, and the productivity spillover. Thus labor supply changes by education, occupation, experience, and nativity must be controlled for. While controlling for this by approximating the changes in relative supply is possible, the granularity of the cells make such an approximation necessarily of high variance, and could potentially introduce bias. To avoid explicit 'backing-out' of supply error, we model each cell as noisy repeated measurements of the same component, with an error covariance structured according to the CES nesting. This not only allows us to indirectly control for supply effects, but also to absorb unmodeled heterogeneity influencing changes in cell wages.

A.2.2 Capital-Labor Ratio, Productivity Growth, and Utilization

As mentioned above, the short sample period renders the standard practice of assuming balanced capital growth untenable. Sluggish capital-labor response to productivity shocks and aggregate labor supply shifts will confound TFP measurement. Additionally, business cycle fluctuations introduce another source of confounding in TFP measurement. To address these concerns we decompose the residual (A.2) with a structural model of productivity and capital growth and adjustments for variable factor utilization.

Capital and Productivity Growth As can be seen from (A.2), if capital intensity growth lags productivity improvements, wage increases will be biased downward relative to the balanced growth path. Recall under the balanced growth assumption, $\kappa_t = ((1-\alpha)A_t/r)^{1/\alpha}$. The log-change with respect to time is then

$$\frac{d}{dt}\log(w_t) = \frac{d}{dt}\log(A_t) + (1-\alpha)\frac{d}{dt}\log(\kappa_t).$$

Note that, under the balanced growth assumption,

$$\frac{d\log(\kappa_t)}{dt} = \frac{1}{\alpha}\frac{d}{dt}\log(A_t) + \frac{1}{\alpha}\frac{d}{dt}\log((1-\alpha)/r) = \frac{1}{\alpha}\frac{d}{dt}\log(A_t). \tag{A.5}$$

We relax the above by allowing for short-term deviations from the long-run equilibrium relationship between a_t and k_t. Denote by a_t and κ_t the log levels of TFP and capital-labor ratio, respectively and by $\dot{a}_t$ and $\dot{\kappa}_t$ the respective change from $t-1$ to t, e.g., $\dot{a}_t = \log(A_t/A_{t-1})$. Casting the model in state space form, we assume state transition equations of form

$$x_t = Fx_{t-1} + \epsilon_t$$
$$F = \begin{pmatrix} \phi_{aa} & 0 \\ \phi_{ka} & \phi_{kk} \end{pmatrix} \tag{A.6}$$
$$x_t^T = \begin{pmatrix} \dot{a}_t & \dot{\kappa}_t \end{pmatrix}$$

where $\epsilon_t \sim \mathcal{N}(0, Q)$. We assume temporally uncorrelated innovations, i.e., $E[\epsilon_t^T \epsilon_{t-\tau}] = 0\ \forall \tau \in \mathbb{Z} \setminus \{0\}$. We assume productivity and capital growth processes stationary and assuming the system is nondegenerate, i.e., $0 < |\phi_{aa}| < 1, 0 < |\phi_{kk}| < 1$. Following standard assumptions in the literature, we further assume $\phi_{aa} > 0$, i.e., technology shocks beget further technology shocks.

Consider the response of κ_t to a contemporaneous innovation in TFP. We re-express ϵ_t by decomposing Q into its Cholesky factorization, $Q = LL^T$, so that

$$\epsilon_t = Le_t$$

where $e_t \sim \mathcal{N}(0, I)$. Parameterize L as

$$L = \begin{pmatrix} \sigma_a & 0 \\ \rho\sigma_a/\alpha & \sigma_\kappa \end{pmatrix}$$

where $0 < \rho < 1$ and $\sigma_a > 0, \sigma_k > 0$. σ_a is the standard deviations of innovations to $\dot{a}_t$ and σ_k is standard deviation of the innovations to $\dot{k}_t$ which are independent of those to $\dot{a}_t$, respectively. Note that when $\rho = 1$, the model implies expected instantaneous capital adjustment to the long-run equilibrium level. Otherwise, ρ defines the strength of contemporaneous adjustment by capital to

TFP shocks.

To guarantee the model in (A.6) defines a system such that TFP and capital converge to the levels implied by the balanced growth path, the system must satisfy the following matrix difference equation:

$$\begin{pmatrix} \frac{1}{1-\phi_{aa}} \\ \frac{1}{\alpha}+\frac{1}{\alpha}\frac{\phi_{aa}}{1-\phi_{aa}} \\ 0 \\ 0 \end{pmatrix} = \lim_{t\to\infty} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & \phi_{aa} & 0 \\ 0 & 0 & \phi_{ak} & \phi_{kk} \end{pmatrix}^t \begin{pmatrix} 0 \\ 0 \\ 1 \\ \rho/\alpha \end{pmatrix} = \lim_{t\to\infty} C^t \begin{pmatrix} 0 \\ 0 \\ 1 \\ \rho/\alpha \end{pmatrix}.$$

where we assume $a_0 = \kappa_0 = 0$ and we approximate (log) levels by the sum of first order changes. The above problem says that, under the assumption of no further shocks, conditional on a shock 1 to a at $t = 0$, that κ_t must converge to the balanced-growth path level implied by contemporaneous shock and to the additional change in TFP resulting from the autocorrelation ϕ_{aa}. The eigendecomposition of the matrix C^t is

$$\begin{pmatrix} 0 & 1 & \frac{\phi_{aa}-\phi_{kk}}{(\phi_{aa}-1)\phi_{ka}} & 0 \\ 1 & 0 & \frac{1}{\phi_{aa}-1} & \frac{1}{\phi_{kk}-1} \\ 0 & 0 & \frac{\phi_{aa}-\phi_{kk}}{\phi_{ka}} & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \phi_{aa}^t & 0 \\ 0 & 0 & 0 & \phi_{kk}^t \end{pmatrix} \begin{pmatrix} 0 & 1 & \frac{\phi_{ka}}{(\phi_{aa}-1)(\phi_{kk}-1)} & \frac{1}{1-\phi_{kk}} \\ 1 & 0 & \frac{1}{1-\phi_{aa}} & 0 \\ 0 & 0 & \frac{\phi_{ka}}{\phi_{aa}-\phi_{kk}} & 0 \\ 0 & 0 & \frac{-\phi_{ka}}{\phi_{aa}-\phi_{kk}} & 1 \end{pmatrix}.$$

From which, under the assumption of stationarity of $\dot{a}_t$ and $\dot{\kappa}_t$, we obtain the identifying conditions:

$$\lim_{t\to\infty} C^t \begin{pmatrix} 0 \\ 0 \\ 1 \\ \rho/\alpha \end{pmatrix} = \begin{pmatrix} \frac{1}{1-\phi_{aa}} \\ \frac{\phi_{ka}}{(\phi_{aa}-1)(\phi_{kk}-1)} + \frac{\rho}{\alpha(1-\phi_{kk})} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{1-\phi_{aa}} \\ \frac{1}{\alpha}\frac{\phi_{aa}}{1-\phi_{aa}} + \frac{1}{\alpha} \\ 0 \\ 0 \end{pmatrix}.$$

For the long-run balanced growth assumption to hold, we need to satisfy

$$\frac{\phi_{ka}}{(\phi_{aa}-1)(\phi_{kk}-1)}+\frac{\rho}{\alpha(1-\phi_{kk})}=\frac{1}{\alpha}\frac{\phi_{aa}}{1-\phi_{aa}}+\frac{1}{\alpha}.$$

Applying stationarity conditions and the above constraint, we identify ϕ_{kk} and ϕ_{ka} in terms of α, ρ, ϕ_{aa}. As we assume $0<\rho<1$, the requirements are satisfied with $\phi_{kk}=1-\rho$, which implies $\phi_{ka}=\frac{\rho\phi_{aa}}{\alpha}$.

We note that the above eigendecomposition of C in fact fails if $\phi_{aa}=\phi_{kk}$. However, the Jordan form exists and it can be verified that the long-run conditions are still satisfied in this case.

As a final remark, we note the resulting autoregressive process for capital-labor growth gives rise to the form of a partial adjustment model, though this independent of our derivation (see, e.g., Sargent (1978)).

Variable Factor Utilization Measuring change in aggregate technology over short periods is complicated by the presence of unobserved systemic non-technological factors in the residual. In particular, the presence of unobserved variations in factor utilization over the business cycle are known to have a confounding effect on estimates of short-run changes in technology Basu, Fernald, and Kimball (2006). To control for these effects, we adapt the procedure of Basu, Fernald, and Kimball (2006) (hereafter, BFK) to estimate utilization adjusted TFP. We begin with a brief overview of the theoretical framework underlying BFK.

The microfoundations of BFK are constructed to provide a simple proxy for variable factor utilization by modeling *changes in unobserved input utilization via changes in observable inputs*. In the words of BFK, 'Since cost-minimizing firms operate on all margins simultaneously, changes in observed inputs can potentially proxy for unobserved utilization changes.' The model assumes the representative firm only minimizes costs and is a price-taker in factor markets, abstracting from firm pricing and output behavior in goods markets. Furthermore, shadow prices of factors are assumed unobserved.

Firms have adjustment costs to labor and capital inputs, such that the inputs are quasi-fixed. In particular, employment (the number of workers) and capital inputs (machines or equipment) cannot be adjusted costless. Instead, firms can adjust the intensity of factor utilization. Labor and capital utilization are linked via a shift premium, whereby the disutility of increasing hours-per-worker is linked to the cost of increasing capital intensity. This is plausible within the framework of a firm optimizing on all margins: increases in labor intensity should be coupled with increases in capital intensity.

BFK augment the Cobb-Douglas production function to incorporate variable labor and capital utilization. In particular, (A.1) becomes

$$Y_t = A_t(E_tL_t)^{\alpha}(Z_tK_t)^{1-\alpha} \tag{A.7}$$

where E_t is the intensity of labor utilization and Z_t the intensity of capital utilization.

The resulting log-linearized Cobb-Douglas production function around its steady state results in a decomposition of output growth as

$$\dot{y}_t = (1-\alpha)\dot{k}_t + \alpha\dot{l}_t + (1-\alpha)\dot{z}_t + \alpha\dot{e}_t + \dot{a}_t = (1-\alpha)\dot{k}_t + \alpha\dot{l}_t + \dot{u}_t + \dot{a}_t \tag{A.8}$$

where $\dot{}$ indicates the log-difference in an input. $\dot{u}_t$, a function of labor and capital utilization rates, accounts for the change in output attributable to utilization.

Take $L_t = N_tH_t$ in (A.7), where N_t is the number of workers and H_t is the hours worked per worker. BFK consider assume firms are price takers and minimize costs. The representative firm minimizes the present value of expected costs

$$E_t\sum_{\tau=t}^{\infty}\prod_{j=t}^{\tau-1}\frac{1}{(1+r_j)}w_\tau G(E_\tau, H_\tau)V(Z_\tau)N_\tau + w_\tau N_\tau\Psi(D_\tau/N_\tau) + p_\tau K_\tau J(I_\tau/K_\tau)$$

which is subject to the constraint on output (A.7) and equations of motion for capital and number of workers. r_j determines the discount factor. Ψ describes the adjustment cost of adding D_τ new hires relative to current number of employees N_τ. J describes the adjustment cost of I_τ capital investment relative to current level of capital K_τ. G describes how worker compensation depends on effort E_τ, while $V(Z_\tau)$ is a shift premium, which is a function of capital utilization Z_τ. Ψ and J are convex functions, with $J(\delta) = \delta$ (where δ is the deprecation rate of capital), $J'(\delta) = 1$, and $\Psi(0) = \Psi'(0) = 0$. Finally N_τ is the number of workers, I_τ is capital investment, D_τ is the number of new hires, and p_τ and w_τ are investment price and wages, respectively. Note the firm faces no adjustment costs for H_t, E_t, and Z_t, and the corresponding first order conditions for these variables are intratemporal, so do not depend on uncertainty.

From the first order conditions for E_t and H_t, BFK apply the implicit relationship between H and E (suppressing time subscripts for convenience),

$$\frac{H\frac{\partial G}{\partial H}}{G(E,H)} = \frac{E\frac{\partial G}{\partial E}}{G(E,H)}$$

and, from assumptions on G implying E can be expressed as a strictly monotonically increasing differentiable function of H, BFK derive the following log-linearized relationship between change in effort and hours

$$\dot{e} = \beta_{eh}\dot{h}$$

where β_{eh} is the elasticity of effort with respect to hours at the steady state value of H.

To proxy for capital utilization in terms of hours, BFK again apply first order conditions at the steady state to obtain the implicit relationship

$$\frac{1-\alpha}{\alpha} = v(Z)/g(H)$$

where $v(Z) = ZV'(Z)/Z$ is the elasticity of labor cost with respect to capital utilization and $g(H) = \frac{H\frac{\partial G}{\partial H}}{G(E,H)}$ is elasticity of labor cost with respect to hours (N.B.: BFK assume capital depreciation is not affected by the level utilization Z). That the LHS is a constant determined by the capital and labor shares follows as the identification is done at the steady state. BFK assume $v(Z)$

is strictly increasing in Z, and log-linearizing about the steady-state, obtain the relationship

$$\dot{z} = \beta_{zh}\dot{h}.$$

Thus the change in utilization can be parameterized in terms of $\dot{u}_t = \beta_{zh}\dot{h}_t + \beta_{eh}\dot{h}_t = \beta\dot{h}_t$, where β is a composite coefficient accounting for labor and capital utilization.

We apply an analogous form to the growth in wage residual. Abstracting from the nesting structure and re-expressing the equation in terms of capital-labor ratio κ_t, the change in log-wage is given by

$$\dot{w}_t = (1-\alpha)\dot{k}_t + (1-\alpha)\dot{z}_t + \alpha\dot{e}_t - (1-\alpha)\dot{l}_t + \dot{a}_t = (1-\alpha)\dot{\kappa}_t + \dot{u}_t + \dot{a}_t.$$

Note that as worker effort, E_t, is unobservable, it enters positively into the residual. Applying BFK's identification of utilization from hours worked per worker,

$$\begin{aligned}\dot{w}_t &= (1-\alpha)\dot{\kappa}_t + (1-\alpha)(\dot{z}_t - \alpha\dot{e}_t) + \dot{a}_t = (1-\alpha)\dot{\kappa}_t + ((1-\alpha)\beta_{zh} + \alpha\beta_{eh})\dot{h}_t + \dot{a}_t \\ &= (1-\alpha)\dot{\kappa}_t + \beta\dot{h}_t + \dot{a}_t\end{aligned}$$

where β_{eh} is the elasticity of effort with respect to hours worked and β_{zh} is the elasticity of capital workweek with respect to hours worked.

To render the problem consistent with the CES labor aggregate, we assume the representative firm hires uniformly across skill groups and apply the constant returns to scale assumption to the labor aggregator to parameterize $L_t = \bar{E}_t\bar{H}_t\bar{N}_t\tilde{L}_t$, so that the cost minimization labor parameters are $\bar{E}_t, \bar{H}_t, \bar{N}_t$, which are scalar factors uniform across cells; $\bar{w}_t$ is the corresponding average wage.

Note that with a labor input itself a composite it is consistent with the BFK framework for utilization to vary across worker type. However, we abstract from this in our estimation, and instead seek only to capture aggregate changes in utilization. This is consistent with our goal, as in controlling for utilization we seeking to control for business cycle fluctuations, and thus to isolate the unexplained change in wages common to all cells. Finally, note that as $\dot{h}_t$ is correlated with technology growth, $\beta\dot{h}_t$ must be estimated via two-stage least squares.

A.2.3 State Space Formulation & Estimation Strategy

We synthesize A.2.1 and A.2.2 with a state space model which controls for noise introduced by labor supply effects, the dynamics of capital adjustment, and utilization via instrumental variables regression.

The full state space form of our model is

$$x_t = Fx_{t-1} + \epsilon_t \tag{A.9}$$

$$y_t = Hx_t + \Pi Z_t + \eta_t \tag{A.10}$$

$$\epsilon_t \sim \mathcal{N}(0, Q) \tag{A.11}$$

$$\eta_t \sim \mathcal{N}(0, R) \tag{A.12}$$

where we make the standard independence assumptions on the process and measurement noise:

$$E[\epsilon_t^T \epsilon_{t-\tau}] = 0$$
$$E[\eta_t \eta_{t-\tau}^T] = 0$$
$$E[\epsilon_t \eta_t^T] = 0$$

for all $\tau \in \mathbb{Z} \setminus \{0\}$. The state transition equation (A.9) is as in section A.2.2.

In the observation equation (A.10), y is a $2n \times 1$ vector, where n is the number of labor cells and the observation matrix H is a $2n \times 2$ matrix. In detail, the observation equation has the following structure:

$$\begin{pmatrix} \tilde{y}_t \\ \dot{h}_t \end{pmatrix} = \begin{pmatrix} \tilde{H} \\ 0 \end{pmatrix} x_t + \begin{pmatrix} \beta D \\ D \end{pmatrix} z_t^T \otimes 1_{2n} + \begin{pmatrix} \beta\omega_t + \tilde{\eta}_t \\ \omega_t. \end{pmatrix}$$

The measurement vector y consist of $\tilde{y}_t$, the n annual changes in the residual by labor cell, and $\dot{h}_t$, the n annual growth in hours worked per worker by labor cell. Each non-zero row of the state observation matrix H, corresponding to $\tilde{H}$, consists of

$$\tilde{H}_{i,\cdot} = \begin{pmatrix} 1 & 1-\alpha \end{pmatrix}.$$

The exogenous regression term ΠZ_t corresponds to the instrumental estimation for utilization. In particular $Z_t = z_t^T \otimes 1_{2n}$, where z_t is the vector of instruments, and

$$\Pi = \begin{pmatrix} \beta D \\ D \end{pmatrix}$$

where $D = d^T \otimes 1_n$, where d are the first stage regression coefficients, and β as the second stage regression coefficient corresponding to the elasticity of wage growth to utilization.

Finally, the measurement error vector

$$\eta_t = \begin{pmatrix} \beta\omega_t + \tilde{\eta}_t \\ \omega_t. \end{pmatrix}$$

has covariance matrix

$$R = \begin{pmatrix} \tilde{R} + \beta^2\Omega & \beta\Omega \\ \beta\Omega & \Omega \end{pmatrix}$$

where $\tilde{\eta}_t \sim \mathcal{N}(0, \tilde{R})$ is the vector of wage growth residual deviations uncorrelated with changes in cell-wise utilization, and $\omega_t \sim \mathcal{N}(0, \Omega)$ is the vector of cell-wise deviations in growth in $\dot{h}_t$. We take Ω to be a diagonal matrix. $\beta\omega_t$ accounts for cell-wise idiosyncratic deviations from the aggregate level of utilization.

$\tilde{R}$ is parameterized in a manner which corresponds to the nested CES structure, such that the diagonal is given by

$$\tilde{R}_{(c_{i_1},\ldots c_{i_N}),(c_{i_1},\ldots c_{i_N})} = \sigma^2 + \sum_{k=1}^{N} \sigma_{c_i}^2.$$

The covariance between cells is

$$\tilde{R}_{(c_{i_1},\ldots c_{i_N}),(c_{j_1},\ldots c_{j_N})} = \sum_{c_{i_k}=c_{j_k}} \sigma_{c_i}^2.$$

Thus $\tilde{\eta}_t$ can be interpreted as a time-varying random effects vector, with each level of CES nesting (skill group, age, and nativity) contributing an additive random effect term.

Estimation To estimate the state space parameters, we maximize the prediction error decomposition form of the log-likelihood:

$$\sum_{t=1}^{T} \log p(y_t|\Theta) = \sum_{t=1}^{T} - \log \det S_t(\Theta) - \nu_t(\Theta)^T S_t(\Theta)^{-1} \nu_t(\Theta) + \text{Constant} \tag{A.13}$$

where Θ is the vector of model parameters and

$$\begin{aligned} S_t &= HP_{t|t-1}H^T + R \\ \nu_t &= y_t - Hx_{t|t-1} - \Pi Z_t \\ P_{t|t-1} &= FP_{t-1|t-1}F^T + Q \end{aligned}$$

where $P_{t-1|t-1}$ is the posterior covariance matrix at $t-1$, $\mathcal{S}_t$ is the innovation covariance matrix at t, and all other variables are defined as in the prequel.

As the state vector is assumed to be stationary, we initialize the filter at the unconditional state mean and covariance. The unconditional state covariance matrix can solved explicitly given the assumptions in A.2.2 and has values

$$\begin{aligned}
\text{Var}(a) &= \frac{\sigma_a^2}{1-\phi_{aa}^2} \\
\text{Cov}(a,k) &= \frac{\rho}{\alpha}\left(\sigma_a^2 + \frac{\text{Var}(a)\phi_{aa}^2}{1-\phi_{aa}\phi_{kk}}\right) \\
\text{Var}(k) &= \frac{\rho^2}{\alpha^2}\sigma_a^2 + \sigma_k^2 + \frac{\rho}{\alpha}\phi_{aa}\left(\frac{\phi_{aa}\rho}{\alpha}\text{Var}(a) + \frac{2\phi_{kk}\,\text{Cov}(k,a)}{1-\phi_{kk}^2}\right).
\end{aligned}$$

Problem (A.13) is a non-linear optimization problem, which we solved numerically via BFGS, introducing reparameterization to constrained parameters to mitigate numerical issues induced by introducing explicit constraints. To initialize the structural parameters for the maximization procedure we borrow from existing estimates in the literature where possible. We obtained our estimated TFP series by applying the Kalman smoother to the estimated system. Smooth state estimates differ from filtered state estimates as smoothed estimates are conditioned on all available sample information, rather than exclusively past sample information.

Table 22 compares our estimated average annual TFP growth over the period with those of Fernald (2025), who derives a measure of aggregate utilization-adjusted TFP following the method of Basu, Fernald, and Kimball 2006 (though with some modifications).

Table 22: Average Utilization-Adjusted Annual TFP Growth Rate (2001-2019)

Our estimate	**Fernald estimate**
0.66	0.73

Note: TFP growth rates are expressed as log changes multiplied by 100. Source: Fernald (2025).

Identification of Utilization The log-likelihood decomposition (A.13) and the structure of H demonstrate the estimation of the coefficients D is independent of the state covariance, and only depends only on the covariance of ω and η and the variance of ω. Under the assumption that the random instruments vector Z_t is assumed uncorrelated with X_t and η, the generated regressors DZ_t are exogenous and β is time invariant, the regressor exogeneity assumption for the Kalman filter is satisfied and the standard maximum likelihood procedure applies.

A.3 Estimation of The Elasticity of TFP to STEM Employment

To estimate the elasticity χ in (A.3), we apply two-stage-least squares regression, instrumenting for annual growth in STEM employment S. The possibility of an endogenous response of STEM employment to productivity growth raises the issue of reverse causality. For instance, if legal immigration channels in U.S. are biased towards STEM workers, and emigrants are attracted to countries with high productivity, the growth of the STEM labor force is driven by productivity. Omitted variable bias is another (in this case, related) concern. Notably, the US technology sector expanded during the study period. In the presence of productivity rises correlated with this expansion, STEM employment will depend on the intensive increase in this sector as a part of the economy. If expansion of this sector is attributable to past innovations, and subsequent productivity growth is attributable to these innovations, a naive regression will overstate the spillover of STEM employment (e.g., a firm develops a new technology which raises productivity in the technical sector; the sector expands and hires more workers, yet the rise in productivity is attributable to the past innovation, not the additional workers). These issues are related to the level and innovation effects discussed above. We note the extent to which we can isolate the level effect even with instrumentation is, however, uncertain.

Thus to isolate the causal effect of STEM workers on TFP care must be taken to ensure the shifts in STEM employment are exogenous. In particular, as the TFP growth series is estimated under an AR(1) process, the instrument must be uncorrelated with past shocks. To accomplish this, we instrument for STEM employment growth in a method inspired by Caiumi and Peri (2024). Caiumi and Peri (2024) seek an estimate of the labor supply elasticity of foreign-born workers, but recognize that changes in the skill-composition of immigrant workers may bias estimation by inducing shifts in the relative productivity of immigrants. Following Caiumi and Peri (2024), we construct different instruments for growth in U.S.-born STEM employment and foreign-born STEM employment, which we then sum to create the instrument for STEM employment. We then estimate the elasticity via two stage least squares.

To capture exogenous changes in U.S.-born employment, we utilize the demographic (age-experience) decomposition of labor supply in the nested CES labor aggregate. In particular, we construct the instrument for U.S.-born STEM employment by assuming the employment size of the j-th five year experience group within the U.S.-born STEM labor force in year $t+5$ is equal to the employment of experience group $j-1$ in year t. To predict the change in the lowest experience group, we fix the STEM skill-group share among those with the corresponding ages in employment at t and apply it to the number of workers with these ages in year $t+5$. Thus the variation in U.S.-born STEM employment depends on the size of the oldest cohort at time t (who exit the labor force), the STEM share of the lowest age-experience group at time t, and number of new workers at time $t+5$. The imputed growth in U.S.-born STEM employment is then

$$\Delta_{S,US}(t_k, t_k+5) = s_{US,\text{exp}_0}\Delta_{\text{exp}_0,US}(t_k, t_k+5) - S_{US,\text{exp}_n}(t_k)$$

where S_{exp_0} is STEM employment in the least experience group, S_{exp_n} STEM employment in the greatest, s_{US,exp_0} is the STEM share of the labor force in the age range corresponding to the least experience group, and $\Delta_{\text{exp}_0}(t, t+5)$ is the aggregate change in labor force employment at this age range. Thus the instrument for U.S.-born employment growth depends only on the past share of lowest-experience STEM workers and aggregate employment growth of the corresponding age group, net of highest experience group exits.

For foreign-born STEM employment we construct a shift-share instrument. In short, we fix 1990 STEM employment distribution by origin, and grow employment in accordance with the growth in population by origin; in particular, we fix the shares of the population employed in STEM occupations (restricted to ages in the lowest potential experience to the highest potential experience) by origin for the top five sending countries (Mexico, Cuba, China, Philippines and Korea) and seven regions (North America, Central America and Caribbean, South America, Europe, Africa, Asia and Oceania) at their 1990 proportions; we then grow each of these populations by the respective by-origin flows over the initial ten year period (1990-2000), then by succeeding five (or four, for the last period) year periods. Let $s_{o,1990}$ denote the initial STEM employment share of population by origin in 1990. The imputed growth for origin o at period t_k is then

$$\Delta_{S,F}(t_k, t_{k+1}) = \sum_{o} \sum_{\ell \leq k} s_{o,1990} \Delta_{o,t_\ell}$$

where Δ_{o,t_ℓ} is the change in population from period t_ℓ to period $t_{\ell+1}$. If $\Delta_{o,t_\ell} < 0$, we follow Caiumi and Peri (2024) in setting the change to 0. The variation in foreign-born STEM labor supply thus arises from the changing patterns in origins over the period. Note that unlike U.S.-born workers, foreign-born workers are more likely to be influenced by further lags in growth, requiring the shares to be fixed at a much earlier date.

The five year ranges are (2000-2005, 2005-2010, 2010-2015, 2015-2019) and are the same for both the foreign and native STEM employment instruments. The aggregate change instrument change between reference periods is the sum of the two components. The employment change instrument is linearly interpolated between reference periods to complete the series.

A.4 Data

A.4.1 State Space Model Inputs

Capital-Labor Ratio The observable for capital-labor ratio was derived from the percent annual growth in the series Labor Statistics, which was rescaled to approximate log-growth.

Utilization Instruments We proxy for exogenous changes is utilization using three instruments from Comin et al. (2025): monetary policy shocks, oil price shocks, and financial shocks. Annual monetary policy shocks are defined to be the averages of the monthly shocks. We define quarterly oil price shocks to be the log difference between the current quarterly real oil price and the highest real oil price over the preceding four quarters; to annualize these, sum the four quarterly price shocks. To reflect asymmetry in oil price shocks, we then define the annual shock to be the maximum of 0 and the annualized raw shock defined above. The instrument of shocks for period t is the vector of shocks for period $t-1$.

A.4.2 STEM Employment

The construction of the STEM employment instrument follows that detailed in A.3; the sample uses American Community Survey (ACS) microdata, following the exposition in B.

B Data Construction and Labor Parameters Estimation

B.1 Sample Restrictions

Our analysis uses American Community Survey (ACS) microdata covering the years 2000 through 2023, excluding 2020 due to pandemic-related data irregularities. All computations employ ACS person weights to ensure population representativeness. The sample is restricted to individuals aged 18 and older who are not residing in group quarters and who are not currently enrolled in school. For employment analyses, we include all individuals who reported working at least one week during the reference year. For wage analyses, we further exclude unpaid family workers and the self-employed, while retaining public-sector employees, and limit the sample to respondents reporting valid, positive wage income. Nominal wages are converted to 2024 dollars using the CPI for All Urban Consumers (CPI-U).

Workers are classified into skill-experience cells based on their educational attainment, STEM occupation status, and potential labor market experience. The resulting skill groups are: less than high school, high school, some college, and college or higher (with the latter split by STEM occupation status into College+ STEM and College+ Non-STEM). Potential experience is computed as age minus the assumed age of labor market entry, which varies by education level: 17 for less than high school, 19 for high school, 21 for some college, 23 for college graduates, and 24 for those with graduate degrees. The sample is restricted to workers with potential experience between 1 and 45 years, grouped into nine five-year intervals (1–5, 6–10, . . ., 41–45). This cell structure follows the education–experience classification in Ottaviano and Peri (2012) and Caiumi and Peri (2024), extended to distinguish STEM from non-STEM occupations among college-educated workers and to include an additional experience bin covering 41–45 years.

For years 2008–2018, when the ACS reported weeks worked only in intervals rather than exact values, we impute weeks worked using interval midpoints. Within each education–experience–nativity cell and year, we construct three measures. First, labor supply is measured in full-time-equivalent (FTE) units: for each individual we compute annual hours (weekly hours multiplied by weeks worked) divided by 2,000, weight by the person weight, and sum across all individuals in the cell. As an alternative supply measure, we also compute cell-level employment as the sum of person weights. Second, the cell-level average weekly wage is constructed as the hours-weighted mean of individual real weekly wages (annual wage income deflated to 2024 dollars and divided by weeks worked), where the weight for each individual is the product of annual hours and the person weight, following the procedure in Ottaviano and Peri (2012) and Caiumi and Peri (2024). All three measures are computed separately by nativity (and gender, if needed) within each cell.

Finally, occupational classifications identifying STEM workers follow the 2010 Standard Occupational Classification (SOC) Policy Committee definition, which groups STEM occupations into four subdomains: (i) life and physical science, engineering, mathematics, and information technology; (ii) social science; (iii) architecture; and (iv) health. Within each subdomain, occupations are further classified by type—research and practitioner, technologist and technician, postsecondary teaching, managerial, and sales positions.

B.2 Nested CES Structure

The labor-input aggregation follows Ottaviano and Peri (2012) and Caiumi and Peri (2024), with an additional nest among college-educated workers distinguishing STEM from non-STEM occupations. The nesting order places broader skill distinctions at the top and finer ones at the bottom, reflecting the assumption that workers who share more observable characteristics are closer substitutes. Aggregate output is produced via a Cobb–Douglas function:

$$Y_t = A_t L_t^{\alpha} K_t^{1-\alpha}, \tag{B.14}$$

where A_t denotes total factor productivity, L_t effective labor, and K_t capital.

Top-level nest: College vs. non-college labor.

$$L_t = \left[\theta_C L_{C,t}^{\frac{\sigma_{COL}-1}{\sigma_{COL}}} + \theta_{NC} L_{NC,t}^{\frac{\sigma_{COL}-1}{\sigma_{COL}}}\right]^{\frac{\sigma_{COL}}{\sigma_{COL}-1}}. \tag{B.15}$$

College nest: STEM vs. non-STEM.

$$L_{C,t} = \left[\theta_{CNS} L_{CNS,t}^{\frac{\sigma_{STEM}-1}{\sigma_{STEM}}} + \theta_{CS} L_{CS,t}^{\frac{\sigma_{STEM}-1}{\sigma_{STEM}}}\right]^{\frac{\sigma_{STEM}}{\sigma_{STEM}-1}}. \tag{B.16}$$

Non-college nest: ND, HSD, SC. To improve the precision of the experience and nativity elasticity estimates, we allow for a finer partition within non-college labor during estimation, splitting workers into three sub-groups: no high school diploma (ND), high school diploma (HSD), and some college (SC):

$$L_{NC,t} = \left[\theta_{ND} L_{ND,t}^{\frac{\sigma_{NC}-1}{\sigma_{NC}}} + \theta_{HSD} L_{HSD,t}^{\frac{\sigma_{NC}-1}{\sigma_{NC}}} + \theta_{SC} L_{SC,t}^{\frac{\sigma_{NC}-1}{\sigma_{NC}}}\right]^{\frac{\sigma_{NC}}{\sigma_{NC}-1}}. \tag{B.17}$$

In the OLG model, non-college labor is treated as a single aggregate ($L_{NC,t}$); the three-way split enters only through the estimation of the parameters below.

Experience groups. Within each of the five education–occupation groups $i \in \{ND, HSD, SC, CNS, CS\}$, nine potential-experience groups (1–5, 6–10, …, 41–45 years) are combined as

$$L_{i,t} = \left[\sum_{j=1}^{9} \theta_{ij} L_{ij,t}^{\frac{\sigma_{EXP_i}-1}{\sigma_{EXP_i}}}\right]^{\frac{\sigma_{EXP_i}}{\sigma_{EXP_i}-1}}. \tag{B.18}$$

We estimate both a pooled σ_{EXP} and skill-group-specific elasticities σ_{EXP_i} for $i \in \{NC, CNS, CS\}$; the OLG model uses the group-specific values.

Nativity groups. At the lowest level of the hierarchy, U.S.-born (US) and foreign-born (F) workers enter as

$$L_{ij,t} = \left[\theta_{iUS}\, L_{ijUS,t}^{\frac{\sigma_{NAT_i}-1}{\sigma_{NAT_i}}} + \theta_{iF}\, L_{ijF,t}^{\frac{\sigma_{NAT_i}-1}{\sigma_{NAT_i}}}\right]^{\frac{\sigma_{NAT_i}}{\sigma_{NAT_i}-1}}. \tag{B.19}$$

We estimate both an overall σ_{NAT} (pooling across all skill groups) and skill-group-specific elasticities σ_{NAT_i} for $i \in \{NC, CNS, CS\}$. The efficiency parameters θ_{iUS} and θ_{iF} vary by skill group but not by experience level. The OLG model uses the skill-group-specific nativity elasticities.

B.3 Relative Productivity Parameters

The elasticity estimation in Section B.4 requires relative productivity parameters (θ) to construct CES-weighted labor composites at each step of the sequential procedure. We estimate these parameters from CPS ASEC microdata. These parameters are distinct from the efficiency parameters calibrated within the OLG model (Section 4.2.2), which are pinned down by matching observed earnings across demographic groups in the model's steady state.

We use CPS ASEC survey years 2001–2024. Because the CPS ASEC reports income earned in the previous calendar year, we match each survey year to its corresponding income-year CPI adjustment factor when deflating wages to 2024 dollars. The sample is restricted to civilians in the labor force aged 21–64 who are salaried employees with positive wages, hours, and weeks worked, excluding top-coded wage observations. Within each year, wages, weeks worked, and usual weekly hours are winsorized at the 0.5th and 99.5th weighted percentiles to limit the influence of outliers. Hourly wages are computed as $\text{WageIncome}/(\text{Weeks}\times\text{Hours})$. STEM occupation status is assigned using the same SOC crosswalk as in the ACS data.

For each pair of groups within a given nest (e.g., U.S.-born vs. foreign-born within a skill group, or STEM vs. non-STEM within college-educated workers), the θ parameters are computed as normalized population-weighted average hourly wages, pooling across all survey years:

$$\theta_i = \frac{\bar{w}_i}{\sum_\ell \bar{w}_\ell}, \tag{B.20}$$

where $\bar{w}_i$ is the population-weighted average hourly wage for group i over the full sample period, and the sum runs over all groups within the nest. This ensures $\sum_i \theta_i = 1$ within each nest. Pooling across years yields time-invariant productivity parameters, an assumption shared with Ottaviano and Peri (2012) and Caiumi and Peri (2024). Experience-group parameters (θ_{kj} for experience group $j = 1, \ldots, 9$) are computed pooling across all skill groups.

B.4 Estimation of Elasticities

Following Caiumi and Peri (2024), elasticities of substitution are estimated sequentially, beginning at the lowest level of the CES hierarchy (nativity) and proceeding upward. At each step, estimates from the preceding level are used to construct CES-weighted labor composites for the next.

Step 1: U.S.-born–foreign-born substitution (σ_{NAT}). The overall elasticity is estimated by OLS:

$$\ln\left(\frac{w_{US,ijt}}{w_{F,ijt}}\right) = \alpha_{ij} + \alpha_t + \frac{1}{\sigma_{NAT}} \ln\left(\frac{Hours_{F,ijt}}{Hours_{US,ijt}}\right) + \varepsilon_{ijt}, \tag{B.21}$$

with education×experience (α_{ij}) and year (α_t) fixed effects, weighted by cell employment, and standard errors clustered at the education×experience level. This OLS specification follows Ottaviano and Peri (2012). Caiumi and Peri (2024) develop a shift-share instrumental variables strategy to address potential omitted variable bias from skill-cell-specific productivity shocks; their preferred 2SLS estimate for 2000–2023 implies $\sigma_{NAT} \approx 15$, close to our OLS estimate of 14.61. At the education-group level, a direct comparison is not possible because our skill partition differs from theirs, but the broad pattern is consistent: Caiumi and Peri (2024) find strong complementarity among college-educated workers ($\sigma_{NAT} \approx 9$); our finer decomposition suggests that this complementarity is concentrated in the STEM margin ($\sigma_{NAT_{CS}} = 5.33$), while college non-STEM workers are substantially more substitutable ($\sigma_{NAT_{CNS}} = 21.92$). For non-college workers, Caiumi and Peri (2024) report considerable heterogeneity across sub-groups (ranging from $\sigma \approx 9$ for workers without a high school diploma to $\sigma \approx 59$ for high school graduates); our aggregate non-college estimate of 30.86 falls between these extremes, reflecting the employment-share-weighted average.

Skill-group-specific elasticities are obtained by interacting the hours ratio with skill group indicators, replacing the single slope with group-specific coefficients while absorbing year and experience fixed effects:

$$\ln\left(\frac{w_{US,ijt}}{w_{F,ijt}}\right) = \alpha_j + \alpha_t + \sum_i \frac{1}{\sigma_{NAT_i}} \mathbf{1}[edu = i] \cdot \ln\left(\frac{Hours_{F,ijt}}{Hours_{US,ijt}}\right) + \varepsilon_{ijt}. \tag{B.22}$$

The non-college elasticity $\sigma_{NAT_{NC}}$ is a weighted average of the three sub-college coefficients (less than high school, high school, some college), with weights proportional to each group's employment share: $\sigma_{NAT_{NC}} = \sum_{i\in\{1,2,3\}} w_i/\hat{\beta}_i$, where w_i are employment shares renormalized within non-college groups. Standard errors for $\sigma_{NAT_{CNS}}$ and $\sigma_{NAT_{CS}}$ are computed via the simple delta method ($SE(1/\hat{\beta}) = SE(\hat{\beta})/\hat{\beta}^2$), while the standard error for $\sigma_{NAT_{NC}}$ uses the multivariate delta method applied to the relevant submatrix of the variance–covariance matrix.

Step 2: Experience group substitution (σ_{EXP}). U.S.-born and foreign-born hours are first aggregated within each education–experience–year cell using the CES composite with the education-specific $\hat{\sigma}_{NAT_i}$ from Step 1:

$$L_{ij,t} = \left[\hat{\theta}_{US,ij}\, Hours_{US,ij,t}^{\frac{\hat{\sigma}_{NAT_i}-1}{\hat{\sigma}_{NAT_i}}} + \hat{\theta}_{F,ij}\, Hours_{F,ij,t}^{\frac{\hat{\sigma}_{NAT_i}-1}{\hat{\sigma}_{NAT_i}}}\right]^{\frac{\hat{\sigma}_{NAT_i}}{\hat{\sigma}_{NAT_i}-1}}. \tag{B.23}$$

At this level, total cell-level labor supply is endogenous to wages, so we move to two-stage least squares (2SLS). The elasticity σ_{EXP} is estimated as:

$$\ln w_{ij,t} = \alpha_{ij} + \alpha_{ti} + \frac{1}{\sigma_{EXP}} \ln L_{ij,t} + \varepsilon_{ij,t}, \tag{B.24}$$

instrumenting $\ln L_{ij,t}$ with $\ln Hours_{F,ij,t}$ (log foreign-born hours), with year×education (α_{ti}) and education×experience (α_{ij}) fixed effects, weighted by cell employment, and standard errors clustered at the education×experience level.

Education-specific experience elasticities ($\sigma_{EXP_{NC}}$, $\sigma_{EXP_{CNS}}$, $\sigma_{EXP_{CS}}$) are obtained by running separate 2SLS regressions within each collapsed skill group (no college, college non-STEM, college STEM).

Step 3: Within-college substitution (σ_{STEM}). Experience groups are aggregated within each education–year cell using the pooled $\hat{\sigma}_{EXP}$ from Step 2, which provides a more stable composite than the noisier skill-group-specific estimates:

$$L_{i,t} = \left[\sum_{j=1}^{9} \hat{\theta}_{ij} \, L_{ij,t}^{\frac{\hat{\sigma}_{EXP}-1}{\hat{\sigma}_{EXP}}} \right]^{\frac{\hat{\sigma}_{EXP}}{\hat{\sigma}_{EXP}-1}} . \tag{B.25}$$

Data are collapsed to the education–year level. Weekly wages are reconstructed as the wage bill divided by total hours. Restricting to college-educated cells (non-STEM $i = CNS$ vs. STEM $i = CS$), data are reshaped to a time series indexed by $t = \text{Year} - 2000$:

$$\ln\!\left(\frac{w_{CNS,t}}{w_{CS,t}}\right) = \alpha + \beta\, t + \frac{1}{\sigma_{STEM}} \ln\!\left(\frac{L_{CS,t}}{L_{CNS,t}}\right) + \varepsilon_t, \tag{B.26}$$

estimated by 2SLS with $\ln(Hours_{F,CS,t}/Hours_{F,CNS,t})$ as instrument for the relative labor composite, a linear time trend to absorb secular changes in relative demand, weighted by total college employment, and heteroskedasticity-robust (HC1) standard errors.

Step 4: College vs. non-college substitution (σ_{COL}). The college aggregate is formed using the $\hat{\sigma}_{STEM}$ estimate:

$$L_{C,t} = \left[\hat{\theta}_{CS} \, L_{CS,t}^{\frac{\hat{\sigma}_{STEM}-1}{\hat{\sigma}_{STEM}}} + \hat{\theta}_{CNS} \, L_{CNS,t}^{\frac{\hat{\sigma}_{STEM}-1}{\hat{\sigma}_{STEM}}} \right]^{\frac{\hat{\sigma}_{STEM}}{\hat{\sigma}_{STEM}-1}} . \tag{B.27}$$

For the non-college aggregate, we assume that non-college education subgroups are perfect substitutes ($\sigma_{NC} \to \infty$), so the CES composite reduces to a productivity-weighted linear combination. The θ parameters still capture productivity differences across subgroups—a unit of high-school labor contributes more to the aggregate than a unit of no-diploma labor—but conditional on these weights the groups substitute one-for-one. This assumption is consistent with the large within-low-education elasticities estimated by Ottaviano and Peri (2012):

$$L_{NC,t} = \hat{\theta}_{ND}\, L_{ND,t} + \hat{\theta}_{HSD}\, L_{HSD,t} + \hat{\theta}_{SC}\, L_{SC,t}. \tag{B.28}$$

Data are reshaped to a time series of non-college vs. college relative wages:

$$\ln\left(\frac{w_{NC,t}}{w_{C,t}}\right) = \alpha + \beta_1 t + \beta_2 t^2 + \frac{1}{\sigma_{COL}} \ln\left(\frac{L_{C,t}}{L_{NC,t}}\right) + \varepsilon_t, \tag{B.29}$$

estimated by 2SLS with $\ln(Hours_{F,C,t}/Hours_{F,NC,t})$ as instrument, quadratic time controls to absorb nonlinear trends in the college wage premium, weighted by total employment, and HC1 standard errors.

Standard errors. All $\sigma = 1/\hat{\beta}$ transformations use the delta method: $SE(\sigma) = SE(\hat{\beta})/\hat{\beta}^2$. For the weighted non-college elasticity $\sigma_{NAT_{NC}} = \sum_i w_i \cdot 1/\hat{\beta}_i$, the gradient vector $\partial g/\partial \beta_i = -w_i/\hat{\beta}_i^2$ is applied to the relevant submatrix of the variance–covariance matrix to obtain $SE(\sigma_{NAT_{NC}}) = \sqrt{\nabla g' \hat{V} \nabla g}$.

B.5 Population Shares by Nativity

In this appendix, we report population shares by nativity in both the Baseline and Green Card Policy. Figure 13 presents results for low-educated workers, Figure 14 reports the corresponding results for high-educated non-STEM workers, and Figure 15 reports the results for high-educated STEM workers.

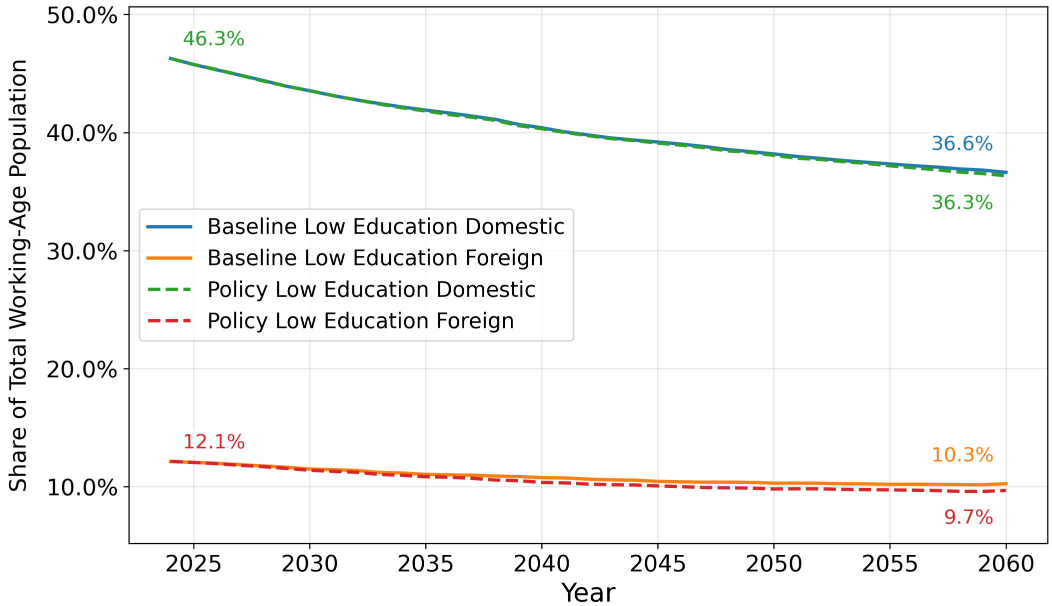


Figure 13: Low-Educated Working-Age Population in the Baseline and Green Card Policy

Notes: The working-age population is defined as individuals aged 21–65.

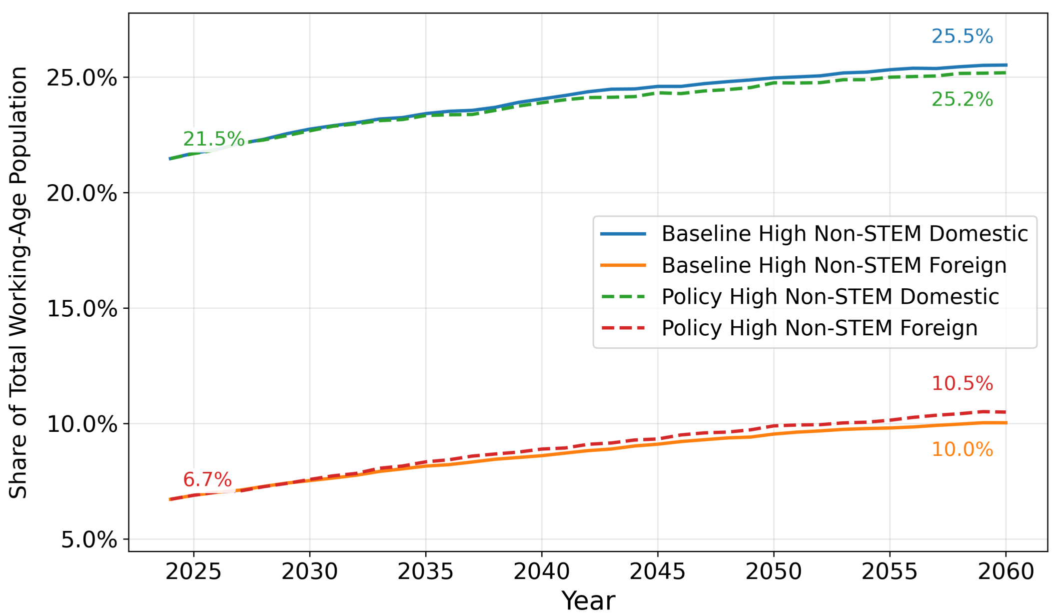


Figure 14: High-Educated Non-STEM Working-Age Population in the Baseline and Green Card Policy

Notes: The working-age population is defined as individuals aged 21–65.

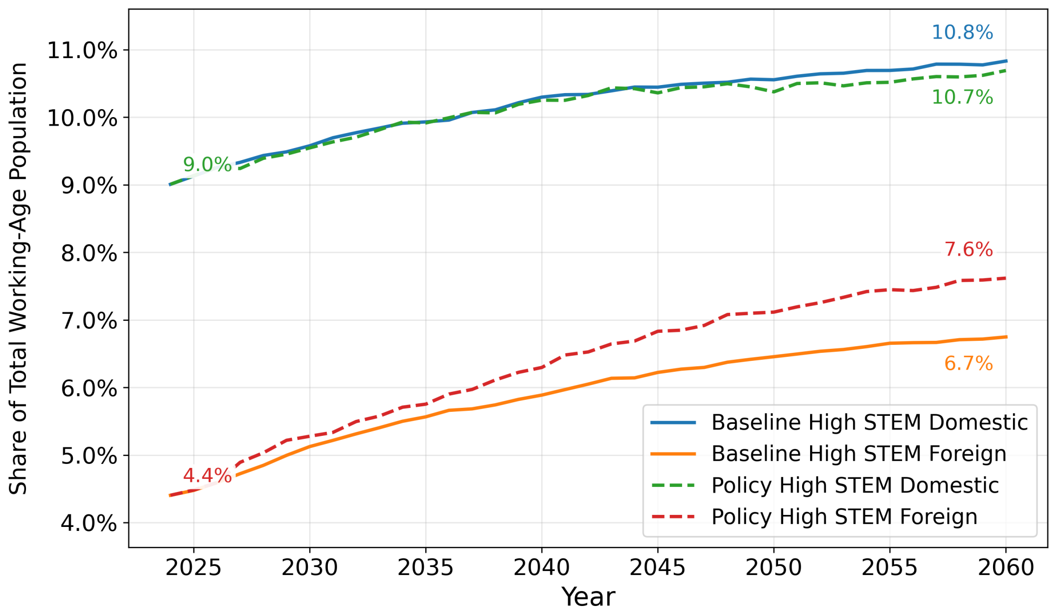


Figure 15: High-Educated STEM Working-Age Population in the Baseline and Green Card Policy

Notes: The working-age population is defined as individuals aged 21–65.

Appendix References

Basu, Susanto, John G. Fernald, and Miles S. Kimball. 2006. "Are Technology Improvements Contractionary?" [In en]. *American Economic Review* 96, no. 5 (December): 1418–1448. https://doi.org/10.1257/aer.96.5.1418.

Bound, John, Gaurav Khanna, and Nicolas Morales. 2017. "Understanding the Economic Impact of the H-1B Program on the United States." In *High-Skilled Migration to the United States and Its Economic Consequences,* 109–175. University of Chicago Press, May.

Caiumi, Alessandro, and Giovanni Peri. 2024. *Immigration's Effect on US Wages and Employment Redux.* Working Paper, April. https://doi.org/10.3386/w32389.

Comin, Diego, Javiar Quintana, Tom Schmitz, and Antonella Trigari. 2025. "Revisiting Productivity Dynamics in Europe: A New Measure of Utilization-Adjusted TFP Growth-Supplementary Material." Comin.etal:2025, *Journal of the European Economic Association* 23, no. 4 (April): 1598–1633.

Fernald, John. 2025. *TFP Data.* https://www.johnfernald.net/TFP, February.

Gunadi, Christian. 2019. "An inquiry on the impact of highly-skilled STEM immigration on the U.S. economy." *Labour Economics* 61 (December): 101751. https://doi.org/10.1016/j.labeco.2019.101751.

Labor Statistics, U.S. Bureau of. *Private Business Sector: Capital Intensity.*

Moretti, Enrico. 2004. "Chapter 51 - Human Capital Externalities in Cities." In *Handbook of Regional and Urban Economics,* 4:2243–2291. https://doi.org/https://doi.org/10.1016/S1574-0080(04)80008-7.

Ottaviano, Gianmarco I. P., and Giovanni Peri. 2012. "Rethinking the Effect of Immigration on Wages." *Journal of the European Economic Association* 10, no. 1 (February): 152–197. https://doi.org/10.1111/j.1542-4774.2011.01052.x.

Peri, Giovanni, Kevin Shih, and Chad Sparber. 2015. "STEM Workers, H-1B Visas, and Productivity in US Cities" [in en]. *Journal of Labor Economics* 33 (S1): S225–S255. https://doi.org/10.1086/679061.

Peri, Giovanni, Kevin Y. Shih, and Chad Sparber. 2014. *Foreign STEM Workers and Native Wages and Employment in U.S. Cities.* Working Paper, May. https://doi.org/10.3386/w20093.

Sargent, Thomas, J. 1978. "Estimation of Dynamic Labor Demand Schedules under Rational Expectations." *Journal of Political Economy* 86, no. 6 (December): 1009–1044.

Online Appendix

A The Model Economy

The model economy consists of a large number of heterogeneous overlapping-generations households, a government that can credibly commit to a fiscal policy, and a representative firm with constant-returns-to-scale technology.[12] Households face uninsurable idiosyncratic income and health risk, and partially insurable health expenditure risk. They maximize welfare by optimally choosing non-health consumption and leisure, whether to enroll in health insurance, and whether to invest in or forgo medical care.

The government finances outlays for Social Security, Medicare, Medicaid, the ACA, and SNAP through income taxes, payroll taxes, corporate income taxes, and estate taxes. Within this framework, a representative health insurance company operates competitively, generating zero profits. We evaluate the general equilibrium and distributional impacts of policy reforms by simulating transition paths between steady states.

Time is discrete and one model period is one year, which is denoted by t.

A.1 Households

A.1.1 Demographics

The economy is populated with households that are ex-ante heterogeneous with respect to the age of the head of household, $j \in J = \{J_B, \ldots, J_D\}$ where J_B is the age of "birth" and J_D is the final year of life in the model; beginning-of-period wealth, $a \in A = [a_{min}, a_{\max}]$; average historical earnings used to calculate Social Security Old-Age and Survivors Insurance (OASI) benefits, $b \in B = [0, b_{\max}]$; labor productivity, $z \in Z$; health status, $h \in H = [1, h_{max}]$; current heath needs, $d \in D = [1, d_{max}]$ and health insurance status $i \in I$.

12. The basic heterogeneous-agent OLG model described in this section is similar to that in Reichling and Nishiyama (2015). The detailed computational procedure to solve those models is described in Nishiyama and Smetters (2014).

An age-j household consists of $\Omega_{a,t}(j)$ adults and $\Omega_{c,t}(j)$ children, and its health status h is measured by the health of the head of household. The probability that an age-j head of household transitions from current health state h to next period's health state h' is given by the Markov transition probability $\pi_h[h'|j, h, d, \iota]$, where ι captures the choice of health expenditures described below. The time-varying period survival probability $\delta_t(j, h)$ depends on current age j and health state h. Household retire and enroll in Medicare at age $j = J_{R,t} = J_{Mcare,t}$.

The individual state space is defined as the Cartesian product of the sets of idiosyncratic characteristics, $S = J \times A \times B \times Z \times H \times D \times I$. A household's specific state at any point in time is summarized by the vector $s \in S$.

A.1.2 Aggregate Variables and Government Policy

Let ψ_t be the government's policy schedule at the beginning of year t,

$$\psi_t = \{c_{G,t}, tr_{OASI,t}(\cdot), Med_{Mcare,t}, \pi^{PremSub}_{ACA,t}(\cdot), Med_{Mcaid,t}, \tag{A.1}$$

$$\tau_{I,t}(\cdot), \tau_{P,t}(\cdot), \tau_{C,t}, \tau_{state,t}, \tau_{E,t}(\cdot), \{\tau^{statutory}_t\}\} \tag{A.2}$$

where $c_{G,t}$ is the government's consumption per household, $tr_{OASI,t}(\cdot)$ is an OASI benefit function, $Med_{Mcare,t}$ is the government Medicare benefits function, $Med_{Mcaid,t}$ is the government Medicaid benefit function, $\tau_{I,t}(\cdot)$ is a income tax function, $\tau_{P,t}(\cdot)$ is a OASI/HI payroll tax function, $\tau_{C,t}$ is a tax function on an exogenously determined share of non-health consumption, $\tau_{state,t}(\cdot)$ is a flat state income tax rate, $\tau_{E,t}(\cdot)$ is an estate tax on bequests net of costs, and $\{\tau^{statutory}_t\}$ are a set of taxes paid by the firm described in section A.4. Government consumption, c_G, does not affect a household's well-being, and the proceeds from the consumption tax, $\tau_{C,t}$, are assumed to approximate federal tax revenues other than those from the income, payroll taxes, and estate taxes, such as revenues from consumption or VAT taxes, excise taxes, and customs duties.

Let $p_t \in P$ be a vector of all prices in the economy at time t and $p^T_t := \{p_i\}^T_{i=t}$ be the sequence of prices from t to T. Let $\psi^T_t := \{\psi_i\}^T_{i=t}$ be a sequence of government policies from t to T. Let

X_t be the distribution of households at time t as defined in A.2. Let $\Psi_t^T = \{p_i, \psi_i, X_i\}_{i=t}^T$.

A.1.3 Utility

Households receive current utility $u(c, l)$ from a numeraire consumption good $c \in C$ and leisure $l \in [0, \Omega_{a,t}(j)l_{max}] \subset L$, where the maximum amount of available leisure time, $\Omega_{a,t}(j)l_{max}$ depends on the number of adults, $\Omega_{a,t}(j)$ in the household. Utility is defined as

$$u(c, l) = \frac{\left[c^{\alpha} l^{1-\alpha}\right]^{1-\gamma}}{1-\gamma}. \tag{A.3}$$

A.1.4 Labor Income

Households have an idiosyncratic Markov ability process z with transition probabilities $\pi_z[z'|z]$ and initial distribution $\pi_z^0(z)$. Household labor productivity is a function $e : J \times H \times Z \to E := [1, e_{\max}]$.

Working households aged $j < J_R$ receive gross wages w_t per working adult so that their total gross income, $y_t^G(j, h, z, l; w_t) = w_t e(j, h, z)(\Omega_{a,t}(j)l_{\max} - l)$.

Retired households receive OASI benefits from the government based on their lifetime average earnings, b, and age: $tr_{OASI,t} : J \times B \to \mathbb{R}^+$.

A.1.5 Health

Medical Expenditures Households face medical expenditures $m_t(j, h, d, i; X)$ that are determined by their age j, the health status of the head of household h, an idiosyncratic expenditure shock $d \in D$, their insurance type i and the distribution of households X. The shock d is i.i.d. and follows a density π_d. We assume that these expenditures can be decomposed as $m_t(j, h, d, i; X) = \nu_t^m(i)\tilde{m}_t(j, h, d; X)$, where $\nu_t^m(i)$ represents the potentially time-varying health care inflation—which varies by insurance type i—and $\tilde{m}_t$ is a time-invariant function of the household's health-relevant state variables and X.

A household's out-of-pocket expenditures $oop_t(m,i)$ are a fraction $\gamma^{oop}(m,i)$ of medical expenditures $m_t(j,h,d,i;X)$. The fraction $\gamma^{oop}(m,i)$ depends on the amount of total medical expenses m that the household faces and the insurance type i the household is enrolled in. If a household chooses not to receive medical treatment (ι=0), total out-of-pocket expenditures are zero. Otherwise, $oop_t(m,i) = \gamma^{oop}(m,i)m(j,h,d,i,X)$.

The transition probability for health depends on age, current health status, health expenditure choice and idiosyncratic health state: $\pi_h(h'|j,h,d,\iota)$.

Health Insurance Enrollment in public insurance is determined by eligibility: households who qualify for Medicaid or Medicare are automatically enrolled. All other households must decide at time t whether to purchase private health insurance for coverage in $t+1$, prior to the realization of their future health state. For those who opt into coverage, premiums are paid in $t+1$.

There are two ways to qualify for Medicaid: Households with sufficiently low income and wealth qualify categorically; those with sufficiently low income and medical expenses that are greater than a household's available resources qualify medically. There are no insurance premiums associated with Medicaid.

Retirees are automatically enrolled in Medicare Parts A, B, and D. We calculate Medicare premiums for Parts B and D following current law and assume that those grow at the same rate as the economy. Both of those premiums are income-based, so that higher income earners pay higher premiums. We denote an age-j household's Medicare premium by $\pi_{Mcare,t}(y,j)$, which depends on income y and the number of adults per household as determined by the age of the head of household.

There is a single competitive private health insurance market, and working-age households may receive subsidies from their employer or the government. These subsidies depend on the household's realized state in $t+1$, including their labor income and resulting leisure choice. Total per-household private insurance premiums are defined as $\Omega_{a,t}(j)p_{p,t}(j)$, where $p_{p,t}(j)$ is the per-person premium for an age-j head of household.

Households with income exceeding specific income thresholds qualify for Employer-Sponsored Insurance (ESI). ESI premiums are fully tax-deductible and subsidized by the firm at a rate of η^{ESI}. The household's portion of the premium is $(1-\eta^{ESI})\Omega_{a,t}(j)p_{p,t}(j)$. Choosing not to enroll in ESI when eligible results in the forfeiture of both the firm subsidy and the associated tax benefits.

In our framework, total compensation consist of two components: cash wages and health insurance subsidies. Paid labor income, $y_t^P(j,h,z,l;w_t)$, equals gross labor income less the ESI employer premium subsidy η^{ESI}, which we describe below. ESI premiums are fully tax deductible in the United States, so that taxable labor income $y_t^T(j,h,z,l;w_t)$ is given by gross labor income less the total cost of the ESI premium $\Omega_{a,t}(j)\,p_{p,t}(j)$:

$$y_t^P(j,h,z,l;w_t) = y_t^G(j,h,z,l;w_t) - \mathbf{1}_{[(i,y,j)=ESI]}\,\eta^{ESI}\,\Omega_{a,t}(j)\,p_{p,t}(j) \tag{A.4}$$

$$y_t^T(j,h,z,l;w_t) = y_t^G(j,h,z,l;w_t) - \mathbf{1}_{[(i,y,j)=ESI]}\,\Omega_{a,t}(j)\,p_{p,t}(j) \tag{A.5}$$

Working-age household who do not qualify for ESI may instead receive subsidies through the ACA marketplace. Households may qualify for premium subsidies from the government based on their family size, assets, and income. As under current law, private insurance premiums not purchased under ESI are not tax deductible. We denote the ACA premium subsidies, $\pi_{ACA,t}^{PremSub}: J \times Y^T \times A \rightarrow \mathbb{R}^+$.

Households may decide to remain uninsured, in which case they do not have to pay any insurance premiums. The uninsured pay their full medical expenditure out-of-pocket if they choose to incur the expenses.

Whether enrolled in health insurance or not, households may choose to forgo medical treatment altogether. While forgoing medical treatment saves on out-of-pocket expenditures, it also has downsides: it increases the probability of moving to a worse health state with higher medical expenses, greater mortality probability, and lower labor productivity.

A.1.6 Assets

Households can save one-period, risk-free shares in a mutual-fund. The mutual fund, described below, pools all household savings and then invests in domestic government debt and in domestic firms. Each of these investments may earn dividends and capital gains, potentially. As described below, the model has both corporate and pass-through firms and their respective dividends may be treated differently for tax purposes. So the household must keep track of the each of the components of the return.

The total return the household receives

$$r_t = r_t^{corp} + r_t^{pass} + r_t^G + r_t^C \tag{A.6}$$

where $r_t^{corp}, r_t^{pass}, r_t^G, r_t^C$ is the dividend on corporate capital, dividend on pass-through capital, interest payment on one-period government debt and capital gains on from changes in the value of the capital stock, respectively.

A.1.7 Bequests and Inheritances

Deceased households leave accidental bequests, which are distributed among the currently alive. We assume that a household that dies at the end of the period with planned savings a' enjoys the same continuation value as if it lived up to a scale parameter ζ:

$$\beta\zeta\mathbb{E}\big[\,V(s', \Psi_{t+1}^{t+J_D-j-1})$$

where V, Γ_h are defined below. Dying after age J_D gives an expected value of 0.

The bequest itself incurs some bequest transactions costs, $\Xi(a')$. After those costs, the government collects the rest as a bequest, taxes those bequests with an estate tax and puts the remainder into an inheritance pool, Q_t, to be distributed to the surviving population next period (with interest):

$$Q_t = \int_s (1-\delta_t(j,h))[\tilde{a}'(s,\Psi_t^{t+J_D-j}) - \tau_{E,t}(\tilde{a}'(s,\Psi_t^{t+J_D-j}))]\,dX_t(\mathbf{s}). \tag{A.7}$$

where $\tilde{a}'(s,\Psi_t^{t+J_D-j}) = a'(s,\Psi_t^{t+J_D-j}) - \Xi(a'(s,\Psi_t^{t+J_D-j}))$, $\tau_{E,t} : A \to \mathbb{R}^+$ is a tax on inheritances and $\Xi : A \to \mathbb{R}^+$ are transaction costs as described in Feiveson and Sabelhaus 2018a. The transaction costs consist of a fixed deduction, capped at the value of the estate to cover basic expenses such as burial, and a share of the residual estate to cover things such as charitable contributions, estate costs, and other end-of-life costs.

In each period, the government fully distributes the pool Q_t to living households according to a function $q_t : J \times Z \to \mathbb{R}^+$.[13]

For simplicity, we assume that the government collects wealth left by deceased households at the end of year t and then distributes it to households in the following year (with interest) but collects any estate taxes in the year t. Because there are no aggregate shocks in the model economy, the government can perfectly predict the sum of accidental bequests (at the end of the year).

The inheritance received by an age-j household is

$$q_t(j,e) = \omega^q(j,e)\left(\int_s \omega^q(j,e)dX_t(\mathbf{s})\right)^{-1} Q_t, \tag{A.8}$$

where $\omega^q(j,e)$ are parameters.

13. While in other settings the way in which bequests are distributed may not matter much, we do not simply distribute bequests lump sum because we want to ensure that households do not receive an unrealistically high minimum consumption floor that would provide insurance against health and health expenditure shocks. Instead we distribute total bequests by age, based on estimates from the Survey of Consumer Finances.

A.1.8 Household Problem

Let $V(s, \Psi_t^{t+J_D-j})$ be the value function of a household at the beginning of age j and year t. The household solves

$$V(s, \Psi_t^{t+J_D-j}) = \max_{c,l,a',i',\iota} \Big\{ u(c,l) + \beta \mathbb{E}\Big[\Big(\delta_t(j,h) + (1-\delta_t(j,h))\zeta \Big) V(s', \Psi_{t+1}^{t+J_D-j-1}) | s, \Gamma_h \Big] \Big\} \tag{A.9}$$

subject to the constraints for the decision variables,

$$c \geq \underline{c}, \qquad 0 \leq l \leq \Omega_{a,t}(j)\, l_{\max}, \qquad a' \geq a'_{\min}(s), \qquad i' \in I, \qquad \iota \in [0,1]$$

and the law of motion of the individual state,

$$s' = (j+1, a', b', z', h', d', i'),$$

$$\begin{aligned} a' = {} & y_t^P + (1+r_t)a + tr_{OASI,t}(j, b, p_{t-j}^t) - \tau_{I,t}(y_t^T, \tilde{r}_t a, tr_{OASI,t}(j, b, p_{t-j}^t)) \\ & - \tau_{P,t}(y_t^T) - \tau_{state}(y_t^T) - \tau_{C,t}(\psi_t^C c) - (1 - \mathbf{1}_{[(i,y,j)=ESI]}\eta^{ESI})\, \Omega_{a,t}(j) p_{i,t} \\ & + \mathbf{1}_{[(i,y,j,a)=ACA]} \pi_{ACA,t}^{PremSub}(y_t^T) - \iota \cdot oop_t(m(j,h,d,i,X), i) + SN_t(s,l) + SSI_t(s,l) \end{aligned} \tag{A.10}$$

$$b' = \mathbf{1}_{\{j<J_R\}} \frac{1}{j-20} \Big[(j-21)\, b \frac{w_t}{w_{t-1}} + \min(\frac{y_t^T}{\Omega_{a,t}(j)}, \vartheta_{\max}) \Big] + \mathbf{1}_{\{j \geq J_R\}} b, \tag{A.11}$$

where the vector $\tilde{r}_t = (r_t^{corp}, r_t^{pass}, r_t^G, r_t^C)$; $\mathbf{1}_{\{\cdot\}}$ is an indicator function that returns 1 if the condition in { } holds and 0 otherwise; and $\vartheta_{\max}$ is the maximum taxable earnings for OASI taxes. The average historical earnings for an individual at the beginning of the following year, b', are calculated recursively as a weighted average of its current (beginning-of-year) average historical earnings, b, adjusted by the growth in the wage, w_t/w_{t-1} (that is, wage-indexed), and of the current OASDI taxable earnings, $\min\{\frac{y_t^T}{\Omega_{a,t}(j)}, \vartheta_{\max}\}$.[14] Nonetheless, we do not model the change in indexing after age 60 in the dynamic OLG model. Note that we deflate the earnings by $\Omega_{a,t}(j)$ in the calculation of AIME so that the earnings history reflects individual earnings. In the subsequent subsection B.9.1, we show how benefits are computed based on individual earnings and then rescale the benefits using the average number of adults to approximate total expected household benefits. SN_t and SSI_t are the government transfer programs, SNAP and SSI, respectively. We gather the household's decision rules and average historical earnings in the next year as $\Gamma_h(s, \Psi_t^{t+J_D-j}) \equiv \{c, l, a', i', \iota, b'\}(s, \Psi_t^{t+J_D-j})$.

A.2 The Distribution of Households

We define $X_t(s)$ as the measure of households at state s at time t, with x_t representing the associated density.

A.2.1 Entry and Initial Distribution

Households enter the economy either at "birth" or through immigration. There are time-varying, exogenous entry flows of households, nb_t, at "birth" (age $j = J_B$) and from immigration at any age, $im_t(j)$. There is a time-varying, exogenous exit flow of households due to emigration at any age, $em_t(j)$. All entrants are assumed to start with no assets ($a = 0$) and no work history ($b = 0$).

Emigration is assumed to occur at the end of period t, while immigration takes place at the beginning of period t. This timing convention follows the Microsim framework, which we replicate in our OLG model to ensure consistency in the evolution of the population.

Households entering through birth We first describe the distribution of native households entering the economy at age $j = J_B \equiv 21$. These agents draw their initial health status h, health

14. Average historical earnings are assumed to remain constant after retirement age. In the actual calculation of AIME, a worker's annual taxable earnings are indexed to reflect the general earnings level in the indexing year, which is age 60 for most workers, the worker's earnings in years after the indexing year are not indexed. See Social Security Administration (2018) for more information.

expenditure state d and labor ability z from the distributions π_h^0, π_d and π_z^0 respectively. We set the initial distribution over health insurance status (i) at entry according to the following algorithm:

Step 1 : For a given Ψ_t^T, we compute the household value and policy functions in A.1.8.

Step 2: For each $(z,h) \in Z \times H$, compute the expected values of having insurance and not having insurance:

$$EV_{0,t}(z,h) = \sum_{d \in D} \pi_d(d) V_t(J_B, 0, 0, z, h, d, 0) \tag{A.12}$$

$$EV_{1,t}(z,h) = \sum_{d \in D} \pi_d(d) V_t(J_B, 0, 0, z, h, d, 1) \tag{A.13}$$

Step 3: Define an indicator function $1_i(z,h) = 1$ if $EV_{1,t}(z,h) > EV_{0,t}(z,h)$ and 0 otherwise.

Define a density $\pi_{NB,t}$ by

$$\pi_{NB,t}(a,b,z,h,d,i) = \begin{cases} \pi_z^0(z)\pi_d(d)\pi_h^0(h), & \text{if } a=0, b=0 \text{ and } i = 1_i(z,h) \\ 0, & \text{otherwise.} \end{cases} \tag{A.14}$$

Note $\pi_{NB,t}$ is time-invariant if we integrate over all insurance choices. In other words,

$$\bar{\pi}_{NB}(a,b,z,h,d) \equiv \sum_{i \in \{0,1\}} \pi_{NB,t}(a,b,z,h,d,i) \tag{A.15}$$

is time-invariant.

Households entering through immigration Similarly to households entering through birth, immigrant households are assumed to enter the economy with no assets ($a = 0$) and no work history ($b = 0$). However, immigrants may enter at any age $j \in J$. Let Π^{immi} denote the distribution of immigrant households over the state space upon arrival, with π^{immi} representing the associated density.

For immigrants entering at age $j = J_B$, we assume that their distribution over individual characteristics coincides with the distribution of native-born entrants described in equation (A.15) and split the immigrants in equal proportion over insurance states:

$$\pi^{\text{immi}}(J_B, a, b, z, h, d, i) = 0.5\bar{\pi}_{NB}(a, b, z, h, d) \qquad \forall\, i \in \{0, 1\}$$

Using this as the starting point, we assume that π^{immi} evolves according to the following law of motion:

$$\pi^{\text{immi}}(j{+}1, a, b, z', h', d, i) = \begin{cases} \delta_t(j, h)\pi^{\text{immi}}(j, 0, 0, z, h, d, i)\pi_z(z'|z)\pi_h(h'|j, h, d, 1), & \text{if } a = 0,\ b = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Notice that the distribution of entering migrants is independent of period t, conditional upon entry. However, the measure of immigrants entering the economy may vary over time (see A.2.3 for details), an important source of variation in the model.

A.2.2 Survival Probabilities

We now describe how survival probabilities are computed for use in the household dynamics. The Microsim dataset provides forecasts of survival probabilities by age j and year t, denoted $\delta_{j,t}^{\text{microsim}}$. In our model, we assume that survival probabilities also depend on the household health state h.

To account for this additional heterogeneity, we use data from the MEPS dataset and compute empirical survival probabilities by age j and health state h, denoted $\delta_{j,h}^{\text{MEPS}}$.

To construct the survival probabilities used in the model, conditional on age j, health state h, and year t, while endogenously matching the survival rates, $\delta_{j,t}^{\text{microsim}}$, we solve for $\varphi_{j,t}$ using the following equation:

$$\sum_h \omega_{j,h,t} \left(1 - \delta_{j,h}^{\text{MEPS}}\right)^{\varphi_{j,t}} = 1 - \delta_{j,t}^{\text{microsim}}, \tag{A.16}$$

for every age j and year t. Intuitively, equation (A.16) represents a weighted average of death rates across health states that matches the aggregate death rate obtained from the Microsim dataset.

The weights $\omega_{j,h,t}$ are based on the population distribution over health states:

$$\omega_{j,h,t} \equiv \frac{\tilde{x}_t(j,h)}{\hat{x}_t(j)},$$

where $\tilde{x}_t(j,h) = \int_{A\times B\times Z\times D,\times I} x(s)ds$ and $\hat{x}_t(j) = \sum_h \tilde{x}_t(j,h)$. These weights ensure that more prevalent health states contribute proportionally to the aggregate survival probability.

Finally, we define the model survival rates as

$$\delta_t(j,h) = \left(\delta_{j,h}^{\text{MEPS}}\right)^{\varphi_{j,t}}. \tag{A.17}$$

Since $\omega_{j,h,t}$ depends on the distribution of households, X, which in turn depends on the choices of households and thus their survival rates, the set of $\omega_{j,h,t}$ are implicitly part of the fixed point equilibrium calculation.

A.2.3 Law of Motion of Household Distribution

In this section, we describe the evolution of the household distribution over the state space S. Before formally presenting the full law of motion, it is helpful to discuss how emigration and immigration affect the evolution of households.

Survival and Emigration Rates The distribution of households in period t is given by $x_t(s)$, and the age-specific emigration measures $em_t(j)$ is exogenous. From these, we can compute the **age-specific emigration rate** for each age j as

$$em_t^{\text{rate}}(j) = \frac{em_t(j)}{\hat{x}_t(j)},$$

where $\hat{x}_t(j)$ is the total population measure conditional on age j in period t.

We treat emigration as a flat rate applied uniformly across the distribution of each cohort. In particular, we can combine the survival probabilities computed in equation (A.17) with the emigration rate $em_t^{\text{rate}}(j)$ to obtain the **adjusted survival probability net of emigration**:

$$\tilde{\delta}_t(j,h) = \delta_t(j,h) - em_t^{\text{rate}}(j). \tag{A.18}$$

Immigration Immigration also affects the distribution of households in each period. We combine the distribution of new entrants from immigration, $\pi^{\text{immi}}(s)$, described in section A.2.1, with the age and period specific immigration measure $im_t(j)$: $\pi^{\text{immi}}(s) \cdot im_t(j)$.

A.2.4 Scaling Vector and the law of motion of households

Our OLG model captures substantial heterogeneity; however, it does not fully replicate all of the richness present in the Microsim dataset. For instance, we do not explicitly model divorce and therefore cannot account for household formation and destruction arising from divorce. To reconcile the model-implied population distribution with the observed data, we iteratively introduce a scaling adjustment to the population distribution, $\pi_{t+1}^s(j)$.

Let $\hat{x}_t^{\text{microsim}}(j)$ denote the total population measure conditional on age j in period t observed in the Microsim dataset. By construction we can set the population entering the model at age $j = J_B$ to $\hat{x}_t^{\text{microsim}}(J_B)$. Call this population $\hat{x}_{t+1}^{\text{forecast}}(J_B)$. Absent rescaling, the model population would evolve iteratively:

$$\begin{aligned}\hat{x}_{t+1}^{\text{forecast}}(j') = \int_S \Big[& \mathbf{1}_{\{a'=a'(),b'=b'(),h'=h'(),i'=i'()\}} \\ & 1_{j=j(s)} \tilde{\delta}_t(j,h)\, \pi_z(z'|\, z)\; \pi_h(h'|j,\; h, d, \iota())\pi_d(d') \Big]\, dX_t(s) \\ & + \pi^{\text{immi}}(j', a', b', z', h', d', i') im_{t+1}(j')\end{aligned} \tag{A.19}$$

where $\mathbf{1}_{\{a'=a'(),b'=b'(),h'=h'(),i'=i'()\}}$ is the indicator function given by the optimal household policy $\Gamma_h(s, \Psi_t^{t+J_D-j})$. Note by construction from (A.16)-(A.18), $\hat{x}_{t+1}^{\text{forecast}}(j')$ can also be computed directly from Microsim survival, immigration and emigration rates. We then define a *scaling vector* in period $t+1$, $\pi_{t+1}^s : J \to \mathbb{R}^{++}$, given by

$$\pi^s_{t+1}(j) = \frac{\hat{x}^{\text{microsim}}_{t+1}(j)}{\hat{x}^{\text{forecast}}_{t+1}(j)}. \tag{A.20}$$

The model density of households then evolves according to: The distribution of households across states s′ at age $j+1$ in year $t+1$ depends on the population distribution over s at age j in year t as well as on households' decisions that influence assets, earnings, health status, and insurance enrollment as follows, for $j = J_B, \ldots, J_D$,

$$x_{t+1}(s') = \pi^s_{j,t}(j') \Bigg[\int_S \mathbf{1}_{\{a'=a'(),b'=b'(),h'=h'(),i'=i'()\}} \tilde{\delta}_t(j,h)\, \pi_z(z'|\, z)\ \pi_h(h'|j,\, h, d, \iota())\pi_d(d')\, dX_t(s) + \pi^{\text{immi}}(j', a', b', z', h', d', i') im_{t+1}(j') \Bigg] \tag{A.21}$$

A.3 Health Insurers

Insurers in the private insurance market are not allowed to condition policies on observables. However we do allow for age-based pricing. The insurance market is competitive, so that the zero profit condition for individual insurance premiums are the sum of reimbursable expenses and administrative costs, divided by those who enroll in private insurance. The equation defining the price of the private insurance is:

$$p_{p,t}(j, \Psi_t^{t+J_D-j}) = \frac{\omega^{HI}_j}{\sum_j \omega^{HI}_j\, \Omega_{a,t}(j)\, \Pi_t(j)(i = private)} \times \frac{\int_S \mathbf{1}_{[i=private,\iota=1]} \Big(m_t(j,h,d,i,X_t) - oop_t(m,i) \Big)\ dX_t(s)}{(1-\kappa_{private})} \tag{A.22}$$

where ι denotes $\iota(s, \Psi_t^{t+J_D-j})$. The weights ω^{HI}_j are age-specific and set such that insurance enrollment closely matches the data. $\Pi_t(j)(i = private)$ is the measure of households of age j at time t enrolled in private health insurance. Administrative expenses associated with providing health insurance are proportional to paid benefits and differ by insurance type. They are denoted κ_i.

A.4 Firms

There is a representative final goods firm, and a set of intermediate good firms, each representative within their sector. First, we describe the zero-profit final goods firm, which takes the goods from n intermediate goods sectors and combines them into a single output using a Cobb-Douglas production technology. Then we describe the n sectors of intermediate good production, which are produced by both a pass-through entity and a corporation.

A.4.1 Final Goods Firm

The final good—which can be directly consumed by the households or converted into an equal number of new investment goods—is produced competitively by combining the output of the n sectors using a Cobb-Douglas production technology:

$$Y_t = \prod_j Y_{j,t}^{\zeta_j} \;\; s.t. \;\; \sum_j \zeta_j = 1$$

where $Y_{j,t}^{\zeta_j}$ is the total production across both business types in sector j at time t. Normalize the price of the final good to one and let $p_{j,t}$ be the price of good $Y_{j,t}$ $\forall j \in \{1, \ldots, n\}$. Cost minimization implies:

$$p_{j,t} = \zeta_j \frac{Y_t}{Y_{j,t}} \tag{A.23}$$

A.4.2 The Intermediate Good Firm

Business Type Within each sector of the economy, businesses can be either (a) corporations or (b) pass-through entities. For each sector $j \in \{1, \ldots, n\}$ and each business type $b \in \{corp, pass\}$, the production technology is:

$$Y_{b,j,t} = A_{j,t} K_{b,j,t}^{\alpha^f_{j,t}} L_{b,j,t}^{1-\alpha^f_{j,t}}$$

Due to constant returns to scale, splitting the "firm" within each sector does not change total production. Note that the TFP ($A_{j,t}$) and the income share to capital ($\alpha^f_{j,t}$) are sector-specific, but not specific to the type of business. The share for business type b is parameter $\zeta^{income}_{b,j,t}$ and is used to divide up capital and labor such that.

$$K_{b,j,t} = \zeta^{income}_{b,j,t} K_j, \quad L_{b,j,t} = \zeta^{income}_{b,j,t} L_{j,t} \tag{A.24}$$

where $\sum_b \zeta^{income}_{b,j,t} = 1$, $K_{j,t}$ is total sector capital, and $L_{j,t}$ is total sector labor. It follows that output, $Y_{b,j,t} = \zeta^{income}_{b,j,t} Y_{j,t}$.

We assume that the corporation has an optimization problem, described in the next subsection and that pass-through firms' capital and labor are determined by the relationships in Equation A.24. Note that the pass-through may have different marginal products of capital and labor than the corporations, and the pass-through pays the same wage rate as the corporation.

Firm Depreciation Allowances When making the decision on how to invest, the user cost of capital uses the current net present value of the tax deductions. Nonetheless, the timing of these deductions matters for the rest of the model as government debt "crowds-out" productive investment in most specifications, and therefore the timing of government debt affects the transition path. Therefore, to accurately model the effect of depreciation schedules on the government budget constraint, each firm keeps track of a schedule of depreciation allowances going back T_f years. For a given firm of business type b and sector j, the allowance schedule for the next $T_f + 1$ years (typically 40 years, plus the current year).

We define $\phi^{exp}_{b,j,t,i}$, for $i = \{0, \ldots, T_f\}$ at time t and sector $j \in \{1, \ldots, n\}$, as the total value of expensing deductions in n years from t for all historical investments up to period t. For example, $\phi^{exp}_{b,j,t,0}$ is the deduction the corporate firm j takes in period t for all investments made in years up to and *including* year t, whereas $\phi^{exp}_{b,j,t+1,1}$ is the deduction this firm will take in $t+1$ for all investments made in years up to and *including* year t but *excluding* the investment made in year $t+1$.

We define every value of $\phi^{exp}_{b,j,t,i}$ for $i = \{0, \dots, T_f\}$ as follows:

$$\phi^{exp}_{b,j,t,i} = \phi^{exp}_{b,j,t-1,i+1} + \phi_{b,j,t,i} q_t [K_{b,j,t} - (1-\delta_{k,j}) K_{b,j,t-1}] \frac{pgdp_t}{pgdp_{t+i}} \tag{A.25}$$

where $\delta_{k,j}$ is the physical depreciation rate of capital used in sector j, $\phi^{exp}_{b,j,t,T_f+1} = 0$ for all j and t, $\phi^{exp}_{b,j,t,i}$ is the deduction in year $t+i$ for an investment made in year t as defined earlier, and $\phi_{b,j,t,i}$ is the parameter for the portion of investment made i periods ago that should be expensed in period t for a corporate firm in sector j. Each value is carried over from the previous year, except for the T_fth year as per the tax code.

In this case, $\phi_{b,j,t,0}$ is the amount of the expense deduction that the firm gets at time t, while $\phi_{b,j,t,1}$ would be the amount that the firm is allowed to deduct in period $t+1$ plus whatever amount is allowed to be expensed out of their investments in period $t+1$.

We treat investment credits as changes to $\phi_{b,j,t,0}$. In this case, the firm gets an extra deduction reflecting the value of the contemporaneous investment credit.

Business Income Subject to Business Taxation Business taxable income starts with the firm's gross revenue ($p_{j,t} Y_{b,j,t}$) minus the wage bill ($w_t L_{b,j,t}$). From that, we immediately deduct a share of the income for "other" deductions ($\zeta^{ded}_{b,j,t}$) from the firm's revenue net of the wage bill. In addition to those deductions, we allow the firm to deduct accrued investment allowances, $\phi^{exp}_{b,j,t,0}$. Lastly, we allow for an interest rate deduction for business debt $\phi^{int}_{b,j,t} \rho(K_{b,j,t})$, where the term $\rho(K_{b,j,t})$ is total cost of borrowing.

$$\Pi^{taxable}_{b,j,t} = (1 - \zeta^{ded}_{b,j,t}) \left[p_{j,t} Y_{b,j,t} - w_t L_{b,j,t}\right] - \phi^{exp}_{b,j,t,0} - \left(\frac{K_{b,j,t}}{K_t}\right) C(I_t, K_t) \tag{A.26}$$

$$- \phi^{int}_{b,t} \underbrace{\left[\rho_{b,j,t} B_{b,j,t} + \nu \left(\frac{B_{b,j,t}}{K_{b,j,t}}\right) q_t \nu_2 K_{b,j,t}\right]}_{\rho(K_{b,j,t})} \tag{A.27}$$

Note that $\zeta^{ded}_{pass,j,t}$ is typically significantly lower than $\zeta^{ded}_{corp,j,t}$, but neither value is zero.

The total business income tax liability is given by

$$T_{b,j,t} = \tau^{statutory}_{b,j,t}\Pi^{taxable}_{b,j,t} - \zeta^{cred}_{b,j,t} p_{j,t}Y_{b,j,t} - \psi_{b,j,t}I_{b,j,t} \tag{A.28}$$

where $\tau^{statutory}_{b,j,t}$ is the statutory business income tax rate, $\zeta^{cred}_{b,j,t}$ is the share of sector and business output which is a tax credit, and $\psi_{b,j,t}$ is the share of investment that can be claimed as a tax credit. In our baseline, $\tau^{statutory}_{pass,t} = 0$ and $\zeta^{cred}_{pass,j,t} = 0$ for all t. In general, the credit amount, driven by the tax share $\psi_{pass,j,t}$, is determined at the entity level and doesn't depend on any individual owners' characteristics. Then the credit amount gets apportioned and distributed to the owners along with the pass-through's income.

Business Dividends The net taxes the firm pays are given by:

$$\begin{aligned} T_{b,j,t} = {} & \tau^{statutory}_{b,j,t}(1-\zeta^{ded}_{b,j,t})\left[p_{j,t}F(K_{b,j,t}, L_{b,j,t}) - w_t L_{b,j,t}\right] - \zeta^{cred}_{b,j,t} p_{j,t}F(K_{b,j,t}, L_{b,j,t}) \\ & - \tau^{statutory}_{b,j,t}\phi^{int}_{b,t}\left[\rho_{b,j,t}B_{b,j,t} + \nu\left(\frac{B_{b,j,t}}{K_{b,j,t}}\right)q_t\nu_2 K_{b,j,t}\right] - \tau^{statutory}_{b,j,t}\phi^{exp}_{b,j,t,0} - \psi_{b,j,t}I_{b,j,t} \\ & - \tau^{statutory}_{b,j,t}\left(\frac{K_{b,j,t}}{K_t}\right)C(I_t, K_t) \end{aligned}$$

The actual money that gets passed back to the mutual fund is equal to:

$$\pi_{b,j,t}K_{b,j,t} = (1-\zeta^{other}_{b,j,t})p_{j,t}Y_{b,j,t} - w_t L_{b,j,t} - \delta_{k,j}K_{b,j,t} - T_{b,j,t} - \left(\frac{K_{b,j,t}}{K_t}\right)C(I_t, K_t) - \nu\left(\frac{B_{b,j,t}}{K_{b,j,t}}\right)q_t\nu_2 K_{b,j,t}$$

which are the total revenues minus the total wage bill, minus the investment, minus the tax bill, minus the fraction of capital adjustment costs, minus the leverage cost. Note that $\pi_{b,j,t}$ is defined as a rate.

It is useful to define the total amount earned by the mutual fund from each business type as:

$$R_{b,t} = \sum_{j=1}^{n} \pi_{b,j,t}K_{b,j,t} \tag{A.29}$$

Price of Capital Since old and new capital are perfect substitutes in production, their marginal acquisition costs must be identical in equilibrium. We compute the price of existing capital so that it the household is indifferent between buying a generic unit of existing capital and a unit of new capital. We assume the output good can be converted into new capital at a one-to-one rate, which makes the price equal to one ($q_t^{new} = 1$).

For new capital, the marginal acquisition cost is 1 unit of the final good less the tax subsidy from investment expensing, which we define as as $\phi^{npv}\tau$, where τ is the marginal tax rate on business income and ϕ^{npv} net present value of the deduction of a single unit of new capital. The price of capital is determined by tax policy alone, since we do not assume capital adjustment costs. Therefore, the price of existing capital must be discounted by the net present value of the tax subsidy.

Each firm in sector j and of business type b has its own a determined schedule of deductions for an investment made at time t. The schedule allows for a series of deductions in the year t, in which the investment is made, and the subsequent T_f years. We call the share of the investment that is deductible in each year $t+n$ as $\phi_{b,j,t,n}$ for $n \in \{0, \ldots, T_f\}$. The net present value of the entire T_f-year stream of deductions for an investment made in time t is defined as:

$$\phi_{b,j,t}^{npv} = \tilde{\tau}_{b,j,t}\phi_{b,j,t,0} + \sum_{n=1}^{T_f}\left(\prod_{i=1}^{n}\frac{1}{1+r_{t+i}^D}\right)\frac{pgdp_t}{pgdp_{t+n}}\tilde{\tau}_{b,j,t+n}\phi_{b,j,t,n} \tag{A.30}$$

where r_t^D is the interest rate on government bonds in period t and $pgdp_t$ is the GDP price index at time t. $\tilde{\tau}_{b,j,t}$ is the tax rate that applies to the income against which the deduction is being taken. For corporations, $\tilde{\tau}_{corp,j,t} = \tau_{b,j,t}^{statutory}$. Pass-through income accrues to all individuals who own assets in the model. For convenience in pricing capital though, we assume that $\tilde{\tau}_{pass,j,t} = \tau_{pit,t}^{top}$, the highest individual marginal tax rate. The fraction of the investment that is deducted in each year of the subsequent T_f years is deflated by both the interest and the inflation rates.

Because the purchaser of new capital in sector j and business type b receives an implicit $\Gamma_{b,j,t} = \psi_{b,j,t} + \phi_{b,j,t}^{npv}$ subsidy from the government while the purchaser of old capital receives no tax subsidy, old capital must also be discounted to make capital purchasers indifferent between old and new capital.[15]

In the case of corporate investment in sector j, the cost of corporate capital is $1 - \Gamma_{corp,j,t}$. However, because there are n sectors with potentially different tax systems and non-corporate capital also exists, we require a more complex formulation. First, there is no single marginal cost of pass-through capital since pass-through capital income enters into the personal income tax, which is a non-linear function. Then, it becomes necessary to reconcile different costs of corporate and pass-through capital and across sectors. To proceed, we are forced to make additional simplifying assumptions or face having to model a capital trade market. The division of capital between pass-through firms and corporations is fixed exogenously by the time-varying parameters $\zeta^{income}_{pass,j,t}$ and $\zeta^{income}_{corp,j,t}$. Under these assumptions, the cost of existing capital becomes:

$$q^{exist}_t = \sum_{j=1}^{n} \frac{K_{j,t}}{K_t} \left(1 - \zeta^{income}_{corp,j,t}\Gamma_{b,j,t} - \zeta^{income}_{pass,j,t}\Gamma_{pass,j,t}\right). \qquad \text{(A.31)}$$

Note that the price of new capital, $q^{new}_t = 1$ as a unit of new capital costs one unit of the final good. Note that the price of capital should not be confused with the user cost of capital, which is derived later in the corporate firm problem.

15. This implied market condition only holds so long as the investment is strictly positive. If the economy is dis-investing in aggregate, then the market will not clear at the price of new capital.

Corporate Sector Problem - with Interest Rate Deductibility We define the problem for a representative corporate firm in sector j and $b = corp$ (we use b for notational convenience):

$$\begin{aligned}
V_{b,j,t}(K_{b,j,t}, \Psi_t^{t+T_f-1}) = & \max_{K_{b,j,t+1}, L_{b,j,t}, B_{b,j,t}} p_{j,t} F(K_{b,j,t}, L_{b,j,t}) - w_t L_{b,j,t} \\
& - \tau_{b,j,t}^{statutory}(1-\zeta_{b,j,t}^{ded}) \left[p_{j,t} F(K_{b,j,t}, L_{b,j,t}) - w_t L_{b,t}\right] + \zeta_{b,j,t}^{cred} p_{j,t} F(K_{b,j,t}, L_{b,j,t}) \\
& + \tau_{b,j,t}^{statutory} \phi_{b,t}^{int} \left[\rho_{b,j,t} B_{b,j,t} + \nu\left(\frac{B_{b,j,t}}{K_{b,j,t}}\right) q_t \nu_2 K_{b,j,t}\right] - I_{b,j,t}(1-\Gamma_{b,j,t}) \\
& - \nu\left(\frac{B_{b,j,t}}{K_{b,j,t}}\right) q_t \nu_2 K_{b,j,t} + \bar{V}_{b,j,t} + \frac{1}{(1+r_t)} V_{b,j,t}(K_{b,j,t+1} \Psi_{t+1}^{t+T_f}) \\
& \text{s.t. } I_{b,j,t} = K_{b,j,t+1} - (1-\delta_{k,j}) K_{b,j,t}
\end{aligned}$$

where q_t is the price of existing capital; $\nu\left(\frac{B_{b,j,t}}{K_{b,j,t}}\right) = \frac{1}{\nu_1}\left(\frac{B_{b,j,t}}{q_t \nu_2 K_{b,j,t}}\right)^{\nu_1}$, where ν_1 is the leverage sensitivity and ν_2 is the scaling of the leverage factor. Note that the price of capital is q_t because that's the average market value of the capital that's collateral; and $\bar{V}_{b,j,t}$ is a sequence of payments taken exogenously[16] by the corporation at time t and do not affect the firm's optimization problem. We assume

$$\bar{V}_{b,j,t} = \tau_{b,j,t}^{statutory} \phi_{b,j,t-1,1}^{exp} - (1-\tau_{b,j,t}^{statutory}) \left(\frac{K_{b,j,t}}{K_t}\right) C(I_t, K_t),$$

where $\phi_{b,j,t-1,1}^{exp}$ corresponds to the total investment expensing allowances for all investments made up until $t-1$ and paid in period t (formally defined later), and $C(I_t, K_t)$ is the (tax-deductible) convex capital adjustment cost function for the economy whereby each firm pays a fraction of such cost proportional to its size in terms of capital. For the total cost of gross investment I_t, given a current aggregate capital stock K_t:

$$C(I_t, K_t) = \frac{\eta}{2}\left(\frac{I_t}{K_t}\right) I_t$$

$\psi_{b,j,t}$ is the initial investment credit.

16. This is by assumption.

The cost of investment is symmetric around zero because we assume that the depreciation schedule applies equally to investment and disinvestment. Firms make decisions treating existing capital the same regardless of whether it is being bought or sold. In the case of disinvestment, the firm may owe additional taxes in the future as these are treated as negative deductions. Let $\Gamma^f_{b,j} : K \times \gneqq^{T_f+1} \rightarrow \mathbb{R}^2_+$ be the optimal policy function for firms.

Wages and the FONC with Respect to Labor The first-order condition with respect to labor requires that the marginal product of labor is equal to the wage (note that we substitute in Equation (A.23) to get this relationship):

$$w_t = \Phi^W_{b,j,t} \zeta_j \frac{Y_t}{L_{j,t}} \;\; \forall j \tag{A.32}$$

where $\Phi^W_{b,j,t} = \frac{[1-\tau^{statutory}_{b,j,t}(1-\zeta^{ded}_{b,j,t})+\zeta^{cred}_{b,j,t}](1-\alpha^f_{j,t})}{1-\tau^{statutory}_{b,j,t}(1-\zeta^{ded}_{b,j,t})}$

The firm subsidizes health insurance premiums for employees that qualify, so that gross and paid labor income differs.

$$\begin{aligned} w_t \, L_t &= \sum_{j=21}^{J_R-1} \int_{A\times B\times E\times H\times I} y^G_t(j,h,z,l,w_t)\, dX_t(s), \\ &= \sum_{j=21}^{J_R-1} \int_{A\times B\times E\times H\times I} \Big(y^P_t(j,h,z,l,\iota,w_t) + \mathbf{1}_{[(i,y,j)=ESI]}\, \eta^{ESI}\, p_{p,t}(j) \Big)\, dX_t(s) \end{aligned}$$

Borrowing and FONC with Respect to Debt The first-order condition with respect to debt is:

$$\frac{V_{b,j,t}}{B_{b,j,t}} : 0 = \tau^{statutory}_{b,j,t} \phi^{int}_{b,j,t} \left[\rho_{b,j,t} + \frac{B^{\nu_1-1}_{b,j,t}}{(q_t \nu_2 K_{b,j,t})^{\nu_1}} q_t \nu_2 K_{b,j,t} \right] - \frac{B^{\nu_1-1}_{b,j,t}}{(q_t \nu_2 K_{b,j,t})^{\nu_1}} q_t \nu_2 K_{b,j,t}$$

Rearranging, we get a level of debt that is optimal for a specific level of firm capital.

$$B_{b,j,t} = \left(\frac{\tau^{statutory}_{b,j,t} \phi^{int}_{b,j,t} \rho_{b,j,t}}{1-\tau^{statutory}_{b,j,t} \phi^{int}_{b,j,t}} \right)^{\frac{1}{\nu_1-1}} q_t \nu_2 K_{b,j,t} \tag{A.33}$$

User Cost of Capital The first-order condition with respect to capital leads to the following relationship:

$$\frac{dV_{b,j,t}}{dK_{b,j,t+1}} : 0 = -\frac{(1-\Gamma_{b,j,t})}{(1+r)^{t-1}} + \frac{1}{(1+r)^{t}} \{ p_{j,t+1} F^{K}_{b,j,t+1} - \tau^{statutory}_{b,j,t+1}(1-\zeta^{ded}_{b,j,t+1}) p_{j,t+1} F^{K}_{b,j,t+1}$$
$$+ \zeta^{cred}_{b,j,t+1} p_{j,t+1} F^{K}_{b,j,t+1} + \left(\tau^{statutory}_{b,j,t+1}\phi^{int}_{b,j,t+1} - 1\right) \frac{1-\nu_1}{\nu_1} \left(\frac{B_{b,j,t+1}}{q_{t+1}\nu_2}\right)^{\nu_1} K^{-\nu_1}_{b,j,t+1} q_{t+1}\nu_2$$
$$+(1-\delta_{k,j})(1-\Gamma_{b,j,t+1}) \}$$

Rearranging the terms, we are able to get a discrete-time definition of the user cost of capital:

$$F^{K}_{b,j,t+1} = \frac{(1-\Gamma_{b,j,t})(1+r) + \left(q_{t+1}\nu_2 \frac{1-\nu_1}{\nu_1}\right) \frac{\left(\tau^{statutory}_{b,j,t+1}\phi^{int}_{b,j,t+1}\rho_{b,j,t+1}\right)^{\frac{\nu_1}{\nu_1-1}}}{\left(1-\tau^{statutory}_{b,j,t+1}\phi^{int}_{b,j,t+1}\right)^{\frac{1}{\nu_1}}} - (1-\delta_{k,j})(1-\Gamma_{b,j,t+1})}{p_{j,t+1}\left[1-\tau^{statutory}_{b,j,t+1}(1-\zeta^{ded}_{b,j,t+1}) + \zeta^{cred}_{b,j,t+1}\right]}$$
$$= \frac{\Phi^{R}_{b,j,t+1}}{p_{j,t+1}} \tag{A.34}$$

where $\Phi^{R}_{b,j,t+1} \equiv \frac{(1-\Gamma_{b,j,t})(1+r) + \left(q_{t+1}\nu_2 \frac{1-\nu_1}{\nu_1}\right) \frac{\left(\tau^{statutory}_{b,j,t+1}\phi^{int}_{b,j,t+1}\rho_{b,j,t+1}\right)^{\frac{\nu_1}{\nu_1-1}}}{\left(1-\tau^{statutory}_{b,j,t+1}\phi^{int}_{b,j,t+1}\right)^{\frac{1}{\nu_1}}} - (1-\delta_{k,j})(1-\Gamma_{b,j,t+1})}{\left[1-\tau^{statutory}_{b,j,t+1}(1-\zeta^{ded}_{b,j,t+1}) + \zeta^{cred}_{b,j,t+1}\right]}$.

A.4.3 Equilibrium Implications

Labor The equalization of the marginal products of labor from Equation (A.32) between any two sectors j and k implies:

$$\Phi^{W}_{corp,j,t} A_{j,t} \frac{\zeta_j}{Y_{j,t}} \left(\frac{K_{j,t}}{L_{j,t}}\right)^{\alpha^f_j} = \Phi^{W}_{corp,k,t} A_{k,t} \frac{\zeta_k}{Y_{k,t}} \left(\frac{K_{k,t}}{L_{k,t}}\right)^{\alpha^f_k}$$

Define the shares of labor allocated to sectors $j \in \{1, \ldots, n-1\}$ as:

$$l_{j,t} = \frac{L_{j,t}}{L_t}$$

We can back out $l_{j,t}$ for each sector, which is equal to:

$$l_{j,t} = \frac{\Phi^W_{corp,j,t}\zeta_j}{\sum_{k=1}^n \Phi^W_{corp,k,t}\zeta_k}. \tag{A.35}$$

and using Equation (A.24) we can back out the amount of labor to the corporate and the pass-through businesses.

Capital We use the FONC with respect to capital, equation A.34, as well as Equations (A.24) and the marginal product of capital for the intermediate good production function in order to identify the share of capital that is allocated to each sector.

On the right side, we have $p_{j,t} = \zeta_{j,t}\frac{Y_t}{Y_{j,t}}$. The sector-specific TFP $A_{j,t+1}$ cancels out on both sides as well.

$$\alpha^f_{j,t+1}\left(\frac{K_{j,t+1}}{L_{j,t+1}}\right)^{\alpha^f_{j,t+1}-1} = \frac{\Phi^R_{b,j,t+1}}{\zeta_{j,t+1}Y_{t+1}}K^{\alpha^f_{j,t+1}}_{j,t+1}L^{1-\alpha^f_{j,t+1}}_{j,t+1}$$

where the capital to labor ratio for the corporation on the left side is replaced with the sector capital to labor ratio (which is the same). This allows us to solve for the ratio of capital in sector j relative to sector 1 for all sectors j:

$$\frac{K_{j,t+1}}{K_{1,t+1}} = \frac{\zeta_{j,t+1}\alpha^f_{j,t+1}(\Phi^R_{j,t+1})^{-1}}{\zeta_{1,t+1}\alpha^f_{1,t+1}(\Phi^R_{1,t+1})^{-1}} \tag{A.36}$$

Add up all of the ratio of capital in the j sector relative to sector 1 for all j and we get:

$$\frac{K_{t+1}}{K_{1,t+1}} = \frac{\sum_{j=1}^n \left(\zeta_{j,t+1}\alpha^f_{j,t+1}(\Phi^R_{j,t+1})^{-1}\right)}{\zeta_{1,t+1}\alpha^f_{1,t+1}(\Phi^R_{1,t+1})^{-1}} \tag{A.37}$$

$$K_{1,t+1} = \frac{\zeta_{1,t+1}\alpha^f_{1,t+1}(\Phi^R_{1,t+1})^{-1}}{\sum_{j=1}^n \left(\zeta_{j,t+1}\alpha^f_{j,t+1}(\Phi^R_{j,t+1})^{-1}\right)} \tag{A.38}$$

which gives capital in sector 1 as a share of all capital, which is the conditional that will allow us to allocate the economy-wide stock of productive capital to each sector.

A.4.4 Pass-through firms

The decisions of a sector j pass-through firm arise as a residual in this economy. Their demand for labor and capital inputs are determined by equations (A.35) and (A.38) and the parameter that determines the split of capital in corporate and pass-through, $\zeta_{b,j,t}^{income}$. Recall that investment and production credits, other expenses, and capital adjustment costs occur at the firm level, all other taxes/subsidies occur at the household level so some parameters in the equations above are zero, capturing that fact.

A.5 The Government

We assume that the government's policy schedule, ψ_t^T, which determines both current and future policy as of year t, is credible. The government collects taxes and makes its consumption and transfer spending as scheduled in ψ_t^T.

The government's income tax revenue, $T_{I,t}$, payroll tax revenue for Social Security and Medicare, $T_{P,t}$, consumption (or other) tax revenue, $T_{C,t}$, corporate or firm tax revenue $T_{corp,t}$ and estate tax revenue $T_{E,t}$ are

$$T_{I,t} = \int_S \tau_{I,t}(y_t^T(j,h,z,l,\iota,w_t), \bar{r}_t a, tr_{OASI,t}(j,b,p_{t-j}^t))\, dX_t(s), \tag{A.39}$$

$$T_{P,t}(\tau_{O,t},\tau_{HI,t}) = \int_S \tau_{P,t}(y_t^T(j,h,z,l,\iota,w_t); \tau_{O,t},\tau_{HI,t})\, dX_t(s), \tag{A.40}$$

$$T_{C,t}(\tau_{C,t}) = \int_S \tau_{C,t}\,(\psi_t^C c)\, dX_t(s), \tag{A.41}$$

$$T_{corp,t} = \sum_b \sum_j T_{b,j,t} \tag{A.42}$$

$$T_{E,t} = \int_S (1-\delta_t(j,h))\tau_{E,t}(a' - \Xi(a'))] dX_t(s), \tag{A.43}$$

$$T_{S,t} = \int_S \tau_{state}\eta_1^y y_t^T(j,h,z,l,\iota,w_t) \tag{A.44}$$

where $\tau_{O,t}$ is an OASDI payroll tax rate and $\tau_{HI,t}$ is a Medicare Hospital Insurance tax rate. $\tau_{corp,t}$ is a function which maps a history of capital stocks $K_{t-T_f+1}^{t+1}$ from $t - T_f + 1$ to $t + 1$ and labor demand to firm tax payments. $T_{S,t}$ are tax revenues collected by the state government.

The government's consumption spending, $C_{G,t}$, Social Security transfer spending, $TR_{OASI,t}$, net of premiums Medicare, $M_{Mcare,t}$, Medicaid, $M_{Mcaid,t}$, spending and SNAP and SSI transfers $SN_t,\ SSI_t$ are

$$C_{G,t}(c_{G,t}) = \int_S c_{G,t}\, dX_t(s) \tag{A.45}$$

$$TR_{OASI,t} = \int_S tr_{OASI,t}(j, b, p_{t-j}^t)\, dX_t(s), \tag{A.46}$$

$$M_{ACA,t} = \int_S \mathbf{1}_{[(i,y,j,a)=ACA]} \pi_{ACA,t}^{PremSub}(y_t^T) dX_t(s) \tag{A.47}$$

$$\begin{aligned} M_{Mcare,t} = \int_S \Big(& \iota \cdot (m_t(j, h, d, i = Mcare) \\ & - oop_t(m, i = Mcare)) - p_{Mcare,t} \Big)\, dX_t(s), \end{aligned} \tag{A.48}$$

$$\begin{aligned} M_{Mcaid,t} = \int_S \Big(& \mathbf{1}_{j<65,\iota=1} \Big(m_t(j, h, d, i = Mcaid, X_t) - oop_t(m, i = Mcaid) \Big) \\ & + \mathbf{1}_{j \geq 65} \Big(p_{Mcare,t} + \iota \cdot (oop_t(m, i = Mcare) - oop_t(m, i = Mcaid)) \Big)\, dX_t(s), \end{aligned} \tag{A.49}$$

$$TR_{SNAP,t} = \int_S SN_t(s, l)\, dX_t(s),$$

$$TR_{SSI,t} = \int_S SSI_t(s, l)\, dX_t(s), \tag{A.50}$$

$$\tag{A.51}$$

where $c_{G,t}$ is government consumption per household; For those who qualify for Medicaid and who are younger than 65, Medicaid covers medical expenses in excess of the out-of-pocket spending requirements. For those who are 65 and older, Medicare is the primary payer, which means that Medicare is responsible for covering the bulk of health care expenses, while Medicaid only covers Medicare premiums, $p_{Mcare,t}$, and the difference in out-of-pocket spending between what Medicaid and Medicare require, $oop_t(m, i = Mcare) - oop_t(m, i = Mcaid)$.

Government Debt and Interest Payments The government's primary net-of-interest deficit pd_t is

$$pd_t = \big[\, C_{G,t}(c_{G,t}) + TR_{OASI,t} + M_{ACA,t} + M_{Mcare,t} + \varrho_t^M \, M_{Mcaid,t} + \varrho_t^{SN} TR_{SNAP,t} + TR_{SSI,t} \tag{A.52}$$

$$- T_{I,t} - T_{P,t}(\tau_{O,t}, \tau_{HI,t}) - T_{C,t}(\tau_{C,t}) - T_{corp,t} - T_{E,t} - T_{exog,t} + mm_t \,\big],$$

where ϱ_t^M (ϱ_t^{SN}) is the share of total Medicaid (SNAP) expenses born by the federal government and $T_{exog,t}$ is the aggregate of a number of exogenous sources of revenue that are not explicitly modeled in the dynamic OLG framework (including gift taxes, excise taxes, and a few other revenue categories). The states' share of Medicaid (SNAP) expenses is thus $1 - \varrho_t^M$ $(1 - \varrho_t^{SN})$ and is financed by the flat state income tax $\tau_{\text{state},t}$. As we will sometimes exogenously forecast the path of government debt, we allow for "magic money", mm_t. When debt is exogenous, mm_t will then be determined by (A.52). When debt is endogenous, then mm_t will be exogenous.

The state government always runs a balanced budget so that in each period:

$$T_{S,t} = (1 - \varrho_t^M) \, M_{Mcaid,t} + (1 - \varrho_t^{SN}) TR_{SNAP,t} \tag{A.53}$$

where η_1^y is the share of taxable labor income that is taxable by states.

Each period the government must finance any deficit by issuing new debt. It must also issue new debt to replace retiring debt. We assume that there are a set of bonds of 1 through T_m maturities that the government can choose to issue. We obtain current yield curves on U.S. government debt and forecast expected future yield curves using Microsim: $yc_t : \mathcal{T}_{\Updownarrow} \rightarrow \mathbb{R}$, where $\mathcal{T}_{\Updownarrow} = \{1, 2, ..., T_m\}$.

We assume that the government issues debt with initial maturities in $IM \subset \mathcal{T}_{\Updownarrow}$. Based on current U.S. government issuance, we set $IM = \{1, 2, 3, 5, 7, 10, 30\}$.

At any time t, suppose the government enters the period with an outstanding debt maturity schedule that is a $(T_m \times 1)$ vector d_t and an effective coupon rate on their outstanding debt that is a $(T_m \times 1)$ vector ey_t. The net interest the government owes that period is $NI_t^G = d_t' ey_t$. Total government debt outstanding is $D_t^G = d_t' 1_{(T_m \times 1)}$ and evolves according to $D_{t+1}^G = NI_t^G + pd_t + D_t^G$. To finance this new debt level D_{t+1}^G the government must issue new debt with face value $d_t(1) + \Delta D_{t+1}$. The share of this new debt that is maturity $im \in IM$ is assumed to be set according to:

$$snd_t(im) = \frac{\sum_{m=IM_{im-1}}^{IM_{im}} d_t(m+1)}{D_t - d_t(1)}, \tag{A.54}$$

where IM_{im} refers to the im'th element of the set IM. The amount of new debt with maturity im is then $nd_t(im) = (d_t(1) + \Delta D_{t+1}) snd_t(im)$. We set $nd_t(m) = 0$ for $m \notin IM$, so that nd_t is defined for all maturities. The law of motion for debt at maturity m is then $d_{t+1}(m) = nd_t(m) + d_t(m+1)$. The effective coupon then evolves as a weighted average of the rate on old debt and new debt:

$$ey_{t+1}(m) = \frac{d_t(m)}{d_{t+1}(m)} ey_t(m+1) + \frac{nd_t(m)}{d_{t+1}(m)} yc_t(m) \tag{A.55}$$

The average government bond yield is then $r_{D,t+1} = \frac{NI_{t+1}^G}{D_{t+1}^G}$, which is in the time t information set.

We assume that government bonds are held by both domestic and foreign investors and that foreign investors hold only government bonds and not domestic private capital. The amount of foreign wealth invested in government bonds at any time t is $W_{F,t}$. We assume that new foreign-owned holdings of U.S. federal debt grow as a share of new, non-rolled-over issues of debt so that foreign holdings obey the following law of motion:

$$W_{F,t+1} = W_{F,t} + \phi_t^D (D_t - D_{t-1}) \tag{A.56}$$

A.5.1 Mutual Fund Assets Markets

Dividends A representative financial intermediary manages a mutual fund, pooling the households' savings (net of depreciation) into deposits to be invested in capital markets, that is, on government bonds and physical capital. Government bonds earn a lower "safe" interest rate relative to the return on productive capital. Since there is no aggregate uncertainty in the model, we assume the share of government bonds in the portfolio managed by the mutual fund is such that it clears the government's need for funding.

Aggregate deposits are:

$$A_{t+1} = \int_{\mathbf{s_t}} a'(s, \Psi_t^{t+J_D-j}) dX_t(\mathbf{s_t}) \tag{A.57}$$

In period t, uses A_{t+1} to purchase K_{t+1} units of physical capital and holds a stock $D^G_{t+1} - W_{F,t+1}$ of government bonds. Given the price of capital fluctuations, the mutual fund realizes capital gains or losses, $\Delta q_t = q_t - q_{t-1}$. The fund pays households the principle and a return on the portfolio chosen in the previous period, $(1 + r_t)A_t$.

The zero-profit condition on the financial intermediary implies that households receive a return on their deposits equal to:

$$\begin{aligned} r_t = & \left.\frac{R_{corp,t}}{A_t}\right\}\text{return on } K_{corp,t} = r_t^{corp} \\ & + \left.\frac{R_{pass,t}}{A_t}\right\}\text{return on } K_{pass,t} = r_t^{pass} \\ & + \left.\frac{r_t^D(D_t^G - W_t^F)}{A_t}\right\}\text{return on } (D_t^G - W_t^F) = r_t^G \\ & + \left.\frac{\Delta q_t K_t}{A_t}\right\}\text{return on capgains } = r_t^C \end{aligned} \tag{A.58}$$

Taxable Income The dividends that the household gets, however, are not the same as their taxable income, which can vary significantly from the actual payments. First, the mutual provides the taxable return on the portfolio that originates from government bonds: r_t^G. This dividend income is entirely taxed at the ordinary rate.

Second, the mutual fund reports the taxable income from corporate sources: r_t^{corp}. Note that these are equal to the dividends paid by the corporation (and capital gains on corporate capital). All credits and deductions have already been applied before the business tax was applied. Whatever is left over is taxable income for the household. Most of this money is taxed at the preferred rate, however, some of it may be taxed at the ordinary rate.

Third, the mutual fund reports the taxable income from pass-through sources of income: r_t^{pass}. Unlike corporations, pass-through taxable income is not the amount of money that is being paid out in dividends, but generally is equal to the dividends minus tax deductions. We do not apply any credits beyond investment credits that may be included in the capital investment allowance schedule. Additional deductions may be applied at the household level. Most of this money is treated as ordinary income for the purposes of taxation, however, the share is determined by the household problem.

Fourth, the mutual fund reports capital gains, $r_t^C = \frac{\Delta q_t K_t}{A_t}$, some fraction of which may or may not be taxed at the preferred rate.

A.6 Aggregation

Total private wealth, A_{t+1}, federal debt held by the public, D_t^G, domestic capital stock, K_t, labor supply in efficiency units, L_t, and aggregate medical expenditures M_t are

$$A_{t+1} = \int_S a' \, dX_t(s), \tag{A.59}$$

$$K_{t+1} = A_{t+1} + W_{F,t+1} - D_{t+1}^G, \tag{A.60}$$

$$L_t = \int_S e(s)(\Omega_{a,t}(j) l_{max} - l(s, \Psi_t^{t+J_D-j})) \, dX_t(s), \tag{A.61}$$

$$M_t = \int_S \frac{\iota \cdot m_t(j,h,d,i,X_t)}{1-\kappa_i} \, dX_t(s), \tag{A.62}$$

A.7 Competitive Equilibrium

Let the set of prices be $p_t = \{r_t, r_t^{corp}, r_t^{pass}, r_t^G, r_t^C, r_t^K, r_t^D, w_t, p_t^p, p_{j,t}\}$. Let the set of government policies be:

$$\psi_t = \{c_{G,t}, tr_{OASI,t}(\cdot), Med_{Mcare,t}, \pi_{ACA,t}^{PremSub}(\cdot), Med_{Mcaid,t}, \tau_{I,t}(\cdot), \tau_{P,t}(\cdot), \tau_{C,t}, \tau_{state,t}, \tau_{E,t}(\cdot), \{\tau_t^{statutory}\}\}.$$

Given an initial distribution at date T_0, X_{T_0}, initial debt $D^G_{T_0}$, effective coupon curve yc_{T_0} and foreign wealth $W^F_{T_0}$, and history of capital choices $\{K_{b,j,t}\}^{T_0}_{t=T_0-T_f}$ and a set of scaling factors and yield curves, $\{\pi_{j,t}, yc_t\}^{\infty}_{t=T_0}$, a competitive equilibrium at time T_0 is a sequence of sets $\{X_{t+1}, K_{t+1}, L_t, \psi_t, p_t, D^G_{t+1}, W^F_{t+}$ and optimal polices for the household and firm, respectively, Γ_h and Γ_f such that:

1. Markets clear: eqs (A.59)-(A.62) $\forall t = T_0, ..., \infty$
2. Γ_h, $\{\Gamma^f_{b,j}\}_{\{corp,pass\},\{1,...,n\}}$, Y_{T_0} solve the household, intermediate and final goods firms' problems, respectively
3. X_{t+1} updates according to eq (A.21) $\forall t = T_0, ..., \infty$
4. The government budget constraint (A.52) holds for all $t = T_0, ..., \infty$, given (A.56) and either D^G_{t+1} or mm_t exogenous.
5. The state government budget constraint (A.53) holds for all $t = T_0, ..., \infty$
6. Mutual Funds earn zero profits $t = T_0, ..., \infty$: (A.58)
7. Health insurers earn zero profits $t = T_0, ..., \infty$: (A.22)

The economy is in a stationary (steady-state) equilibrium at time t if, in addition, $X_k = X_{k+1}$, $p_k = p_{k+1}$, $\pi_{j,k} = \pi_{j,k+1}$, $D^G_{k+1} = D^G_k$, $mm_{k+1} = mm_k$ and $\psi_k = \psi_{k+1}$ for all $k = t, \ldots, \infty$.

A.7.1 Competitive Equilibrium with Terminal Steady State

Given an initial distribution at date T_0, X_{T_0}, initial debt $D^G_{T_0}$ and foreign wealth $W^F_{T_0}$, and history of capital choices $\{K_{b,j,t}\}^{T_0}_{t=T_0-T_f}$ and a set of scaling factors $\{\pi_{j,t}\}^{\infty}_{t=T_0}$, a competitive equilibrium at time T_0 with Terminal Steady State is a Competitive Equilibrium such that there exists a T such that for $X_k = X_{k+1}$, $p_k = p_{k+1}$, $\pi_{j,k} = \pi_{j,k+1}$, $D^G_{k+1} = D^G_k$, $mm_{k+1} = mm_k$ and $\psi_k = \psi_{k+1}$ for all $k > T$.

An Initial Steady State is simply a Competitive Equilibrium with Terminal Steady State such that the steady state date, $T = T_0$.

A.8 Computing an Initial Steady State

Given a set of parameters θ, a set of government policies, ψ_{T_0}, and a desired debt level $D^G_{T_0}$ and level of foreign investment $W^F_{T_0}$, to compute an Initial Steady State we follow these steps:

1. Set the scaling factors, $\pi_{j,T_0} \,\forall\, j \in J$
2. Guess a price vector p_{T_0}, a state tax τ_{state}, a survival function δ and a household distribution X_{T_0}.
3. Solve the household's problem.
4. Compute law of motion for households and update the distribution of households.
5. Aggregate household policies to get aggregate savings, labor and insurance demand.
6. Back out Capital Supply from Aggregate Savings using $D^G_{T_0}$ and $W^F_{T_0}$.
7. Apportion labor and capital to intermediate firms using (A.24),(A.35) and (A.36)-(A.38). Solve for Aggregate Output Y_{T_0} and interest rate vector $\tilde{r}_{T_0}$
8. Find "new guess" of prices, $p_{T_0} = \{r_{T_0}, \tilde{r}_{T_0}, w_{T_0}, p^p_{T_0}, p_{j,T_0}\}$ using (A.58),(A.32),(A.23) and (A.22). Update the state tax rate using (A.53). Check if the new guesses are close to guesses in Step 2. Likewise check if model survival rates (conditional on age but not health) are close to those in the data. If either of these checks fail, take new guesses and go to Step 2.
9. Back out a magic money amount mm_{T_0} using (A.52)

Let the operator ISS take vectors of parameters, θ, into vectors in $\mathbb{R}^{N^I}$ as described by this implementation. In other words, for a given set of parameters θ, $ISS(\theta)$ will find an Initial Steady State and then output a set of simulated data from that Steady State.

To calibrate θ for the initial steady state, we seek:

$$\theta^I = \arg\min_{\theta \in \Theta} \Xi(ISS(\theta), \mathcal{M}^I) \tag{A.63}$$

where $\Xi : \mathbb{R}^{NI} \times \mathbb{R}^{NI} \rightarrow \mathbb{R}^{+}$ is a distance function and $\mathcal{M}^I \subset \mathbb{R}^{NI}$. To compute θ^I, we guess an initial value, θ_0, compute $ISS(\theta_0)$ and then update each element of our guess, $\theta_{1,i} = \theta_{0,1} + k_i(ISS(\theta_0)_i - \mathcal{M}^I_i)$, where $k_i \in \mathbb{R}$ is chosen by the user. We iterate on this until $\Xi(ISS(\theta_n)_i - \mathcal{M}^I_i) < \epsilon_{tol}$. We describe θ^I and $\mathcal{M}^I$ in the following section.

A.9 Computing an Baseline Transition Path

Given an initial year, T_0, a closure year T_c, and a final steady state year, T_f, a set of parameters $\{\theta_t\}_{t=T_0}^{T_f}$ (including health insurance premium growth rates ν_i), a set of government policies, $\{\psi_t\}_{t=T_0}^{T_c}$, a desired age-dependent population level $\{x_{j,t}^{est}\}_{t=T_0}^{T_c} \quad \forall$j$\in [J_B, J_D]$, a desired debt level $\{D_t^G\}_{t=T_0}^{T_c}$ and an initial level of foreign investment $W_{T_0}^F$, to compute a Baseline Transition Path we follow these steps:

1. Compute an ISS using steps above. This will provide an initial distribution X_{T_0} and history of capital choices $\{K_{b,j,t}\}_{t=T_0-T_f}^{T_0}$.
2. Set $D_t^G = D_{t-1}^G$ and $W_t^F = W_{t-1}^F$ for all $T_f \geq t > T_c$.
3. Compute the final steady state at time T_f using the same steps as the ISS above. This will provide a set of prices and policies p_{T_f} and ψ_{T_f}.
4. Guess survival rate function $\{\delta_t\}_{t=T_0}^{T_f}$[17]
5. Guess a path of prices, household distributions and state tax rates for all $t \in \{T_0, ..., T_f - 1\}$.
6. Iterate backwards from T_f to T_0, computing the optimal household and firm's policy at each year: $\{\Gamma_{h,t}, \Gamma_{f,t}\}$.
7. Use $\{\Gamma_{h,t}, \Gamma_{f,t}\}_{t=T_0}^{T_f}$ to update X_t and capital choices $\{K_{b,j,t}\}_{t=T_0}^{T_f}$.
8. Check market clearing conditions for each $t \in \{T_0, ..., T_f - 1\}$ by: aggregating household policies to get aggregate savings, labor and insurance demand; backing out Capital Supply

17. The δs are included in the set of parameters $\{\theta_t\}_{t=T_0}^{T_f}$ and δ_{T_0} from this set is used to compute ISS. For subsequent steps though we will adjust the δs.

from Aggregate Savings using D_t^G and W_t^F; apportioning labor and capital to intermediate firms using (A.24),(A.35) and (A.36)-(A.38) and solving for Aggregate Output Y_t and interest rate vector $\tilde{r}_t$.

9. Find "new guess" of prices, $p_t = \{r_t, \tilde{r}_t, w_t, p_t^p, p_{j,t}\}$ using (A.58),(A.32),(A.23) and (A.22). Likewise, update guesses of state tax rates and survival functions δ. Check if new guesses are close to guess in Steps 4-5. If not, take new guess and go to Step 4.

10. Back out a magic money amount $\{mm_t\}_{t=T_0}^{T_f}$ using (A.52)

Let the operator BT take sequences of vectors of parameters, $\{\theta_t\}_{t=T_0}^{T_f}$, into sequences of vectors in $\{\mathbb{R}^{N^I}\}_{t=T_0}^{T_f}$ as described by this implementation. In other words, for a given sequence of parameters $\{\theta_t\}_{t=T_0}^{T_f}$, $BT(\{\theta_t\}_{t=T_0}^{T_f})$ will find an Initial Steady State, compute a Competitive Equilibrium with Terminal Steady State and magic money and then output a sequence of simulated data.

A.10 Computing a Policy Transition Path

We wish to compute the effects of a change in taxes, transfers and government spending. We will keep mm_i and ν_i at their baseline levels. We assume that the new policy parameters $\hat{\psi}_t$ can vary arbitrarily from the baseline parameters ψ_t. During closure we require $\hat{\psi}_t = \hat{\psi}_{t-1}\ \forall t > T_c$ with one exception: we will allow a tax rate (currently a consumption tax, $\hat{\tau}_c$) to vary year by year during and after closure in order to ensure the budget balances period by period with no change in government debt for each year $t \geq T_c$.

After computing a Baseline Transition Path, we can compute a Policy Transition Path with the same given initial year, T_0, closure year T_c, final steady state year, T_f, set of parameters $\{\theta_t\}_{t=T_0}^{T_f}$, a desired age-dependent population level $\{x_t^{est}\}_{t=T_0}^{T_c}$ and new set of government policies: $\{\hat{\psi}_t\}_{t=T_0}^{T_c}$ by following these steps:

1. Obtain the path of magic money, scaling factors and survival rates, $\{mm_t, \pi_{j,t}, \delta_t\}_{t=T_0}^{T_f} \in$

$BT(\{\theta_t\}_{t=T_0}^{T_f})$ as well as the initial distribution X_{T_0}, history of capital choices $\{K_{b,j,t}\}_{t=T_0-T_f}^{T_0}$, and initial debt level $D^G_{T_0}$ from the baseline's Transition path.

2. Guess a $D^G_{T_c}$ and $W^F_{T_c}$ and set $D^G_t = D^G_{t-1}$ and $W^F_t = W^F_{t-1}$ for all $T_f \geq t > T_c$.

3. Compute the final steady state at time T_f by following:

   (a) Guess a price vector p_{T_f}, a state tax rate τ_{state,T_f}, a household distribution X_{T_f} and a consumption tax rate T_{c,T_f}.

   (b) Solve the household's problem.

   (c) Compute law of motion for households and update the steady state distribution of households.

   (d) Aggregate household policies to get aggregate savings, labor and insurance demand.

   (e) Back out Capital Supply from Aggregate Savings using $D^G_{T_f}$ and $W^F_{T_f}$.

   (f) Apportion labor and capital to intermediate firms using (A.24),(A.35) and (A.36)-(A.38). Solve for Aggregate Output Y_{T_f} and interest rate vector $\tilde{r}_{T_f}$

   (g) Find "new guess" of prices, $p_{T_f} = \{r_{T_f}, \tilde{r}_{T_f}, w_{T_f}, p^p_{T_f}, p_{j,T_f}\}$ using (A.58),(A.32),(A.23) and (A.22) and similarly for state tax rates. Check if new guesses are close to guess in Step 3a. If not, take new guess and go to Step 3a.

4. Guess a path of prices, household distributions, state taxes and consumption taxes for all $t \in \{T_0, ..., T_f - 1\}$.

5. Iterate backwards from T_f to T_0, computing the optimal household and firm's policies at each year: $\{\Gamma_{h,t}, \Gamma_{f,t}\}$.

6. Use $\{\Gamma_{h,t}, \Gamma_{f,t}\}_{t=T_0}^{T_f}$ to update X_t and obtain capital choices $\{K_{b,j,t}\}_{t=T_0}^{T_f}$.

7. Check market clearing conditions for each $t \in \{T_0, ..., T_f - 1\}$ by: aggregating household policies to get aggregate savings, labor and insurance demand; backing out Capital Supply

from Aggregate Savings using D_t^G and W_t^F; apportioning labor and capital to intermediate firms using (A.24),(A.35) and (A.36)-(A.38) and solving for Aggregate Output Y_t and interest rate vector $\tilde{r}_t$.

8. Find "new guess" of prices, $p_t = \{r_t, \tilde{r}_t, w_t, p_t^p, p_{j,t}\}$ using (A.58),(A.32),(A.23) and (A.22). Likewise, update guesses of $\tau_{state,t}$ and $\tau_{c,t}$ using (A.53) and (A.52). Check if new guesses are close to guess in step 3a. If not, take new guess and go to Step 3.

B Calibration and Model Performance

B.1 Computational Parameterization of Initial Steady State

Define the function $M^I : \mathbb{R}^{N^I} \times \mathbb{R}^{N^I} \rightarrow \mathbb{R}$ be a criterion function. A calibrated initial steady state ("ISS") is a set of internal parameters θ^I such that $M^I(ISS(\theta^I), M^*) = 0$.[18]

For an initial steady-state, some model variables, such as government outlays and government debt, are set to match data from PWBM's Microsim model. In some cases, the OLG model takes data from the Microsim model directly. In other cases, the OLG model will first project Microsim data onto a small dimension before using it in the model. This is done, for example, for any data that is conditional on demographic variables.

In this section, we discuss the calibration of the parameters and exogenous variables that are not based solely on the Microsim and mention in passing those that are set using the Microsim model.

We calibrate the model's initial steady state to the 2024 U.S. economy with a fiscal policy that is close to the policy prevailing in 2024. In the initial steady state, the economy is assumed to be on a balanced growth path with a constant labor-augmenting productivity growth rate, μ. Individual variables other than working hours are thus growth-adjusted by $(1+\mu)^{-t}$ and aggregate variables are adjusted by $(1+\mu)^{-t}$. We set μ to 0.62 percent, which is based on Microsim projections.

B.2 Demographics

We normalize the measure of households in the initial steady state ($t = T_0$) to unity:

$$\int_S dX_{T_0} = 1.0,$$

Households enter the economy and start working at age $J_B = 21$ and can live to age $J_D = 100$. In the baseline, we assume that households retire and start claiming OASI benefits at current law

18. We could index M^I by all non-internal parameters. For notational simplicity, we do not.

full-retirement age $J_R = 67$, and they become eligible to enroll in Medicare at $J_{Mcare} = 65$ (see Table 23). Household's OASI benefits are calculated on the basis of its growth-adjusted earnings between ages 21 and $J_R - 1$.[19]

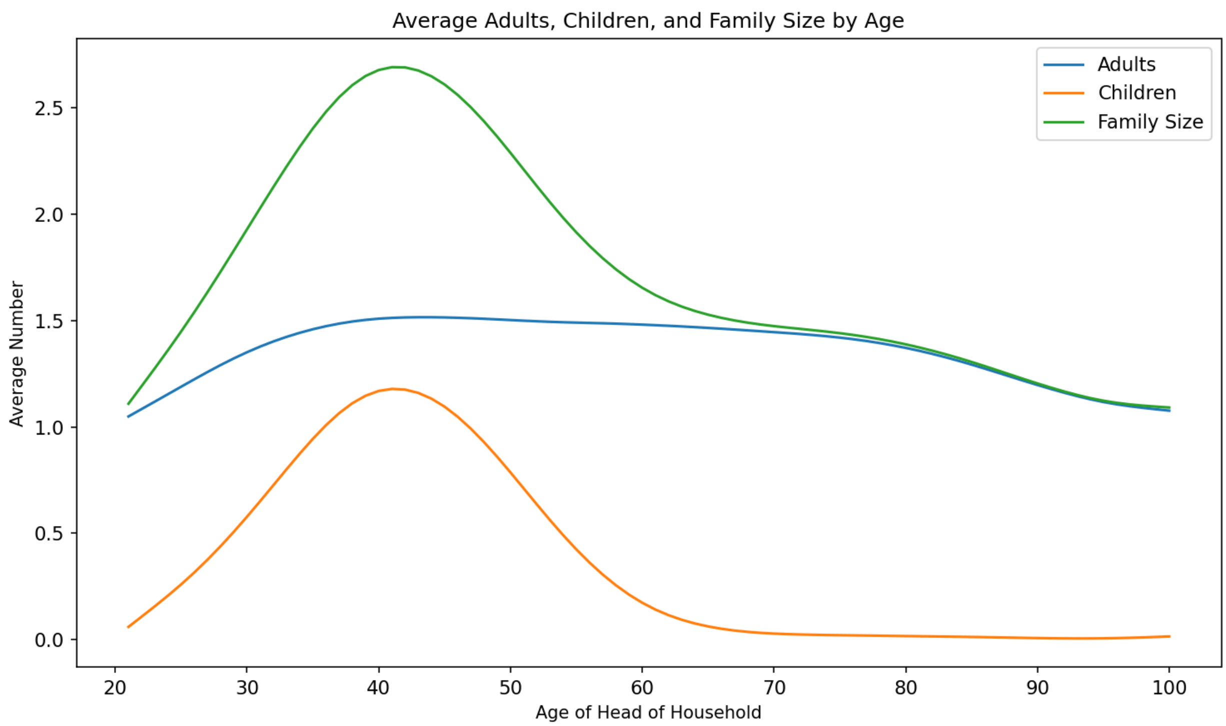


Figure 16: Average Number of Adults and Children per Household by Age of Head of Household in the the initial steady state

Source: Microsim, Interface calculations for 2024

To estimate the average household size, we use data from the Microsim to estimate the average number of adults Ω_{a,T_0} and the average number of children Ω_{c,T_0} by age of the head of household. Figure 16 shows the resulting average number of adults and children by the age of the head of household from the data.

From the Microsim data, for each year $t \in T_0, ..., T_f$, we then create $(J \times 1)$ vector estimates of household population counts, pop_t, the number of households that died $died_t$, emigrations, em_t, and immigrations, im_t. The model explicitly accounts for death, emigration and immigration but

19. In the current Social Security system, AIME is calculated as the average of the highest 35 years of growth-adjusted earnings. However, keeping the previous 35 highest earnings as the household's state variables would make the household problem computationally intractable. In the model economy, therefore, AIME is approximated by the average of growth-adjusted earnings of all ages before J_R.

Table 23: Demographic and Preferences Parameters for the initial steady state

Parameter		Value	Comment
Demographics			
Maximum age	J^{max}	100	
Minimum age to receive OASI benefits	J_R	67	Full retirement age for OASI benefits
Minimum age to receive Medicare benefits	J_{Mcare}	65	Medicare eligibility age
Productivity growth rate	μ	0.006	
Health transition probabilities	P		
Conditional survival rates	s		
Average number of Adults by age	$\Omega_{a,ISS}$		
Average number of Children by age	$\Omega_{c,ISS}$		
Preferences			
Risk aversion parameter	γ	2.722	Target: CRRA = 2.0
Consumption share parameter	α	0.581	
Maximum leisure hours	l^{max}	1.196	
Discount factor	β	0.981	Target: K/Y = 3.0
Growth-adjusted discount factor	$\tilde{\beta}$	0.975	
Minimum consumption	$\underline{c}$	0.070	$8,500 annual minimum consumption
$-per-model unit	dpu	120,825	Matches average labor earnings in nominal dollars
Scale value of death	ζ	2.704	

not for other entry and exit of households (such as marriage, divorce or death of the head but not the spouse). We use the scaling vector, $\pi^s_{T_0}$ to adjust the model number of households at each age so that the resulting relative number of households at each age in the model matches the data using (A.20), where $\hat{x}^{microsim}_{t+1}(j+1) = pop_t(j) - emi_t(j) - died_t(j) + imm_{t+1}(j+1)$. For the initial and final steady states ($T = T_0, T_f$, respectively) we instead let $\hat{x}^{microsim}_{T}(j+1) = pop_T(j) - emi_T(j) - died_T(j) + imm_T(j+1)$.

B.3 Preferences

We assume that the utility function is a combination of Cobb-Douglas and constant relative risk aversion (CRRA) preferences, which is consistent with a growth economy

$$u(c,l) = \frac{\left[c^{\alpha} l^{1-\alpha}\right]^{1-\gamma}}{1-\gamma}. \tag{B.64}$$

We internally calibrate the maximum number of leisure hours per adult, l^{max}, to normalize the average hours worked in the baseline steady state at 1.0. Given that the average hours worked will be 1.0, we can jointly choose the preference parameters α and γ to match a coefficient of relative risk aversion of 2.0 and an average Frisch elasticity of labor supply of 0.5.[20] Our choice for risk aversion of 2.0 is roughly in the middle of the range typically used in the macro public finance literature, and requires setting $\gamma = 2.7626$.[21]

We find that $\alpha = 0.5673$ so that the average household-weighted Frisch elasticity is 0.5 in the model, which is consistent with the literature surveyed in **Reichling.Whalen:2012**.[22] To normalize the average number of working hours to 1.0, we set the maximum time endowment per adult, $l^{\max}$, to 1.2028.

We choose the discount factor, β, so that the capital–output ratio, K_t/Y_t, in the benchmark economy is 3.0, which is commonly used in the literature. The growth adjusted discount factor $\tilde{\beta} = \beta(1+\mu)^{\alpha(1-\gamma)}$ can then be calculated to equal 0.9724.

Government programs provide a consumption floor $\underline{c}$ that, depending on the reason for low resource, will pay benefits through Medicaid and/or SNAP. Medicaid pays if usually high medical expenditures are the cause of low resources and SNAP provides benefits if low resources are the result of a low income realization. We set the consumption floor to $8,500 in 2024 dollars, which is informed by estimates from Moffitt 2020.[23]

20. The coefficient of relative risk aversion (CRRA) can be calculated as $(1 - \alpha + \alpha * \gamma)$. The average Frisch elasticity is $\frac{(1-\alpha+\alpha*\gamma)}{\gamma} \frac{l^{max}\bar{\Omega}-\bar{h}}{\bar{h}}$ where $\bar{h}$ is the average hours worked in the model economy and $l^{max}\bar{\Omega}$ is the average leisure maximum.

21. For example, **Domeij.Heathcote:2004** use 1.0; **Imrohoroglu.etal:1995b** use 2.0; and **Auerbach.Kotlikoff:1987** and **Conesa.etal:2009** use 4.0.

22. The Frisch elasticity for a household in the model is generally positively correlated with a household's wealth, average earnings history and health state (worse health means a higher Frisch elasticity) and negatively correlated with a household's age, productivity, and medical expenditures.

23. Moffitt 2020 estimates the average benefits that non-elderly and non-disabled SNAP families receive from different sources, including the Child Tax Credit (CTC), Supplemental Security Income (SSI), subsidized housing, Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), Temporary Assistance for Needy Families (TANF), and Unemployment Insurance (UI). Estimates of consumption floors from other authors include: Hubbard, Skinner, and Zeldes 1994 use $20,900, Scholz, Seshadri, and Khitatrakun 2006 use $18,000, De Nardi and Yang 2014 use $16,000, **Ozkan:2013** uses $13,000, Moffitt 2020 calculates $10,400, French and Jones 2011 uses $8,400, and De Nardi, French, and Jones 2010 use $5,100 (all numbers adjusted to 2024 dollars using the CPI-U).

B.4 Earnings and Income

minipage=scale=0.89

Figure 17: Earnings Potential by Age and Health Status Change
Source: ASEC 1997-2024 and Penn Wharton Budget Model.

Our earnings process is based on data from the 1997-2024 Annual Social and Economic Supplements (ASEC) of the Current Population Survey (CPS) and the 1996-2023 waves of the Panel Study on Income Dyanamics (PSID) . The calibration of our earnings process generates an earnings distribution that is broadly consistent with the data. As Table 24 shows, the model tracks the data almost exactly up to the 90th percentile and even at the 99th percentile the model performs quite well. The same is true for income and wealth, although the model is less successful at matching the wealth data at the 99th percentile of the distribution, and as a result, is also unable to match minipage=scale=0.89

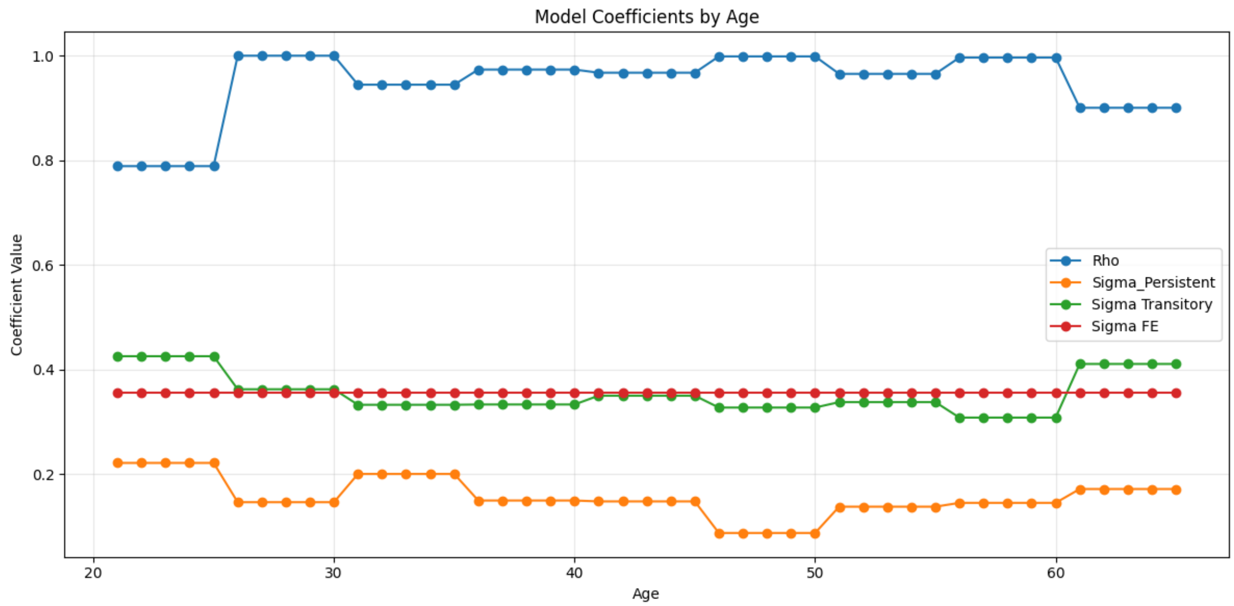


Figure 18: Stochastic Earnings Potential Coefficients by Age
Source: PSID 1996-2023 and Penn Wharton Budget Model.

the income distribution at the very top of the distribution.[24]

Hourly wages in the economy are the product of the efficiency wage w_t, labor productivity e_j, and the health state h_t. We set TFP, $A = 0.845$ to normalize the efficiency wage in the initial steady state to $w_{T_0} = 1$. The working ability $e_j(h)$ of an age-j household in health state h satisfies

$$\ln e(j, h, z) = \ln \bar{e}(j, h) + \ln z$$

for $j = 21, \ldots, 64$, where $\bar{e}(j, h)$ is the average working ability at age j for somebody who is in health state h, and z is the persistent shock that follows an age-dependent AR(1) process,

$$\ln z' = \rho(j) \ln z + \epsilon_z$$

for $j = 21, \ldots, 64$. The shock, ϵ_z, is normally distributed, $\epsilon_z \sim N(0, \sigma_z^2(j))$. The initial distribu-

24. Typically, these kinds of models are unable to generate the amount of wealth at the very top of the distribution because the only savings motive is precautionary. Other model features not related to health are required to make the model more successful at matching the top of the wealth distribution. See Cagetti and De Nardi 2008 for a discussion of the kinds of models that could accomplish that.

minipage=scale=0.89

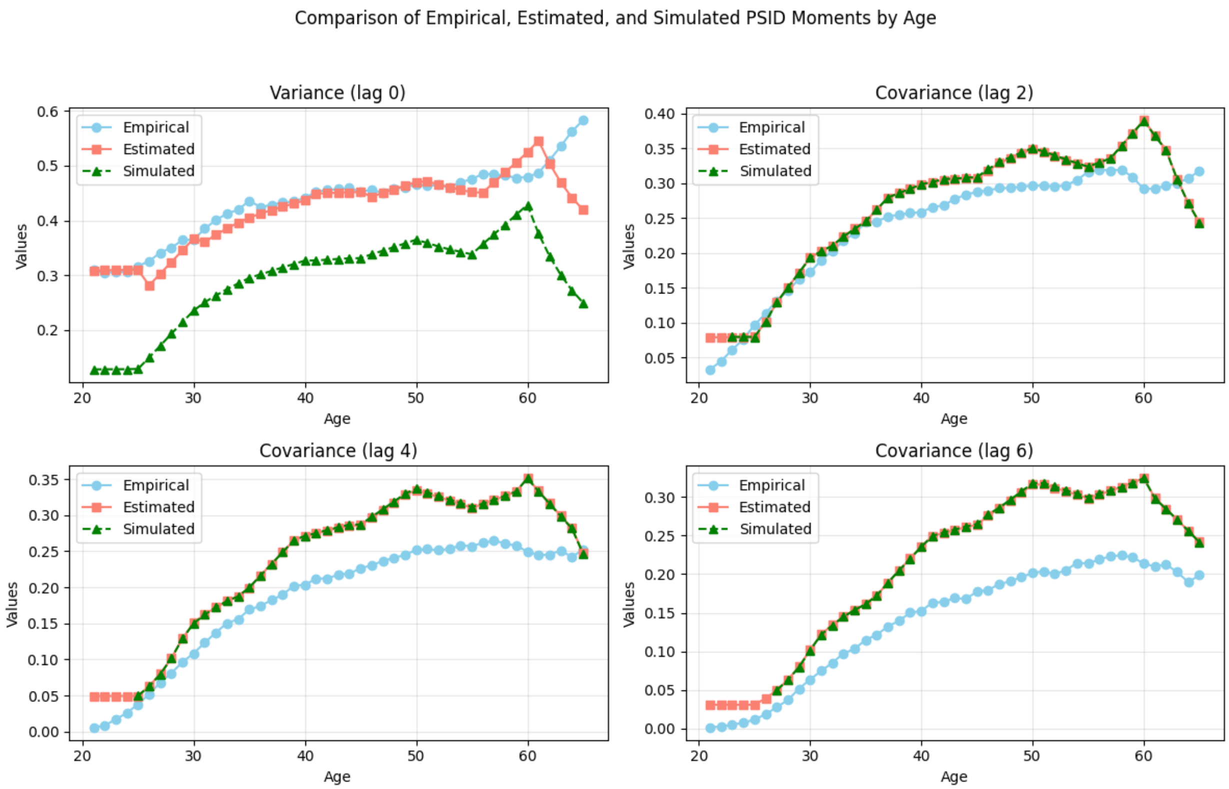


Figure 19: Simulated, Estimated and Data moments by Age

Note: Simulated model profiles do not include transitory shock.
Source: PSID 1996-2023 and Penn Wharton Budget Model.

tion π_z^0 satisfies $\ln z_{21} \sim N(0, \sigma^2_{\ln z_{FE}})$.

We assume average earnings potential, $\bar{e}(j,h)$, follows the specification

$$\ln \bar{e}(j,h) = \alpha^e + \sum_{i=1}^{4} \beta^e_{i,h} \; j^i \tag{B.65}$$

for ages $j = 21, ..., 64$ and include year fixed effects. We use individual-level earnings per hour (total wage income divided by total hours worked). For the ASEC, we use only individuals working in the private sector. The profile of average earnings potential by age and health is estimated by ordinary least squares regression. We estimate the same specification using both the PSID and the ASEC data. Figure 17 shows the estimated values from the ASEC data, which are the ones we

Table 24: Earnings, Income, and Wealth Distributions, Model versus Data

	Percentile of Distribution					
	0.10	0.25	0.50	0.75	0.90	0.99
				Earnings		
Model	18.4	36.7	60.8	119.1	203.1	554.6
Data	21.6	43.2	80.0	140.5	250.8	1091.7
				Income		
Model	23.7	42.9	72.1	138.1	223.8	630.9
Data	21.6	43.2	80.0	140.5	256.2	1149.0
				Wealth		
Model	0.1	17.3	151.1	569.7	1048.8	3185.8
Data	0.6	26.3	168.9	584.4	1669.3	29295.6

Notes: In thousands of U.S. dollars. Data samples are households with head of household under 65 and at least one adult in the labor force. Data from 2022 converted to 2024 dollars using the CPI-U from December 2021 and December 2024.
Source: 2022 Survey of Consumer Finance.

calibrate the model with.

A decrease in health leads to lower wages, and those decreases get larger the bigger the drop from the current health state is. Increases in health status, on the other hand, lead to wage increases.

We estimate the age-dependent parameters ρ and σ_z using the PSID by first estimating the same process in (B.65) and extracting the residuals and then using GMM to fit 0th, 2nd, 4th and 6th order autocovariances of the residuals. While we do not allow for transitory shocks in the OLG model, we allow for a transitory shock with age-dependent variance when estimating rho and σ_z. We estimate the parameters using 5-year age bins (21-26, 26-30, ..., 60-64), thereby assuming that the parameters are constant within-bin but free to vary across bins. Figure 18 plots the age-profiles of the parameters, including the standard deviation of the transitory shock which is not included in the calibrated model. The profiles are similar to those found in the literature.[25]

The log-persistent shock is discretized into nine nodes and the Markov transition matrices are constructed using using the minimum entropy methods of Tanaka and Toda (2013), Farmer and Toda (2017), Fella, Gallipoli, and Pan (2019), and Kirkby (2025) using a 9-point state space. We then add a 10th point into the earnings process to better capture earnings and wealth at the very top of the distribution. The 10th point is defined as $\log 10$ times the value of the 9th point. The transition probability from the 9th to the 10th point is 0.16, while the probability of exiting the 10th to the 9th point is 0.74; all other transition probabilities are 0.

In the baseline economy, the average labor income of all working-age households is 0.9069. The average labor-income of working-age households was $121,044 in 2024, according to the 2024 Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS). Therefore, in the baseline economy, one model unit is approximately equal to $121,044/0.9069 = $133,485 in 2024 growth-adjusted dollars. The model uses that ratio to convert some policy variables into model units.

The calibration of our earnings process generates aggregate earnings, income, and wealth distributions that are broadly consistent with the data. As is typical for these models, we cannot generate the amount of wealth at the very top of the distribution, which would require other model features not related to health.[26] However, the model performs very well up to the 90th percentile of the distributions.

B.5 Inheritances

In the model, household have bequest motives and value having bequeathal assets when dying. ζ, which we calibrate in B.6, helps determine the value. Some of those assets are used to pay for estate costs and some of it may be taxed. The transactions cost function follows Feiveson and

25. For example, Karahan and Ozkan (2013) and De Nardi, Fella, and Paz-Pardo (2020).

26. See Cagetti and De Nardi 2008 for a discussion of the economic models that can explain the high concentration of wealth at the very top.

Sabelhaus (2018a):

$$\Xi(a') = 0.5 \times \max\{0, a' - \$10{,}000\}$$

while tax policy determines the tax function $\tau_{E,t}$.

We distribute inheritances Q according to empirical age and income profiles derived from the 2001-2019 waves of the Survey of Consumer Finances (SCF). That distribution is represented by the PDF $\omega^q(j)$ (see Equation A.8).[27] The model generates an inheritances-to-GDP ratio of 0.6 percent in the initial steady state, which is below the estimates provided by Feiveson and Sabelhaus 2018b (1.5 percent), Hendricks 2002 (2.0 percent), and **Gale.Scholz:1994** (2.6 percent).

B.6 Health

B.6.1 Average Medical Spending by Age

We estimate total medical expenditures using MEPS data for the 2010-2022 period for the 21–84 year old population, controlling for age and health state.[28] For those 85 to 100 years old we linearly extrapolate the spending of 80 to 84 year olds. For each age and health state, we calculate the average medical expenses for three bins, corresponding to the bottom 40 percent ($d = 1$), the middle 50 percent ($d = 2$), and the top 10 percent ($d = 3$). Figure 20 shows the estimated (for 21-84 year olds) and projected (for 85-100 year olds) total households medical expenditures by age and health state of the head of household.

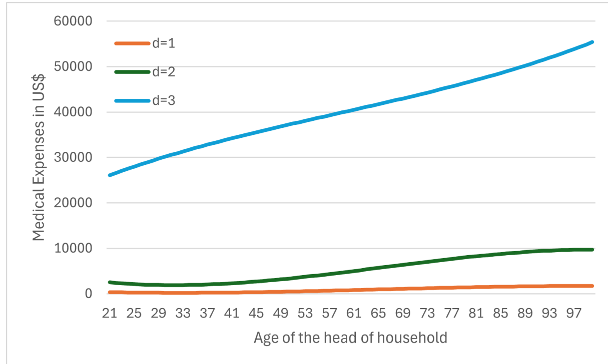
Let $m^H(j, h, d)$ represent the total medical expenses for a head of household of age j, health state h, and expenditure bin d, as derived from the MEPS data. We assume that spouses incur average health expenditures based on the head's age j, while all children incur a fixed average expenditure m^C. Accordingly, spousal medical expenses are defined as $m^S(j) = \mathbb{E}[m^H|j]$, where $\mathbb{E}$ uses the model's probability distribution. Total household medical expenditures are then calculated

27. For more information see https://budgetmodel.wharton.upenn.edu/issues/2021/7/16/inheritances-by-age-and-income-group.

28. The publicly available MEPS data groups all data for those above age 85 into the category for 85 year olds.

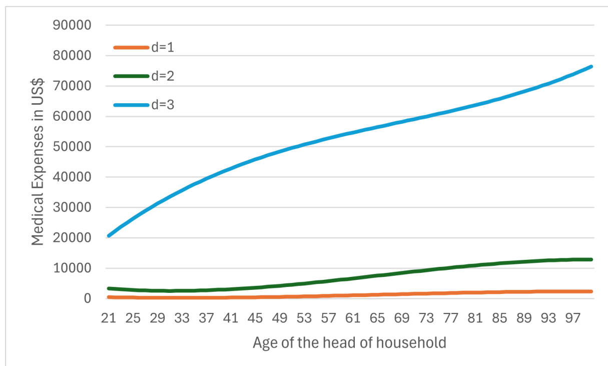
as:

$$m(j,h,d) = m^H(j,h,d) + (\Omega_{a,T_0}(j) - 1)m^S(j) + \Omega_{c,T_0}(j)m^C.$$

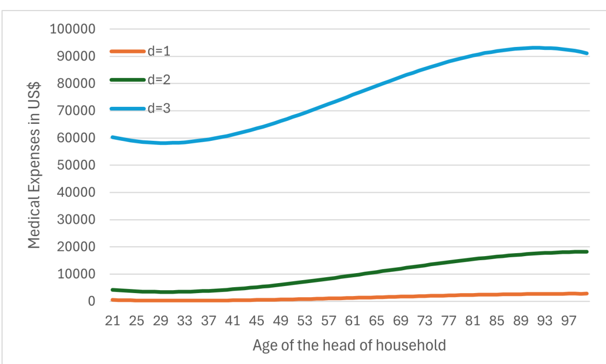


minipage=scale=0.89

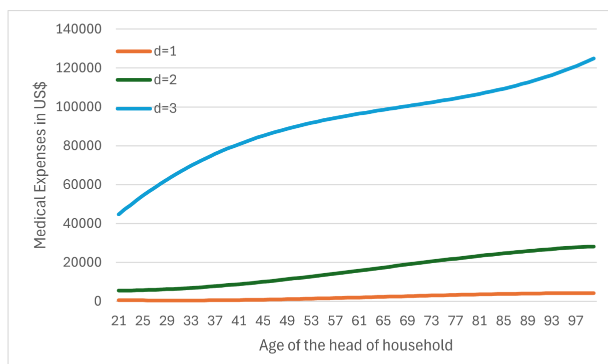
(a) Health State 1



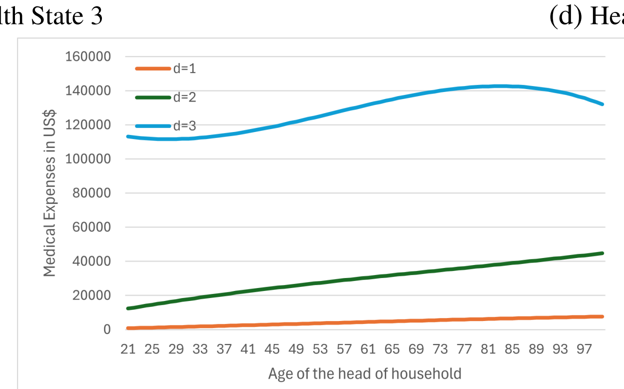
(b) Health State 2



(c) Health State 3



(d) Health State 4



(e) Health State 5

Figure 20: Total Health Expenditures by Age and Health State

Source: PWBM estimates from the Medical Expenditure Panel Surveys.

Table 25: Average Individual Medical Expenditures by Age

Age of Head	Model	Data
21-25	$6,480	$5,800
26-34	$7,750	$6,430
35-44	$8,460	$7,870
45-54	$10,420	$10,210
55-64	$12,750	$13,230
65-74	$15,530	$16,640
75-84	$18,470	$20,000
21-64	$9,290	$8,670
65+	$17,220	$20,250
21+	$11,370	$12,450

Notes: All dollar values are adjusted to 2024 dollars. Dollar values are rounded.
Source: MEPS data for 2010-2022.

Average medical spending in our model matches that in the MEPS very closely. Table 25 shows average individual medical expenses by age groups and compares those to MEPS data. Households with health insurance only pay a fraction of the amounts shown out-of-pocket.

We restrict the scale value of dying, $\zeta > 1$,[29] and calibrate the parameter ζ so that the initial steady state matches the data on the average medical care expenditures of the uninsured in health state 5 relative to those in that health state who are insured (see Table 28):

$$\frac{\int_{S(i=0)} \iota(s(i=0), p_{T_0}, \psi_{T_0}) m(j,h,d) dX_{T_0}}{\int_{S(i=1)} \iota(s(i=1), p_{T_0}, \psi_{T_0}) m(j,h,d) dX_{T_0}} \tag{B.66}$$

where i indicates whether household have health insurance ($i = 1$) or are uninsured ($i = 0$).

B.6.2 Out-of-Pocket Spending

We use MEPS data to estimate the fraction of total medical expenses that households pay as OOP expenses for each insurance type, or $\gamma^{oop}(m, i)$. We focus on households that received insurance

29. $u()$ is strictly negative, so we need $\zeta > 1$ to ensure that households prefer living to death. Hall.Jones:2007, for example, solve that problem by adding a constant to the per-period utility function to ensure that it is a positive number (see their Equation (6) on page 47). We take a different approach here and do not change the standard utility function.

coverage from a single source to ensure that we capture the proper HI reimbursement policy. We find that Medicaid is the most generous with required OOP shares generally below 6 percent, and that Medicare is as generous as Private Health Insurance for large claims, but more generous for smaller claims. Out-of-pocket expenditures are then calculated as $oop_t(m, i) = \gamma^{oop}(m, i) \times m_t(j, h, d)$.

For each insurance product i, we estimate the out-of-pocket share function $\gamma^{oop}(m, i)$ using a fourth-order polynomial. Specifically, we assume:

$$\log(\gamma^{oop}(m, i)) = \sum_{j=0}^{4} \left[(\log(10 \times m))^j \times \text{inscoef}_i(j) \right]$$

where the vector inscoef_i contains five parameters estimated via Ordinary Least Squares (OLS). These parameters are obtained by fitting the polynomial to MEPS data on out-of-pocket expenditure shares across the distribution of total medical expenses.

The results are shown in Figure 21.

B.6.3 Excess Cost Growth

To parameterize the annual excess cost growth rates, ν_t^m for the three different insurance types, we rely on estimates by Congressional Budget Office 2018, which are summarized in Table 26. Over the 2025-2055 period, excess cost growth in private health insurance cumulates to 56.3 percent, 8.9 percentage points more than in Medicaid (47.4 percent), and 16.9 percentage points more than Medicare (39.4 percent).

Note in any steady-state, all excess cost growth rates are set to 1.

B.6.4 Health Transition and Survival Probabilities, and Forgone Treatment

We estimate health transition probabilities, $\pi_h[h'|j, h, d, \iota = 1]$, and survival probabilities, $\delta(j, h)$, from MEPS data (for those younger than 55) and HRS data (for those 55 and older). We use a multinomial logit where health next year, h', and survival to next year depend on current health

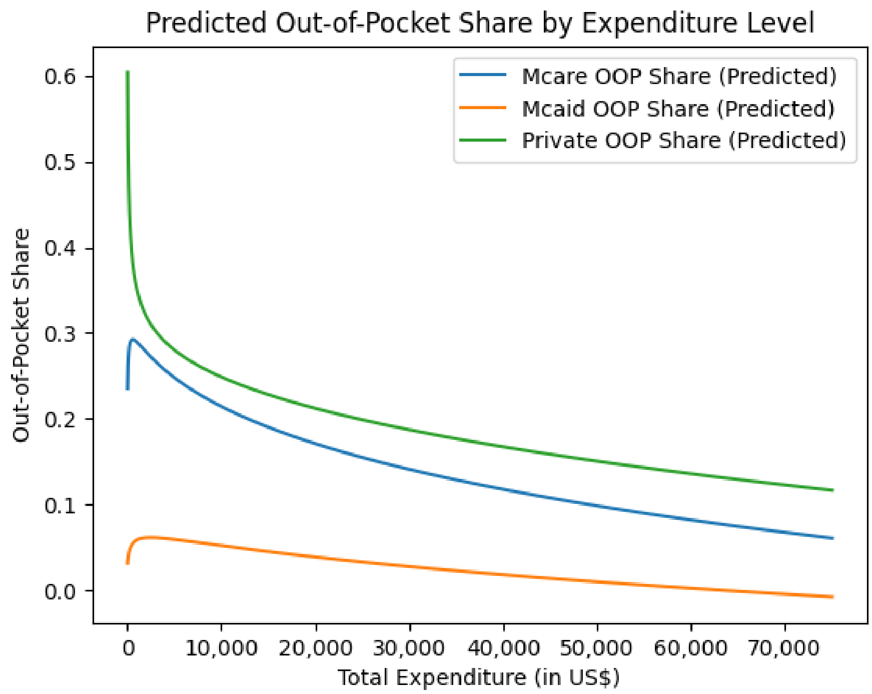


Figure 21: Required OOP shares by total medical expenditures and insurance plan

Source: PWBM estimates from the Medical Expenditure Panel Survey.

h and a quadratic in age j. We use the following self-reported health states: 1=excellent, 2=very good, 3=good, 4=fair, and 5=poor.

Households opting for medical treatment ($\iota = 1$) follow the estimated transition probabilities, while those forgoing treatment ($\iota = 0$) face adjusted probabilities reflecting unmitigated health risks. A primary identification challenge arises because the data does not explicitly label households that experienced a health shock but chose to forgo care. Categorizing all households with zero expenditures as "choosing not to be treated" would introduce significant bias; for many, zero expenditure simply reflects the absence of a health shock rather than a deliberate decision to avoid healthcare.

To resolve this identification problem, we rely on the following key assumptions:

1. **Low health expenditures:** We assume that households in the lowest medical expenditure group ($d = 1$) who forgo treatment face health transitions comparable to those observed

Table 26: Excess Cost Growth By Insurance Provider

Year	Private Insurance	Medicaid	Medicare
2025	2.0	1.4	1.0
2030	1.9	1.6	1.2
2035	1.7	1.4	1.1
2040	1.4	1.3	1.1
2045	1.2	1.1	1.0
2050	1.0	1.0	1.0
2055	1.0	1.0	1.0
2025-55	56.3	47.4	39.4

Notes: Excess cost growth after 2055 is assumed to be 1 percent until it disappears starting in 2080. The 2025-55 total shows the cumulative excess cost growth over that period.
Source: Congressional Budget Office 2018

among the uninsured population.

2. **Expenditure proxy for severity:** Medical expenditures serve as a signal for unobserved heterogeneity within health states. We assume that the magnitude of the expenditure shock (d) is positively correlated with the severity of the underlying health condition.

3. **The marginal cost of forgone care:** The probability of transitioning to a worse health state is an increasing function of the "avoided" expenditure shock. Consequently, forgoing a larger required medical expenditure results in a higher likelihood of a negative health transition.

To implement these assumptions, we calibrate the transition probabilities for forgoing care, $\pi_h[h'|j,h,d,0]$, by reweighting the treatment transitions, $\pi_h[h'|j,h,d,1]$, using the following weighting function:

$$\pi_h[h'|j,h,d,0] = \frac{\omega^{hp}(h,h',j,m)\pi_h[h'|j,h,d,1]}{\sum_{h'}\left(\omega^{hp}(h,h',j,m)\pi_h[h'|j,h,d,1]\right)} \tag{B.67}$$

where the weights are:

$$\omega^{hp}(h,h',j,m) = \gamma^{ht}(h,h')^{\delta^{ht}(h)\frac{m_{T_0}(j,h,d)}{m_{T_0}(j,h,1)}}, \quad \forall h,h' \in H \tag{B.68}$$

Table 27: Estimates of $\gamma^{ht}(h, h')$

h/h'	h'_1	h'_2	h'_3	h'_4	h'_5
h_1	0.87303	0.88084	0.94511	1.0429	1.25811
h_2	0.9368	0.94202	0.96556	1.01205	1.14356
h_3	0.97089	0.97611	0.99078	1.01509	1.04714
h_4	0.98956	0.99478	1.0	1.00522	1.01044
h_5	0.0	0.0	0.0	0.0	1.0

In this specification, the $h^{max} \times h^{max}$ matrix γ^{ht} determines the direction of redistribution (e.g., from healthy to sick), while the parameters $\delta^{ht}(h)$, in conjunction with the ratio of medical expenditures $\tilde{m}(j, h, d)/\tilde{m}(j, h, 1)$, determine the intensity of that redistribution.

We identify $\gamma^{ht}(h, h')$ and $\delta^{ht}(h)$ through a multi-step calibration. First, we estimate $\pi_h[h'|j, h, d, 1]$ for the insured and a separate transition profile, $\pi_h^{unins}[h'|j, h, d, 1]$, for the uninsured. We then estimate $\gamma^h(h, h')$ to minimize the sum of square residuals:

$$\sum(\pi_h[h'|j, h, 1, 0] - \pi_h^{unins}[h'|j, h, 1, 1])^2$$

We subject this minimization to the following structural restrictions:

- **No spontaneous recovery:** $\gamma^{ht}(h, h'') \geq \gamma^{ht}(h, h')\ \forall h'' > h'$. This ensures that forgoing care never increases the probability of transitioning into a better health state relative to sicker states.

- **Critical shocks:** For households in the second worse health state who receive the highest expenditure shocks, forgoing care results in a transition to the worst health state with certainty: $\pi_h[h^{max}|j, h^{max-1}, 1, 0] = 1$.

- **Absorbing state:** Households in health state h^{max} who forgo care remain in state h^{max} with certainty.

Empirically, the uninsured spend roughly 60 percent less on medical care than the insured.[30] We set $\delta^{ht}(h)$ to match this ratio for each health state, targeting moments from the MEPS data.

30. In fact, based on data from the National Health Interview Survey, the uninsured are more than four times as

Table 28: Calibration of the ratio of medical spending of the uninsured to the insured

	Health State					
	1	2	3	4	5	Total
Model	48	45	36	33	37	39
Data	49	44	37	34	38	42
$\delta(h)$	6.79	7.09	6.81	14.30	1.00	
ζ						2.70

Notes: Model and Data for the medical spending ratio of the uninsured to the insured expressed as percentages. The parameters $\delta(h)$ depend on health state h, while the parameter ζ is the same across all health states. Those parameters are set such that the spending ratios of the uninsured to the insured in the model match those in the data. *Source:* Data from the Medical Expenditure Panel Survey for working-age households 2012-2022.

Because the model fails to converge during joint estimation, we fix δ^h to estimate γ^h, then iterate only over δ^h. Finally, we set the parameter $\zeta = 2.84$, the scale value of death in the value function, to match the average ratio of medical spending between uninsured and insured households aged 26 to 64 in poor health ($h = 5$).

Table 27 shows the resulting estimates for γ^{ht}, and Table 28 shows that our model matches the data for the ratio of medical spending of the uninsured to the insured in each health state very well.

We assume that households' health investment choices do not affect their survival probabilities directly, only through their effects of moving to a worse health state. However, a bigger likelihood of transitioning into worse health also decreases households' life expectancy, because those in lower health states have lower survival rates.

Table 31 shows that our approach generates a distribution of health status in the population that matches the data well, particularly for working-age households.

Our approach to modeling health departs from the seminal human capital framework of Grossman:1972. While Grossman treats health as a capital stock that depreciates deterministically and grows through discretionary investment, we model health as a stochastic process governed by a

likely as the insured to forgo or delay needed medical treatment as a result of costs. See NationalCenterforHealthStatistics:2018.

household's current health state, age, and the decision to obtain medical treatment. This distinction allows our framework to capture the fundamental uncertainty inherent in health shocks that a deterministic capital-accumulation model may overlook. In this setting, households that forgo medical treatment face a higher probability of transitioning into sicker health states in subsequent periods. These transitions carry significant economic consequences, including increased mortality risk, diminished labor productivity through lower wages, and higher future medical liabilities.

Our framework offers several distinct advantages over the Grossman model by grounding the simulation in empirical health states. We match health states directly to observed health expenditures, transition probabilities, and mortality rates, whereas the Grossman model often treats health spending as discretionary and frequently fails to link that spending directly to mortality outcomes. By tying health transitions to mortality, we avoid the pitfall noted by Ehrlich.Chuma:1990, who observe that the Grossman model struggles to generate realistic life-cycle mortality patterns because households can theoretically invest indefinitely to forestall death. Our model ensures that biological aging and stochastic shocks eventually dominate, regardless of investment levels. Furthermore, we relax the direct coupling between spending and health improvement. While the basic Grossman model implies that higher spending necessarily yields better health, empirical evidence from Bajari.etal:2014, Khwaja:2010, and Card.etal:2009 suggests a more subtle relationship. Our framework acknowledges that medical spending often addresses the severity of a shock rather than simply purchasing health increments, thereby forcing households to face realistic trade-offs between current consumption and health maintenance.

The theoretical and empirical limitations of treating health strictly as a depreciating asset inform our shift toward a stochastic specification. Ehrlich.Chuma:1990 highlight that Grossman's assumption of a constant-returns-to-scale health production function introduces indeterminacy regarding optimal investment choices, which defies a systematic resolution of the choice of both optimal health paths and longevity. Similarly, Hall.Jones:2007 and Grossman:2000 discuss the difficulty of reconciling the capital-stock approach with observed longevity choices. In the basic Grossman model, population health also relates directly to out-of-pocket expenses; if these were set to zero, households would theoretically consume all available healthcare services to maximize their health stock. Empirically, the implications of the Grossman model have yielded mixed results at best, as noted in the micro- and macro-based reviews by Zweifel:2012 and Hartwig.etal:2018. By shifting the focus to stochastic transitions conditioned on treatment, we provide a framework more consistent with the observed volatility in medical expenditures and the persistence of health states.

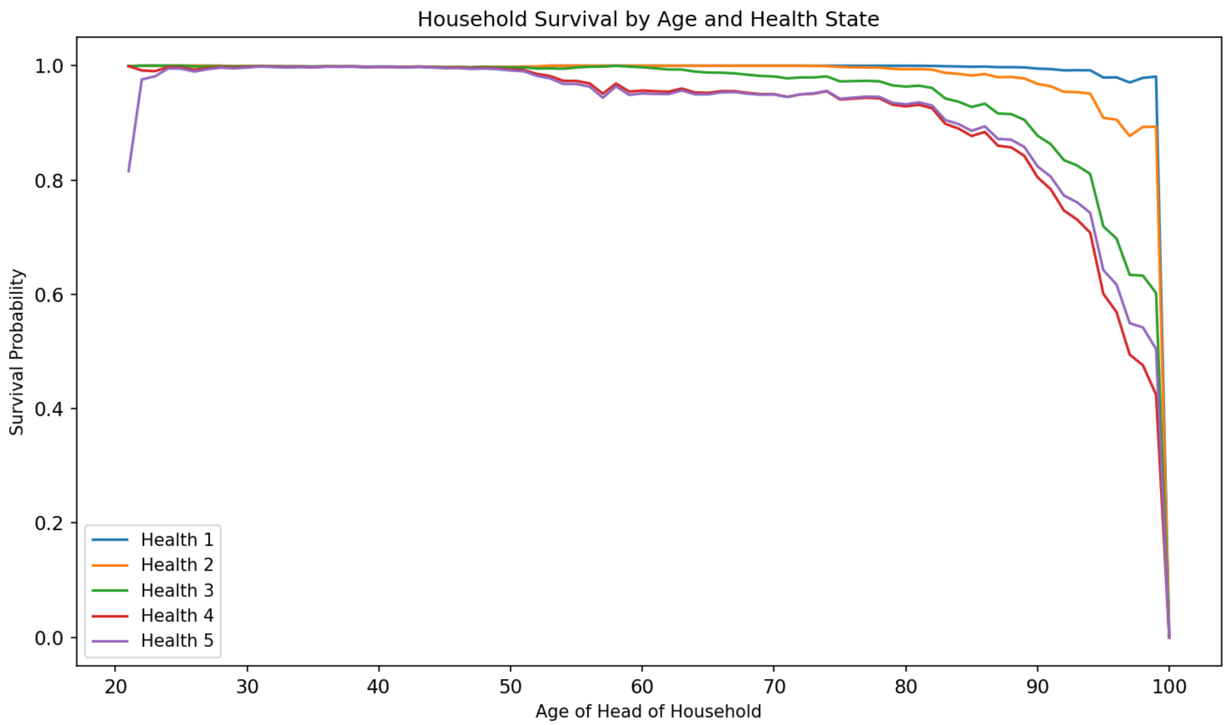


Figure 22: Survival Probabilities by Current Health State and Age for 2024 initial steady state

B.6.5 Administrative and Overhead Spending

Health insurance administrative and overhead spending varies between payers and typically includes spending on advertising, utilization management (including programs to combat fraud and abuse), claims and payment processing, profits, and taxes and fees.[31] Including an estimate for administrative spending is important, because it directly increases private health insurance premiums and the cost of providing public health insurance programs.

Private health insurance companies spend the most on administrative and overhead costs, with estimates ranging from 11 percent in the large-group market, 16 percent in the small-group market, to 20 percent in the nongroup market.[32] In a more recent study, CBO estimated total administrative expenses of ESI to be 12 percent, which is the value we use for all private insurance.[33]

Medicare generally has the lowest administrative costs. According to the Medicare Trustees, Medicare fee-for-service (FFS) spends about 1.4 percent on administrative costs,[34] while CBO estimates that number to be 2 percent. Including Medicare Advantage (Medicare Part C) and Medicare Part D plans, which are operated by private insurance companies, CBO estimates that total Medicare administrative costs are 8 percent. Data reported in the NHEA suggest that the total administrative costs for Medicare are about 7.6 percent in 2019. Administrative costs of Medicaid are roughly in line with those of total Medicare. CBO estimates that administrative spending on Medicaid and Chip is 8 percent, while NHEA data suggests that it is 12.4 percent.[35] We use CBO's estimates for administrative spending, which are summarized in Table 29.[36]

31. Note that these estimate ignore large administrative costs born by providers. For a more complete account of administrative costs, see, for example, Woolhandler, Campbell, and Himmelstein 2003 and Woolhandler and Himmelstein 2017.

32. See Congressional Budget Office 2016.

33. See Congressional Budget Office 2020a. That number is also consistent with the 11 percent administrative spending in private health insurance reported in the National Health Expenditure Account (NHEA). See National Health Expenditure Accounts, https://www.cms.gov/files/zip/national-health-expenditures-type-service-and-source-funds-cy-1960-2019.zip

34. See The Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds 2020.

35. The Centers for Medicare & Medicaid Services report average Medicaid administrative costs for 2018 of 4.6 percent, while the NHEA estimates Medicaid administrative costs are 12.2 percent without CHIP. See, https://www.medicaid.gov/state-overviews/scorecard/annual-medicaid-chip-expenditures/index.html (accessed October 28, 2021).

36. Recent work by Scheinker et al. 2021 find that billing and insurance related costs in the U.S. could be decreased by between 30 and 60 percent.

Table 29: Administrative Expenses by Insurer

Insurer	CBO	NHEA
Private Insurance	12.0	12.0
Medicare fee-for-service	2.0	
Total Medicare	8.0	6.9
Medicaid	8.0	10.4

Sources: Congressional Budget Office (2020a) and 2024 data from the National Health Expenditure Account.
Notes: Data expressed as a percent of totla health expenditures. Total Medicare includes the fee-for-service program, Medicare Advantage, and Medicare Part D plans. Medicaid includes Mediciad and CHIP.

B.6.6 Medicare and Medicaid Premiums

While Medicare Part A is partially financed through OASI HI payroll taxes, Medicare Part B (medical insurance) and Part D (prescription drug coverage) are partially financed through income-based premiums. We assume that all retirees enroll in Medicare Parts B and D, which is largely consistent with the data and require households to pay the income-based premiums for those programs.[37] Premiums for Medicare Part B are paid to the government. Total Medicare Part D premiums consist of two parts: a plan premium that is paid directly to the chosen prescription drug insurance plan, and an income based portion that is paid to the federal government. We use the 2020 statutory premiums, which we show in Table 30.

Consistent with current law, Medicaid does not charge insurance premiums.[38]

Our model of health transition and survival generates an aggregate population distribution across health states that matches the data quite well, as Table 31 shows.

37. In 2023, the large majority of Medicare beneficiaries were enrolled in Medicare Parts B (91 percent) and D (78 percent). See CMS Program Statistics, Tab MDCR ENROLL AB 3_CPS_02ENR and Tab MDCR UTLZN D 2_CPS_12UPD.

38. Under current law, the federal government generally prohibits states from charging premiums to Medicaid enrollees with income below 150 percent of the federal poverty line and limits the extend to which states can charge premiums and co-payments. See Guth, Ammula, and Hinton 2021.

Table 30: 2024 Annual Medicare Part B and D Premiums

Income Threshold	Part B	Part D
$y_t^T \leq$ \$103,000	\$2,096.40	\$0.00
\$103,000$<$ y_t^T $\leq$\$129,000	\$2,935.20	\$154.80
\$129,000$<$ y_t^T $\leq$\$161,000	\$4,192.80	\$399.60
\$161,000$<$ y_t^T $\leq$\$193,000	\$5,450.40	\$645.60
\$193,000$<$ y_t^T $\leq$\$500,000	\$6,708.00	\$890.40
y_t^T $>$\$500,000	\$7,128.00	\$972.00

Notes: y_t^T is taxable income. Amounts are per person. Income threshold is for individual tax filers. The average Part D plan premium in 2020 was \$42.0/month × 12 = \$504.00/year.
Sources: Medicare.gov, available at https://www.medicare.gov/your-medicare-costs/part-b-costs. Archived 2024 data available in the Federal Register (for Plan B premiums) and https://www.medicare.gov/drug-coverage-part-d/costs-for-medicare-drug-coverage/monthly-premium-for-drug-plans (for Part D premiums); Average Plan D premium available at Kaiser Family Foundation, available at https://www.kff.org/report-section/medicare-part-d-a-first-look-at-prescription-drug-plans-in-2020-issue-brief/

Table 31: Distribution by Age and Health State

	Health State									
	1	2	3	4	5	1	2	3	4	5
	Working Age (21-64)					**Retirees (65+)**				
Model	24	36	29	9	2	13	34	33	16	5
Data	25	35	27	10	3	18	31	31	15	5

Notes: The values represent the percentage distribution across health states within each age group.
Source: 2024 MEPS data.

B.6.7 Health Insurance Enrollment

Our calibration strategy produces a distribution of health insurance enrollment, both in aggregate and by age group, that is consistent with data from the 2025 Current Population Survey (CPS). Table 32 shows that our model matches somewhat well the data for the fraction that is uninsured and enrolled in Medicaid. Our model matches the age distribution of those receiving ACA premium support slightly less well, especially for the group of 21-25 year olds, who under current law are

Table 32: Health Insurance Enrollment by Age Group

	Uninsured		Medicaid		ACA	
	Model	Data	Model	Data	Model	Data
21-25	14.3	14.1	18.2	18.3	13.1	3.2
26-34	12.1	12.5	17.4	17.6	0.7	3.4
35-44	12.3	11.6	15.5	15.4	1.0	3.5
45-54	9.0	9.2	13.2	13.2	4.4	2.9
55-64	8.0	7.9	15.1	15.0	8.5	4.2
21-64	11.0	1.1	15.8	7.6	4.9	4.8
26-64	10.4	0.5	15.3	6.2	3.5	6.8

Notes: Data show percentages of the population.
Source: Data for 2024 from 2025 CPS. Model data from 2024 initial steady state.

able to stay on their parent's insurance plan, a feature we do not model. However, we still match the aggregate fraction of households receiving premium subsidies and the general trend by age.

We first assume that everybody who qualifies for Medicare and Medicaid enrolls in those programs. We then set a schedule, by age, of assets that are excluded from Medicaid countable assets to calibrate to the Medicaid enrollment by age group we observe in the data. Next we choose an earnings level above which households are offered ESI to match the aggregate uninsurance rate for working age households and set the age-based weights ω_j^{HI} from Equation A.22 such that the uninsurance rate by age group in the model matches the data. Finally, we choose cut-off earnings for below which households qualify for premium subsidies in the ACA in order to match data on ACA enrollment of working age households.[39]

All retirees are assumed to enroll in Medicare Parts A, B, and D and pay the associated Medicare premiums as shown in Table 30.

Households may qualify for Medicaid categorically if they have low income and low assets. Those with low income but larger assets may qualify medically if their healthcare expenses exceed

39. For our calibration, we target the population of 26 to 64 year olds, because the ACA introduced a rule by which young adults up to age 26 may receive health insurance coverage through their parent's ESI. That makes it difficult to exactly match the enrollment data for the 21-25 year old group. However, we do report enrollment numbers for that age group.

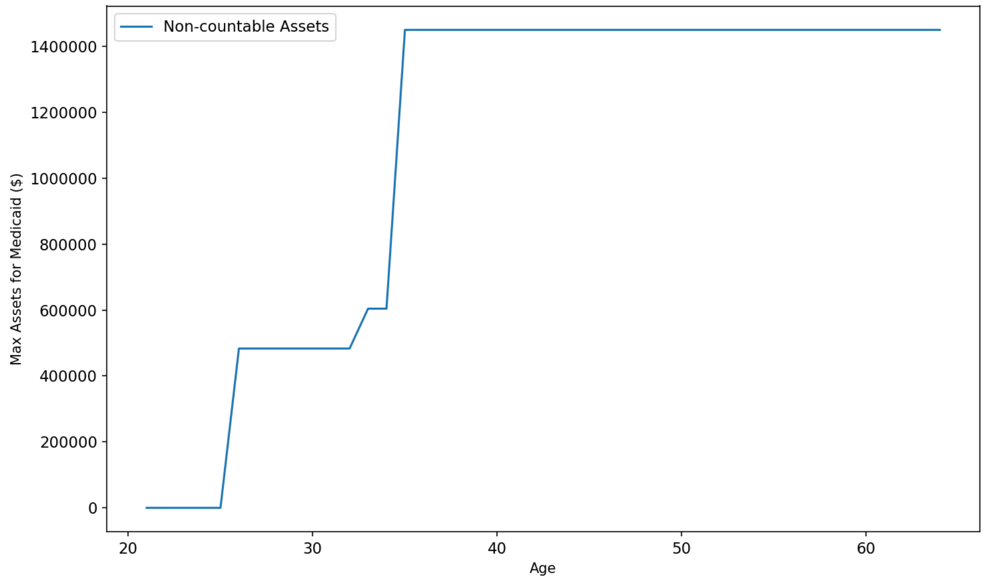


Figure 23: Assets that are excluded from the Medicaid asset test in Model for 2024 initial steady state

their total resources.[40] To qualify households must earn less than 138 percent of the federal poverty level (FPL) and own no more than $2,000 in assets, excluding non-countable assets such as a primary residence, a car, personal effects, household goods and furnishings. We adjust the amount of non-countable assets by age to match the Medicaid enrollment rates in the data by age group (see Figure 23 for those parameters). As Table 32 shows, this approach generates Medicaid enrollment that is consistent with the data but does tend to over-estimate Medicare enrollment. In our model, 21.3 percent of 21 to 64 year olds are enrolled in Medicaid compared to 15.7 percent in the data.

We calibrate our model to match an uninsured rate of the 21 to 64 year old population of 10.8 percent (see Table 32). That requires setting the income threshold above which households are offered ESI to $\underline{y}^{ESI} = \$16,446$ and results in an uninsured rate of 8.9 percent in the model.

We calibrate the employer's share of ESI insurance premiums, $\eta^{ESI} = 0.7749$.

40. Medicaid eligibility requirements are too complex to model in detail here. Our approach captures the idea that Medicaid insured the low-income households with the greatest need. Mitchell et al. 2021 provide a detail description of the Medicaid program, including a comprehensive discussion of the many different eligibility pathways. According to Congressional Budget Office 2020b, roughly 17 percent of the uninsured in 2019 could enroll in Medicaid but do not. The Congressional Budget Office cites unawareness of households' eligibility or the complexity of the enrollment process that may prevent the eligible uninsured from applying for new coverage or renew their coverage. In addition, the Congressional Budget Office notes, recent immigrants may be discouraged from applying for Medicaid coverage for their citizen children because they fear it could prevent them from becoming permanent legal residents. In the model, relatively high wealth may also proxy for an unwillingness to endure the complexity of the enrollment process.

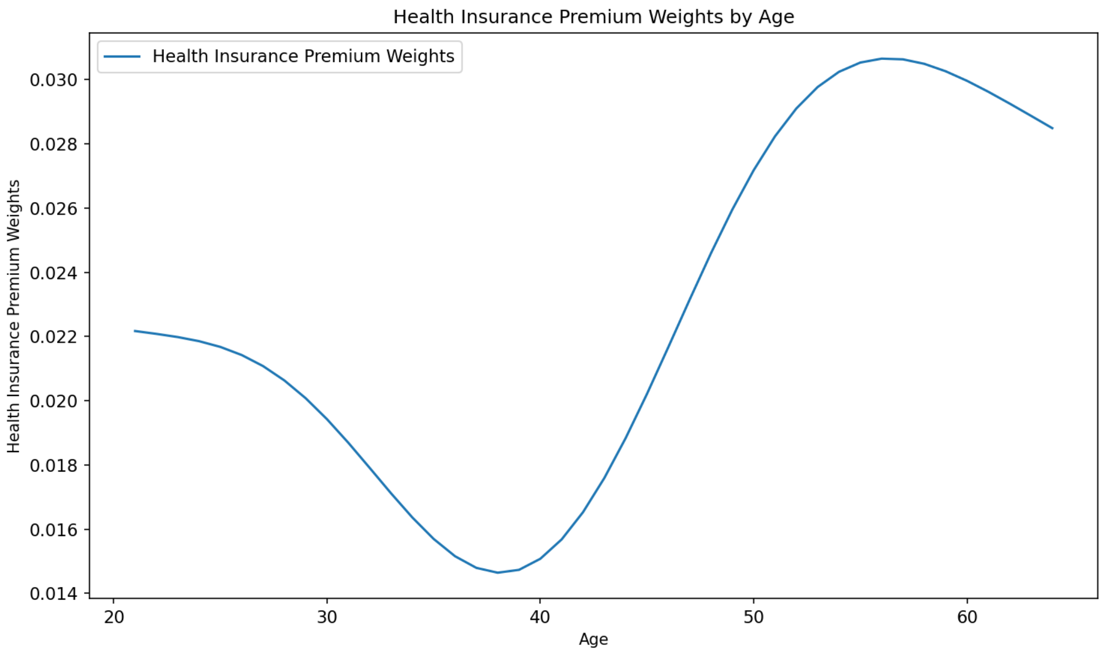


Figure 24: Weights used in calculating health insurance premiums by age

To calibrate the model to match the fraction of households without health insurance by age group, we set the weight ω_j^{HI} from Equation A.22 by minimizing the sum of squares residuals of the difference between the rates in the data and the model. Figure 24 shows the resulting weights ω_j^{HI}. Table 32 shows that the model performs well in matching the data by age groups. The calibration requires that those age 60 to 64 pay ESI premiums that are about double than those paid by 21 to 25 year olds.[41]

Under 2024 law, the Required Contribution Percentage ranged from 0 percent of income for those below 133 percent of the federal poverty line (FPL), to 8.5 percent of income for those who make over 400 percent of the FPL. Under current law, for 2026 and beyond, households earning more than 400 percent of the FPL are ineligible for premium tax credits.[42] We set required contribution percentages in the initial steady state according to 2024 law. We calculate the ACA premium subsidy $\pi_{ACA,t}^{PremSub}(y_t^T) = \max\{\varsigma p_{p,t} - RCP(y_t^T)y_t^T, 0\}$. ς is the average marketplace benchmark premium in 2024 of \$540 per month divided by the average annual single premium per

41. Age-dependent insurance premiums can be rationalized with firms offering a variety of health plans and households of different age groups choosing plans with different benefits and premiums. In 2021, 62 percent of workers whose employer offered health insurance had two or more plans to choose from. See Kaiser Family Foundation 2021.
42. For more information, see https://tinyurl.com/2fewkerd .

enrolled employee in 2024 of $587 or 0.92.[43]

We calibrate the model to an ACA marketplace enrollment rate of 3.5 percent for 26 to 64 year olds by setting the maximum labor income (aca_{cutoff}) to be eligible for the ACA market at $16,446.41. Enrollment is limited to households that do not qualify for Medicaid and have incomes below below the cutoff. As Table 32 shows, we are able to match the aggregate for the 26-64 year old population pretty closely, but we are less successful at matching enrollment by age group, which we do not target. The largest differences are for the group of 21-25 year olds. In the data, that group is likely enrolled in their parents' health plan, or some other group plan (such as through their college), which we do not model.

B.6.8 Health Investment

Table 33 shows compares data on the fraction of the population that forgoes or delays medical treatment.[44] In the model 4.3 percent of the population forgoes medical treatment compared to 12.1 percent in the data.[45] Our model also captures that uninsured households are much more likely to forgo medical treatment than are insured households, although the model overshoots slightly. In the model 64.8 percent of uninsured forgo medical treatment compared to 44.4 percent in the data.[46] Those results are particularly striking given that we did not attempt to calibrate to data on forgone medical spending. It is the result of our overall calibration strategy and suggests that our model captures households' incentives to insure and seek medical treatment that are consistent with the data.

43. See https://tinyurl.com/mryzvk3w.

44. While there might be a difference between whether somebody delays or forgoes medical treatment, our model cannot capture that nuance as it runs at an annual frequency.

45. Radley et al. 2021 document large geographic, racial, and ethnic differences in forgone care due to cost. The proportion of white people who cited cost as a barrier to receive care ranged from 6 percent in the District of Columbia and Hawaii to 14 percent in Georgia, Oklahoma, Alabama, and Mississippi. For Black and Latinx people, state rates vary between 9 percent in California (Black) to a high of 30 percent in Tennessee (Latinx).

46. Collins, Gunja, and Aboulafia 2020 document that under-insured and uninsured are 2 to 3.5 times more likely to forgo needed medical care and medications and that 56 percent of the uninsured avoided needed medical care.

Table 33: Forgone and Delayed Medical Care Due to Cost

	Model	NHIS Data
All	6.4%	10.1%
Uninsured	57.7%	24.9%

Notes: Data are for 21- to 64-year-old individuals.
Sources: Data from the 2024 IPUMS National Health Interview Survey. See Blewett et al. (2025).

Table 34: Life Expectancy, Model versus Data

	Life Expectancy at Age	
	21	65
Model Benchmark Economy	61.4	22.0
Census Bureau	59.9	20.1
Centers for Disease Control and Prevention	58.3	19.5
Congressional Budget Office		19.7
Social Security Administration	57.3	18.9

Sources: Census Bureau data for 2020 from https://www.census.gov/content/dam/Census/library/publications/2020/demo/p25-1145-supplemental-tables.pdf ; CDC data for 2023 from https://www.cdc.gov/nchs/data/nvsr/nvsr74/nvsr74-06.pdf, Table 1; Congressional Budget Office projections for 2025 from https://www.cbo.gov/publication/61164, Table A-1; SSA data for 2022 as of 2025 from https://www.ssa.gov/oact/STATS/table4c6.html

B.6.9 Life Expectancy

Our model projects life expectancies that are generally consistent with US government projections as shown in Table 34. Our model projects that 21-year-olds can expect to live 60.9 years, on average, compared to 59.9 years based on projects by the Census Bureau, 58.3 years by the CDC, and 57.3 years by the SSA. Our projections of the life expectancy for 65-year-olds is 21.5 years, compared to a range of 18.9 to 20.1 years by the Census Bureau, the CDC, the CBO, and the SSA.

B.6.10 Aggregate Health Spending

Table 35 shows total health spending by payer as a share of GDP and compares those to 2023 data from the National Health Expenditure Account (NHEA).[47] Overall, the numbers match up very closely, even though none of them were calibration targets. Total health spending is XXX percent of GDP in the model and 14.2 percent in the NHEA data. Medicaid is lower than in the data, which is largely due to our model not capturing long-term support services. The model generates average private insurance premiums of $11,150, which is somewhat more than the $8,951 that the Kaiser Family Foundation reports for 2024.

Table 35: Aggregate Health Spending as a Share of GDP

	Model	NHEA Data
Private insurance	4.6	5.0
Medicare	3.6	3.6
Medicaid	1.4	1.9
Out-of-pocket	1.9	1.9
Total administration	0.9	1.0
Total health expenses	12.0	13.4
Average private health insurance premium	$9,161	$8,486

Notes: Private health insurance premiums are in US$.
Sources: Spending data for 2024 from the National Health Expenditure Account. Data is for the ”Personal Health Care” and the ”Total Administration and Total Non-Medical Insurance Expenditures” categories. It includes data for Private Health Insurance, Medicare, and the federal portion of Medicaid. It excludes spending on Public Health Activity, Investments in Research, and Structures and Equipment. Data for the 2024 average annual single premium per enrolled employee for employer-based health insurance from the Kaiser Family Foundation. GDP data: https://fred.stlouisfed.org/release/tables?rid=53&eid=41047&od=# NHEA data: https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/historical , https://www.cms.gov/files/zip/national-health-expenditures-type-service-source-funds-cy-1960-2024.zip . Insurance Premiums: https://www.kff.org/private-insurance/state-indicator/single-coverage/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D.

47. We compare our data to the Personal Health Care and Total Administration and Total Net Cost of Health Insurance Expenditures categories of the 2023 National Health Expenditures. Those two categories make up more than 90 percent of the 2023 National Health Expenditures. The remainder includes Public Health Activities and Investments in Research, Structures and Equipment, which we do not explicitly model.

Table 36: Values of Government Policy Parameters in the Benchmark Economy

Parameters		Value	Comment
Income taxes			
Taxable labor income ratio (change to Payroll Tax Base Share)	η_1^y	0.8641	T_P / GDP = 6.01%
Payroll Base Deduction	η_0^y	$6000	?
Social Security system			
Medicare pay roll tax rate: HI	τ_{HI}	0.029	
: HI surtax	τ_{HI2}	0.009	
HI surtax threshold	ϑ_{HI}	1.9053	0.4 x $200,000 + 0.6 x $250,000

B.7 The Social Security and Hospital Insurance System

The Social Security's Old-Age and Survivors Insurance and Medicare's Hospital Insurance (OASHI) system closely mirrors current U.S. law. The payroll tax function includes social security payroll taxes that are levied on taxable earnings up to the statutory maximum, HI payroll taxes levied on all taxable income and a HI surtax levied on income above $200,000 for single filers and $250,000 for joint filers. The OASHI primary insurance amount is calculated based on a household's average historical earnings. We report the parameters of the OASHI system in Table 36.

We define the OASHI payroll tax function to be

$$\tau_{P,t}(y_t^T) = \tau_{O,t}\min(\eta_1^y(y_t^T - \eta_0^y), \vartheta_{\max})$$
$$+ \tau_{HI,t}\,\eta_1^y(y_t^T - \eta_0^y) + \tau_{HI2}\max(\eta_1^y(y_t^T - \eta_0^y) - \vartheta_{HI}, 0),$$

where $\tau_{O,t}$ is the flat Old-Age and Survivors Insurance tax rate, $\tau_{HI,t}$ is a Hospital Insurance (Part A of Medicare) tax rate, and τ_{HI2} is an HI surtax rate for households with high labor income (covered earnings greater than $200,000 for single taxpayers and $250,000 for married couples filing jointly). The first three tax rates include the portion paid by employers. When labor income is below the threshold of maximum taxable earnings, the statutory OASDI tax rate is 12.4 percent, including 6.2 percent paid by employers; thus, $\tau_{O,t} = 0.124$. The effective HI tax and HI surtax rates are set at $\tau_{HI,t} = 0.029$, and $\tau_{HI2} = 0.009$, respectively. The maximum taxable earnings per worker for the OASDI payroll tax were $128,400 in 2018 (**SocialSecurityAdministration:2019a**). Because the model is based on household units, that single-worker figure must be translated to the household level. We assume that 60 percent of households are married households, of which two-thirds are two-earner households—meaning that 40 percent of all households are two-earner households.

B.8 Production Technology and Wages

In our baseline model specification, we assume a single intermediate goods sector ($n = 1, \zeta_1 = 1$). We currently keep α_1^f fixed across time and set it to 0.45 based on Microsim calculations. In the section below, the subscript b denotes the business-type (corporate or pass-through), j denotes the business sector (the type of output, set to 1 in the default case of a single intermediate-goods sector), and t is the year.

B.8.1 Business income

U.S. business taxation has two main treatments: corporate and pass-through. Corporations are business entities which are subject to taxation at the entity level while pass-throughs are business entities (such as partnerships, some limited-liability companies, subchapter S corporations, and sole proprietorships) for which tax liability is shifted onto the pass-through entity's owners. All business tax parameters are generated from from historical data and projections in the PWBM tax simulator. The choice between investing in pass-through versus corporate business is not modeled in the OLG model, which takes these decisions as exogenous, and all households hold the same proportion of pass-through and corporate assets.

All of the main parameters in the dynamic OLG model for business taxation come from either historical data or projections contained in the PWBM tax simulator. The parameter $\zeta_{b,1,t}^{ded}$ is the share of firm GDP which is the base for other tax deductions. It is calculated as the residual that equalizes OLG-basis taxable income and conventional-basis taxable income for each business type (corporate or pass-through) b. For the 2024 initial steady state it is set to 0.084 for the corporate sector and 0.213 for the pass-through sector. The parameter $\zeta^{cred}_{b,1,t}$ denotes the share of production that is subject to the production tax credit. For 2024 initial steady state the share is set to 0.0078 for the corporate sector and 0 for pass-throughs. $\psi_{b,1,t}$, the share of investment credits is set to zero for all firms. The parameter $\zeta^{other}_{b,1,t}$ is other business expenses is set to 0.091 and 0.082 for corporate and pass-through firms, respectively, and the economic depreciation rate of capital $\delta_{1,t} = 0.056$; these values are all kept constant across time in the baseline. The share of borrowing or interest expenses that are eligible to be deducted is $\phi^{int}_{b,j,t}$, and it is the same for both pass-throughs and corporations. We set it to 0.915 in the 2024 initial steady state.

We currently exogenously set interest rate that the firm pays on business debt is $\rho_{corp,1,t} = 7.0$ percent. We assume that all deductions, $\phi_{b,j,t,i}$ occur over a $T_f = 40$ year schedule that comes from the PWBM tax simulator. The tax deduction schedule is computed separately for each sector and business-type and is based on approximating the capital mixes in the economy and the various depreciation schedules in the tax code that are applied to different types of capital.

In the current model, we take $\nu_1 = 3.5$ following Glover, Gomes, and Yaron (2015) and the leverage ratio $B_{corp}/K_{corp} = 0.32$ from Barro and Furman (2018) and Graham, Leary, and Roberts (2015). Using these values, we derive (and assuming a time invariant statutory tax rate, $\tau^{statutory} = 0.21$) $\nu_2 = 1.61$ from (A.33),

$$\nu_2 = \left(\frac{\tau^{statutory} \phi^{int}_{corp} \rho_{corp,1,t}}{1 - \tau^{statutory} \phi^{int}_{corp}} \right)^{-\frac{1}{\nu_1 - 1}} \frac{B_{corp}}{K_{corp}} \tag{B.69}$$

These time-invariant values of ν_1 and ν_2 are used for all policy runs under the assumption that the leverage cost function depends only on the leverage ratio.

Given α^f_1 and a targeted capital to labor ratio of 3, we normalize $A_{1,t} = 0.845$. In the initial steady state this implies wages $w = 1$. η, the adjustment cost of capital parameter is assumed to be 1.

B.9 Transfers

B.9.1 OASI Benefits

One of the key features of the dynamic OLG model is its model of OASI benefits, $tr_{OASI,t}(j, b, p^t_{t-j})$. The OASI benefit is a time-varying function of age (j); the average number of adults in the household ($\Omega_{a,T_0}(j)$), which itself is a function of age; and the average real wage in the economy in the year the household turns 60 years of age ($w_{t(j=60)}$), itself an element of the set of historical prices ($p^{T_F}_{T_0}$).

The schedule of OASI benefits $tr_{OASI,t}(j, b, p^t_{t-j})$ is a piecewise linear function with Z vertices (including 0), which we call primary insurance amount (PIA) bend points. Specifically, for a household with $b \geq 0$ earnings history with PIA bend points $PIA_{z,j,t}$ and PIA income replacement rates $PIA^r_{z,j,t}$ for age j and time t, the OASI benefit is defined as:

$$tr_{OASI,t}(j, b, p^t_{t-j}) = \sum_{z=1}^{Z} max[0, min(b - PIA_{z,j,t,p^t_{t-j}}, PIA_{z+1,j,t,p^t_{t-j}} - PIA_{z,j,t,p^t_{t-j}})]PIA^r_{z,j,t}\Omega_{a,t}(j)$$

where $PIA_{Z+1,j,t,p^t_{t-j}}$ is not an actual vertex on the piecewise linear line, but rather set to be higher than the largest value on the support of b. In the baseline and any policy function, We multiply these rates by the number of adults ($\Omega_{a,t}(j)$) to approximate the total household OASI benefit. Under current law, there are four PIA bend points (including 0) with corresponding PIA income replacement rates of 0.9, 0.32, 0.15, and 0.0.

The primary insurance amount (PIA) parameters—both bend points ($PIA_{z,j,t,p^t_{t-j}}$) and PIA income replacement rates ($PIA^r_{z,j,t}$)—are drawn from the PWBM OASI calculator, which provides this information in each year and for each cohort. The OASI calculator provides these values based on current and forecast economic projections; the forecast projections underlying these values may differ from the projections in the dynamic OLG model. In particular, the average wage index (AWI)—both real and nominal—may differ from the projections used to construct the PIA brackets, especially as economic conditions change in response to changes in other economic assumptions or changes in policies.

Therefore, we construct the PIA brackets used in the dynamic OLG model for each bracket z, age j, and year t

$$PIA_{z,j,t,p_{t-j}^{t}} = \frac{PIA_{z,j,t}^{OASI\_calc}}{AWI_{t(j=60)}^{OASI\_calc}} w_{t(j=60)}$$

where $PIA_{z,j,t}^{OASI\_calc}$ is the PIA bend point provided by the OASI calculator, $AWI_{t(j=60)}^{OASI\_calc}$ is the nominal AWI series used by the OASI calcualator, and $w_{t(j=60)}$ is the real average wage (wage income divided by hours worked) used by the dynamic OLG model. The real average wage, $w_{t(j=60)}$, is endogenous, so as policy changes the average wages for households at age 60, the OASI benefits change as well.

B.9.2 Supplemental Nutrition Assistance Program (SNAP)

This transfer program has been design to help low income households achieve a nutritionally adequate diet.

Eligibility Requirements Eligibility is determined at the household level. Legislated SNAP requirements are based on household size. To accommodate the non-integer family sizes we have in the model, we approximate the legislated eligibility rules linear.

1. Gross income limit (GIL): $GIL_t(j) = G\bar{I}L_t + G\hat{I}L_t(\Omega_{a,t}(j) + \Omega_{c,t}(j))$. For 2024, $G\bar{I}L_t = \$11,286$ and $G\hat{I}L_t = \$6143$.

2. Net income limit (NIL): $NIL_t(j) = N\bar{I}L_t + G\hat{I}L_t(\Omega_{a,t}(j) + \Omega_{c,t}(j))$. For 2024, $G\bar{I}L_t = \$8,689$ and $G\hat{I}L_t = \$4719$.

3. Assets: For 2024, households must have \$2,861 or less in assets, a. If the head of the household age is at least 60 years old, the limit goes up to \$4,422.

For households over 59 years old, the only eligibility requirement is the asset requirement; the income limits do not apply. More generally, eligibility and benefit levels depend on net income and gross income as defined by the SNAP rules. Gross income is the sum of paid labor income, dividend income, OASI benefits, SSI benefits and SSDI benefits (which are set to 0 in the baseline economy). Section A income, y^{secA}, is 80 percent of gross income less a standard deduction. In 2024, the standard deduction is \$2,210. Net income, y_t^{NI} is then Section A income less a shelter deduction, $SD_{snap,t}$, where:

$$SD_{snap,t} = \min\{SC_t - 0.5y_t^{secA}, \nVdash_{a<60} SC_t^{MAX}\} \tag{B.70}$$

For 2024, $SC_t = \$13,97$ and $SC_t^{MAX} = \$7,791$.

The maximum benefit that a household can receive depends on its size. To accommodate for average household size, we perform linear interpolation: Benefit limit, $BL_t(j) = \bar{BL}_t + \hat{BL}_t(\Omega_{a,t}(j) + \Omega_{c,t}(j))$. For 2024, $\bar{BL}_t = \$774$ and $\hat{BL}_t = \$2,357$.

Eligible households receive a SNAP benefit $\tilde{SN}_t(s,l) = \max\{0, BL_t - 0.3y_t^{NI}\}$. We top up this benefit if the total resources a household has falls below \$8500. Total resources are defined the net of assets, a, paid income, y_t^P, SNAP benefits $\tilde{SN}_t(s,l)$ and any tax credits for insurance premia less insurance premia paid and tax liabilities. The total transfer the household receives is $SN_t(s,l) = \tilde{SN}_t(s,l) +$ resource shortfall.

B.9.3 Supplemental Security Income (SSI)

This federal assistance program provides cash payments to aged, blind, or disable individuals with limited income and resources. To be eligible in the model, the household must (1) be aged 65 or over; (2) have assets, a, below a limit (\$2,080.80 in 2024).

The household SSI benefit, $SSI_t(s,l)$ is difference between the maximum SSI benefit (\$9718.56 in 2024) and a households "countable income." If paid labor income y_t^P is less than the SSI general deduction (\$244.80 in 2024), the household's countable income is the maximum of the sum of y_t^P,

Table 37: The Government's Budget in the Benchmark Economy

	Model	Data		Model	Data
Revenues			**Outlays**		
Income and corporate taxes	8.6	9.1	Mandatory		
Payroll Taxes	5.7	5.5	Social Security	6.0	5.0
Other	1.0	0.8	Medicare	3.6	3.7
			Medicaid	1.0	2.1
			Other	2.6	3.0
			Subtotal	13.1	13.8
			Discretionary		
			Net interest	2.9	3.0
Total Revenues	**15.2**	**15.5**	**Total Outlays**	**22.3**	**23.0**

Notes: In percent of gross domestic product (GDP). Data for 2024.
Source: Data from Penn Wharton Budget Model

dividend income, OASI benefits and SSDI benefits less the SSI general deduction or \$0. If y_t^P is greater than the SSI general deduction, countable income is the sum of ”countable labor income” and dividend income, OASI benefits and SSDI benefits, where countable labor income takes y_t^P less both the SSI general deduction and the SSI earned income deduction (\$811.51 in 2024) and multiplies the results by 50%. If countable labor income is less than 0 as a result of this calculation, it is set at \$0.

B.10 Government Budget

Most of the previously-described parameters were calibrated so that so that large categories in the the government budget were broadly consistent between the model and the observed U.S. federal budget in the model's starting year. Table 37 in the main text shows the government's budget in the stationary-population benchmark economy and the data.

We set discretionary spending to its value in the data and make adjustments (under the category ”Adjustments”) so that total primary deficits match the data.

We set federal debt so that the distribution of federal debt and the effective coupon rate for each year's vintage of debt matches its empirical distribution of marketable securities (excluding TIPS) in the Monthly Statement of Public Debt at the end of December of the steady-state year. The total amount of federal debt is set so that it matches Debt Held by the Public at the beginning of the year.

B.10.1 Personal Income Taxes

The dynamic OLG model's tax schedule is derived from the PWBM tax simulator. The PWBM tax simulator generates a huge number of tax filings (records) that are then distilled into (1) an effective deduction schedule broken down into 19 brackets and deduction shares; (2) an effective marginal ordinary tax schedule with 16 brackets and rates; and (3) an effective marginal preferred income tax schedule with 5 brackets and rates. All brackets, rates, and burdens are fixed in real values (or fixed percentages) after 2055.

We start with a schedule of deductions. The first step is to define the basis of compensation in the OLG model to which the deduction schedule is applied. In the dynamic OLG model, we apply deductions to a very broad definition of compensation: all taxable compensation that could be taxed at the non-preferred rate. This includes all OASI income, labor income, business income taxed at the ordinary rate, and dividend income from U.S. federal debt. Note that this excludes non-OASI transfers such as SNAP or ACA subsidies. The deduction schedule is meant to approximate all major deductions from taxable total compensation excluding employer-sponsored healthcare premia deductions, which are handled separately in the model.

To find a comparable measure of income in the PWBM household tax simulator to use as our basis for approximating a deduction schedule to be used in the dynamic OLG model, we take adjusted gross income and add back elective deferrals (doubled) and payroll deductions (doubled).[48] The deduction share is computed as the sum of elective deferrals (doubled), above-the-line deductions, and the itemized or standard deduction. We compute deductions as a share of this income in every year across five different age groups (30-, 31-40, 41-50, 51-65, 65+) for 16 income brackets

48. Electrive deferrals and payroll deductions are reported as the household expense only; we double these values to capture the employer contributions to both of these categories.

whose thresholds are determined by every decile, plus the 2.5th, 5th, 7.5th, 95th, 96th, 97th, 98th, 99th, and 99.9th percentiles. The deduction share for each of these bins is equal to a weighted average of the deduction share among all filers in these bins. This schedule is imported and applied in the OLG model to each household based on the compensation measure defined above.

In the dynamic OLG model, after applying this deduction share plus any endogenous healthcare insurance premia deduction that the household may be eligible to receive, we are left with a measure of income that closely approximates income taxed at the ordinary rate, or "ordinary income." The next step is to import a set of marginal tax rates that are approximated from data in the PWBM household tax simulator, also based on ordinary taxable income.

In order to construct the same 19-bracket, 5-age-group, yearly schedule described above for ordinary marginal tax rates, we start by computing effective marginal tax rates in the PWBM household tax simulator. The PWBM tax simulator estimates an effective marginal rate for every filer by adding ordinary income (either labor for working-age people or OASI benefits for retirees) and recomputing the filer's tax liability. The change in tax liability over the change in income is the marginal effective rate. Note that this effective tax rate may include the effects of certain tax credits, most notably the Earned Income Tax Credit, whose value changes significantly as ordinary income changes.

As before, we take a weighted average of the marginal effective tax rate for all of the filers in each of the bins defined by year, age group, and 16-income groups computed using the same percentiles but applied to ordinary taxable. We pass the schedule to the dynamic OLG model. The dynamic OLG model takes this schedule of marginal effective tax rate to the household ordinary income, which results in the household ordinary tax liability.

Finally, we need a separate tax schedule for preferred income in the OLG model. In the dynamic OLG model, a share of total compensation comes from business income, and a large share of corporate business income is taxed at the preferred rate. We use the PWBM household tax simulator to generate a marginal effective tax rate on this preferred income. The age bins for the preferred schedule remain the same as before, however, instead of 19 brackets, we have 5 brackets based on 50th, 90th, 95th, 99th, and 99.9th percentiles of income based on household preferred distributions. As before, the PWBM household tax simulator computes the marginal effective rate for each filer for preferred income. We construct a weighted average of this marginal effective preferred tax rate for each of the bins to create the preferred tax schedule used by the OLG model. The preferred tax schedule in the dynamic OLG model is applied to only the preferred share of household income.

B.10.2 Payroll Taxes

Payroll taxes brackets and rates are set to their statutory values in each period.

B.10.3 Estate Taxes

The estate tax has two brackets and rates; a rate of 0.0 percent for the first bracket and a rate of 40.0 percent for the second bracket, both set to their statutory values. The threshold for the bracket, currently set at $0.9 million. This generates about $26 billion in estate tax receipts in the dynamic OLG model in 2024, consistent with 2024 statistics on the revenues from estate and gift taxes.

B.10.4 Consumption / VAT Taxes

In the default model, the consumption or VAT tax is set to 0.0 percent in all cases. The number of brackets and rates can be adjusted for policy analysis.

B.10.5 Business Taxes

Many of the parameters for the corporate / business tax were defined in section B.8. The corporate statutory tax rate, $\tau^{statutory}_{corp,j,t}$, is set to its current-law value of 21.0 percent in the baseline in all years.

B.10.6 State Income Taxes

The value for the state taxes are set to fully fund state obligations for SNAP and Medicaid. The parameter, $\tau_{State,t}$, is the state tax applied only to a share of taxable labor income (total taxable labor income times a percent that is allowed to be deducted). This value is set endogenously in the model.

B.10.7 Interest Rates on U.S. Federal Debt

In this initial steady state, the distribution of coupon rates on federal debt is sorted by vintage. For example, debt maturing in two years has a coupon rate that is the weighted average of all outstanding marketable issues (excluding TIPS) that mature in two years. We compute this average effective coupon rate in each maturity year for all federal debt maturing in the next 30-year period from the Monthly Public Statement of Debt from the U.S. Treasury for December in the steady state year. The yield curve for future debt is estimated as part of the PWBM Microsim forecast process.

B.11 Foreign holding of U.S. Federal Debt

We set $\phi_t^D = 0.3$. For each dollar of new net issues, which is defined as the total new debt issues minus the retired debt (debt with a maturity of one year at the beginning of the period), $\phi_t^D = 0.3$ get purchased by foreign investors. PWBM has published its estimates of foreign take-up of U.S. federal debt at (https://budgetmodel.wharton.upenn.edu/issues/2020/9/21/capital-flows-update). Historically, this value has been closer to 40 percent, however, in recent years, a decline in foreign ownership of U.S. debt has reduced the long-run average of foreign take-up of debt to about 30 percent.

B.12 The maturity structure

In the first year of the simulation (T_0), we need the vector for the effective coupon rates ey_{T_0} and the vector of the amount of outstanding debt at the end of the previous period d_{T_0-1}. To construct the beginning-of-year distribution in year time T_0, we use information from the December issue of the Monthly Statement of the Public Debt of the United States in year $T_0 - 1$. To construct the distribution, we limit ourselves to marketable securities. From these the marketable securities, we exclude Treasury Inflation Protected Securities (TIPS) and Floating Rate Notes. We define this set of securities $TREAS$, which are the set of treasury bonds, bills, and notes. For a single issue in this set of securities $\Lambda \in TREAS$ we define $mat(\Lambda)$ as the security's maturity in years, $out(\Lambda)$ as the security's outstanding value, and $int(\Lambda)$ as the security's coupon rate (the coupon for bonds and notes, which are typically sold close to par, and yield for bills, which are typically sold below par). From this population of Treasury bills, bonds, and notes, we construct each element of d_{T_0-1}, which corresponds to each maturity i for the next 30 years:

$$d^i_{T_0-1} = \frac{\sum_{\Lambda}^{TREAS} out(\Lambda) I[mat(\Lambda) = i)]}{\sum_{\Lambda}^{TREAS} out(\Lambda)} \tag{B.71}$$

where $I(.)$ is an indicator function.

We construct the average coupon rate using the weighted average of the interest rates from each maturiy of debt:

$$ey^i_{T_0-1} = \sum_{\Lambda}^{TREAS} int(\Lambda) \frac{out(\Lambda)}{\sum_{\Lambda}^{TREAS} out(\Lambda) I[mat(\Lambda) = i)]} I[mat(\Lambda) = i)] \tag{B.72}$$

Online Appendix References

Blewett, Lynn A., Julia A. Rivera Drew, Miriam L. King, Kari C.W. Williams, Daniel Backman, Etienne Breton, Grace Cooper, and Stephanie Richards. 2025. *IPUMS health surveys: National health interview survey, version 8.1 [dataset].* Place: Minneapolis, MN. https://doi.org/10.18128/D070.V8.1.

Cagetti, Marco, and Mariacristina De Nardi. 2008. "Wealth Inequality: Data and Models." Num Pages: 29 Place: Cambridge, United Kingdom, *Macroeconomic Dynamics* 12 (S2): 285–313. https://doi.org/http://dx.doi.org/10.1017/S1365100507070150.

Collins, Sara R., Munira Z. Gunja, and Gabriella N. Aboulafia. 2020. *U.S. Health Insurance Coverage in 2020: A Looming Crisis in Affordability.* August. https://doi.org/10.26099/6aj3-n655.

Congressional Budget Office. 2016. *Private Health Insurance Premiums and Federal Policy.* February.

———. 2018. *The 2018 Long-Term Budget Outlook.* June.

———. 2020a. "How CBO Analyzes the Costs of Proposals for Single-Payer Health Care Systems That Are Based on Medicare's Fee-for-Service Program." *CBO Working Paper* (December 10, 2020): 209.

———. 2020b. *Who Went Without Health Insurance in 2019, and Why?* September.

De Nardi, Mariacristina, Giulio Fella, and Gonzalo Paz-Pardo. 2020. "Nonlinear Household Earnings Dynamics, Self-Insurance, and Welfare." *Journal of the European Economic Association* 18, no. 2 (April): 890–926. https://doi.org/10.1093/jeea/jvz010.

De Nardi, Mariacristina, Eric French, and John Bailey Jones. 2010. "Why Do the Elderly Save? The Role of Medical Expenses." *Journal of Political Economy* 118, no. 1 (February 1, 2010): 39–75. https://doi.org/10.1086/651674.

De Nardi, Mariacristina, and Fang Yang. 2014. "Bequests and heterogeneity in retirement wealth." *European Economic Review* 72 (November 1, 2014): 182–196. https://doi.org/10.1016/j.euroecorev.2014.09.004.

Farmer, Leland E., and Alexis Akira Toda. 2017. "Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments: Discretizing Markov processes" [in en]. *Quantitative Economics* 8, no. 2 (July): 651–683. https://doi.org/10.3982/QE737.

Feiveson, Laura, and John Sabelhaus. 2018a. *Intergenerational Transfers and Wealth Concentration.* May.

———. 2018b. *Intergenerational Transfers and Wealth Concentration.* Technical report. May.

Fella, Giulio, Giovanni Gallipoli, and Jutong Pan. 2019. "Markov-chain approximations for life-cycle models" [in en]. *Review of Economic Dynamics* 34 (October): 183–201. https://doi.org/10.1016/j.red.2019.03.013.

French, Eric, and John Bailey Jones. 2011. "The effects of health insurance and self-insurance on retirement behavior." *Econometrica* 79 (3): 693–732.

Guth, Madeline, Meghana Ammula, and Elizabeth Hinton. 2021. *Understanding the Impact of Medicaid Premiums & Cost-Sharing: Updated Evidence from the Literature and Section 1115 Waivers.* Kaiser Family Foundation.

Hendricks, Lutz. 2002. *Intended and Accidental Bequests in a Life-cycle Economy.* February.

Hubbard, R. Glenn, Jonathan Skinner, and Stephen P. Zeldes. 1994. "Expanding the Life-Cycle Model: Precautionary Saving and Public Policy." *The American Economic Review* 84 (2): 174–179.

Kaiser Family Foundation. 2021. *Employer Health Benefits 2021.*

Karahan, Fatih, and Serdar Ozkan. 2013. “On the persistence of income shocks over the life cycle: Evidence, theory, and implications.” *Review of Economic Dynamics* 16, no. 3 (July): 452–476. https://doi.org/10.1016/j.red.2012.08.001.

Kirkby, Robert. 2025. “Discretizing earnings dynamics: implications of Gaussian-mixture shocks for life-cycle models.” *The Japanese Economic Review* 76, no. 2 (April): 493–519. https://doi.org/10.1007/s42973-025-00196-7.

Mitchell, Alison, Evelyne P Baumrucker, Kirsten J Colello, Angela Napili, Cliff Binder, and Julia A Keyser. 2021. *Medicaid: An Overview.* Congressional Research Service, February.

Moffitt, Robert A. 2020. “Multiple Program Participation and the SNAP Program.” In *SNAP Matters,* edited by Judith Bartfeld, Craig Gundersen, Timothy Smeeding, and James P. Ziliak, 213–242. Stanford University Press, December 31, 2020. https://doi.org/10.1515/9780804796873-011.

Nishiyama, Shinichi, and Kent Smetters. 2014. “Chapter 3 - Analyzing Fiscal Policies in a Heterogeneous-Agent Overlapping-Generations Economy.” In *Handbook of Computational Economics,* edited by Karl Schmedders {and} Kenneth L. Judd, 3:117–160. Handbook of Computational Economics Vol. 3. Elsevier. https://doi.org/10.1016/B978-0-444-52980-0.00003-7.

Radley, David C, Jesse C Baumgartner, Sara R Collins, Laurie Zephyrin, and Eric C Schneider. 2021. “Achieving Racial and Ethnic Equity in U.S. Health Care.” *The Commonwealth Fund,* 36.

Reichling, Felix, and Shinichi Nishiyama. 2015. *The Costs to Different Generations of Policies That Close the Fiscal Gap.* Technical report. Congressional Budget Office, December.

Scheinker, David, Barak D. Richman, Arnold Milstein, and Kevin A. Schulman. 2021. “Reducing administrative costs in US health care: Assessing single payer and its alternatives.” _Eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-6773.13649, *Health Services Research* n/a (n/a). https://doi.org/https://doi.org/10.1111/1475-6773.13649.

Scholz, John Karl, Ananth Seshadri, and Surachai Khitatrakun. 2006. "Are Americans Saving "Optimally" for Retirement?" *Journal of Political Economy* 11 (4): 37.

Social Security Administration. 2018. *Annual Statistical Supplement to the Social Security Bulletin, 2017.* Technical report. March.

Tanaka, Ken'ichiro, and Alexis Akira Toda. 2013. "Discrete approximations of continuous distributions by maximum entropy" [in en]. *Economics Letters* 118, no. 3 (March): 445–450. https://doi.org/10.1016/j.econlet.2012.12.020.

The Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. 2020. *2020 annual report of the Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds.*

Woolhandler, Steffie, Terry Campbell, and David U. Himmelstein. 2003. "Costs of Health Care Administration in the United States and Canada." Citation Key Alias: Woolhandler.etal:2003, *New England Journal of Medicine* 349, no. 8 (August 21, 2003): 768–775. https://doi.org/10.1056/NEJMsa022033.

Woolhandler, Steffie, and David U. Himmelstein. 2017. "Single-Payer Reform: The Only Way to Fulfill the President's Pledge of More Coverage, Better Benefits, and Lower Costs." *Annals of Internal Medicine* 166, no. 8 (February 21, 2017): 587–588. https://doi.org/10.7326/M17-0302.