CALIBRATION AS A CHANNEL FOR DISCRETIZATION ERROR

Orhan Torul*

Boğaziçi University

August 27, 2026

Abstract

Discretization error reaches a calibrated model's conclusions by two routes: through the solution of the agent's problem, and through the calibrated parameters themselves, because the target moments are computed on the same grid. Standard diagnostics measure only the first. I derive the decomposition and a screening statistic, A, computable from the calibration Jacobian, two extra solves per calibrated parameter, and one solve at a finer grid. In a heterogeneous-agent model of education finance, $A = 14$: a grid that hits every target overstates a reform's welfare value by a factor of nine, with 94% of the gap due to the calibration. Under conventional aggregate targets — in the Aiyagari model, two further economies, and a published replication archive at its own discretization — the statistic validates standard practice. Retargeting the same parameters to distributional moments raises A by orders of magnitude: the channel belongs to the pairing of model, conclusion, and target.

Keywords: Computational methods, calibration, numerical convergence, sensitivity analysis, heterogeneous agents.

JEL Codes: C63, C68, D63, E24, H20, I22.

*Department of Economics, Boğaziçi University, Natuk Birkan Building, Bebek, 34342 Istanbul, Türkiye. E-mail: orhan.torul@bogazici.edu.tr. A replication archive containing every production script and the processed output behind each table and figure accompanies this submission.

1 Introduction

Consider a quantitative model of education finance, calibrated in the ordinary way to match each region's public share of education spending, and asked whether a European-style fiscal regime would raise welfare in the United States. Solved on a uniform grid of sixty points for human capital, the model hits every calibration target, converges by its own tolerances, and answers that the reform is worth 3.8% of lifetime consumption, with every household in favor. Solved on a grid four times finer and calibrated again the same way, it answers 0.4%, with a quarter of the population opposed. The coarse version also fails an untargeted test that the fine version passes: it predicts that Europe produces more per capita than the United States.

The obvious reading is that the coarse grid solves the household problem badly. It does not. Holding the calibrated parameters fixed and refining the grid moves the welfare number by 0.11 percentage points. Holding the grid fixed and swapping in the parameters the finer grid selects moves it by 3.17. Under that ordering the parameters account for 3.17 of the 3.36 percentage-point gap, the solver for 0.11, and their interaction for the remaining 0.08. What changed between the two runs was the calibration, not the solution of the agent's problem. The moment identifying the scale of government maps to $\bar{T}/\bar{Y} = 0.108$ when the model is solved coarsely and 0.097 when it is solved accurately, and because the aggregate response to fiscal scale is $\cup$-shaped with the calibrated economies near its trough, that difference carries them across the turning point.

The channel that produced this is generic, even where its magnitude is not. Discretization can reach a calibrated conclusion two ways. It distorts the solution of the agent's problem at given parameters, and it distorts the parameters themselves, because the moment used to pin a parameter down is computed from the same discretized model. Tolerance checks, Bellman residuals and fixed-parameter convergence studies address the first route. Unless the calibration is repeated at every resolution, which is the expense they exist to avoid, they are silent on the second. Writing θ for the calibrated parameters, m for the target moments, W for the reported conclusion and g for the discretization, the implicit function theorem applied to the calibration condition $m(\hat{\theta}(g); g) = m^*$ splits the total effect of the grid on the conclusion into

$$\frac{dW}{dg} = \underbrace{\frac{\partial W}{\partial g}}_{\text{solver}} \underbrace{- \left(\frac{\partial W}{\partial \theta}\right)' \left(\frac{\partial m}{\partial \theta}\right)^{-1} \frac{\partial m}{\partial g}}_{\text{calibration}}.$$

The second term is what no reported diagnostic sees. It is large when three conditions hold at once: the conclusion is sensitive to a parameter, the mapping from that parameter's target moment to the parameter is locally flat, and the target moment is itself grid-sensitive. None of the three is visible in the checks a paper ordinarily runs.

The ratio of the two terms is a number a researcher can compute before deciding whether to re-solve anything. Call it A. It requires the calibration Jacobian, which a bisection or Newton search produces as a by-product; two extra solves per calibrated parameter, at the calibration already in hand; and one solve at a finer grid holding parameters fixed. It does not require a second calibration at the finer grid, and that is the expensive step it exists to let a researcher skip. A is a screening statistic and not a forecast: it says which of the two channels deserves the effort, conditional on the discretization moving the conclusion at all, and Section 6.2 shows that it cannot be trusted for the size or even the sign of the change over a step that crosses a turning point. It should always be read next to the magnitude it decomposes.

The paper then asks whether the calibration channel matters in practice, and answers in three steps. First,

in the education-finance model that opens this introduction, run on nothing but the coarse calibration and one finer solve, $A = 14$: a convergence study at fixed parameters is close to uninformative about this conclusion, and the two-by-two experiment confirms it. Second, in the incomplete-markets economy of Aiyagari (1994), with the discount factor set to a capital–output ratio as that literature sets it, $A = 0.15$: the solver carries seven-eighths of the discretization error, refining the asset grid at fixed parameters is the right check, and conventional practice is vindicated. The two verdicts differ by a factor of about a hundred, and the difference lies in how sharply the target moment pins the parameter down. Third, because the Aiyagari economy admits several natural targets for the same parameter, the comparison can be run inside a single model. Holding the environment, the grids and the conclusion fixed and changing only the moment the discount factor is calibrated to moves A by two orders of magnitude, in a model with a monotone aggregate response and no turning point anywhere near the calibration. The calibration channel is therefore a property of the pairing between a model, a conclusion and a target moment, and a researcher chooses the last of the three. A final block of evidence runs the identical protocol in three further environments: an exchange economy in the manner of Huggett (1993), a life-cycle economy in the tradition of Huggett (1996), and the replication archive of Conesa et al. (2009), compiled and screened at its published discretization. The division repeats. Conventional aggregate targets give small A with the signed prediction validated by measured two-by-twos; distributional targets give an implied parameter step that leaves the admissible range, or a coarse-grid target value that no parameter can attain on the fine grid; and the published configuration passes the screen, with both channels below two hundredths of a percentage point on a conclusion of 1.33%.

Two things should be said about the education model before it is used. It is the fragile case rather than a typical one, and choosing it that way is deliberate, since a diagnostic earns its keep by separating the fragile from the benign in advance. And the coarse grid used to expose the mechanism is coarser than the one the antecedent paper actually used. Solved on the grid documented in that paper's replication archive, this environment lands close to the converged answer (Section 7.4). No published conclusion is shown here to be wrong. What is shown is that nothing in the diagnostics a paper ordinarily reports would reveal it if one were.

Running the check routinely requires cheap converged solves, and the paper supplies one. Two features of the case-study environment let the household's intra-period problem be maximized in sequence rather than jointly. Proposition 1 states the conditions in a form that covers any dynastic or life-cycle problem in which one choice moves only the current payoff and another moves only the continuation state. Exploiting them shrinks the candidate array by a factor of forty and cuts a full stationary solve by a factor of twenty, with policies verified identical to the brute-force formulation.

Finally, on the economics of the case study: the converged model matches the sign of every untargeted cross-Atlantic gap and between 18 and 46 percent of the magnitudes, and fourteen discretizations put the residual numerical variation in the welfare figure at 0.22 percentage points. The economic sensitivities are larger than that. The welfare figure reverses to -2.5% once non-labor income measured on the Panel Study of Income Dynamics enters household budgets (Appendix D.1), so $[-2.5\%, +0.4\%]$ is a better summary of what this environment says than any point estimate. The substantive claim about transatlantic welfare is correspondingly modest, and the methodological claim does not depend on which end of that interval is right.

Section 2 places the argument in the literature. Section 3 derives the decomposition, defines the statistic, and states the protocol for computing it. Section 4 sets out the case-study model, its solution method and calibration, and the equilibrium objects the conclusion is built from. Section 5 is the case: two resolutions, two conclusions,

and the two-by-two that separates the channels. Section 6 computes the statistic in that model, including under the informational restriction an actual user faces. Section 7 computes it in the Aiyagari economy, runs the controlled experiment across target moments, extends the screen to the further environments just described, and solves the case-study environment on the antecedent's documented grid. Section 8 concludes. The Supplemental Appendix collects the computational detail, the data work, additional properties of the case-study economy, and the economic sensitivity analysis (Appendices A–G).

2 Related Literature

The paper's methodological point falls between two literatures that rarely meet. One studies the accuracy of solution methods, asking how finely to discretize, how to interpolate, and when projection or perturbation beats value-function iteration. Judd (1992, 1998) set out the projection and approximation toolkit, Santos (2000) shows that Euler-equation residuals bound the approximation error in the policy function and so supplies the accuracy check the field now reports as a matter of course, and Aruoba et al. (2006) rank the leading methods on a common model by speed and accuracy together. In heterogeneous-agent economies the same programme runs from Krusell and Smith (1998) through the comparison exercise Den Haan (2010) organized around their model, the non-stochastic simulation of Young (2010), the continuous-time formulation of Achdou et al. (2022), and the survey of Maliar and Maliar (2014). What these share is the object being approximated: each judges a method by how closely it reproduces the solution of the agent's problem at given parameters. The other studies how the moments a model is asked to match determine its parameters, and how those parameters in turn determine counterfactuals. Andrews et al. (2017) formalize the sensitivity of estimated parameters to the moments that identify them, and their statistic is the inverse Jacobian that appears in the decomposition of Section 6. The two concerns compose. Solution accuracy governs how much the grid moves the target moment; moment sensitivity governs how much of that movement the calibrated parameter must absorb; and their product is what carries discretization error into the parameter and from there into whatever the paper concludes. Neither literature on its own would prompt a researcher to compute that product, because an accuracy check holds the parameters fixed and a sensitivity analysis holds the solution fixed. The composition turns out to vary by about a factor of a hundred between the incomplete-markets economy of Aiyagari (1994) and the model studied here, so it is the composed object, and not either factor, that a researcher needs to look at.

The closest antecedent lies on the estimation side. Fernández-Villaverde et al. (2006) give conditions under which the likelihood implied by a numerically approximated policy function converges to the exact likelihood as the approximation improves, and show that second-order approximation errors in the policy function have first-order effects on that likelihood, so estimates computed from an approximated model carry the error into the parameters. The concern is the same in kind as the one raised here, since approximation error reaches the parameters rather than stopping at the policy function, and their finding that a small error in the policy can matter disproportionately further on is what the statistic below is built to measure. Their result is a limit theorem, while the object here is a magnitude computed at the grid a researcher is standing on. Their setting is likelihood estimation on a sample, whereas exact-identification calibration to a handful of moments leaves the Jacobian square and the decomposition available in closed form. And the question asked here is not whether the error vanishes in the limit but whether, at the resolution in use, it travels mainly through the parameters or mainly through the solver. That has different answers in different models, which is what the statistic is for.

The environment comes from Torul (2020). The two-cohort dynastic structure, the (h,ξ) state, the public-versus-private schooling margin with take-it-or-leave-it public provision, the human-capital law of motion including the spillover parameter ν, the calibration of the fiscal instrument to observed education-financing ratios, and the resulting $\cup$-shape in aggregates are all established there, along with the double-peakedness of preferences over taxes and the possibility that either transatlantic regime commands support. I take that environment as given and do not re-derive its positive results. New here are idiosyncratic earnings risk, the separation of fiscal scale from progressivity, the replacement of steady-state comparisons by perfect-foresight transitions, the pairwise-majority analysis that Torul leaves to future research, and the empirical work on the Panel Study of Income Dynamics. Beyond those ingredients, what this paper adds is the demonstration that in this environment the untargeted output ranking, and the sign and magnitude of a transition-based welfare comparison, depend on the discretization mainly through the calibrated parameters rather than through the accuracy of the household solution.

Four further literatures supply ingredients. Prescott (2004) originates the argument that differences in labor-income taxation account for the cross-Atlantic gap in hours worked, and the intensive-margin channel of Section 4.2.2 is his. Birinci et al. (2026) revisit that argument with two further decades of data and a much richer labor-supply environment, adding heterogeneity across individuals, an extensive margin, multi-member households and a detailed tax-and-benefit system, and find the raw hours gap substantially narrowed. This paper keeps the labor-supply block deliberately simple and adds an education-financing margin instead, because the question here is not how much of the hours gap taxes explain but whether one pair of fiscal instruments can move hours, the public-private composition of education spending, and post-tax inequality in the observed directions at once. The treatment of non-labor income in Appendix D.1 follows their strategy of identifying non-labor resources from cross-sectional micro data rather than endogenizing them.

The human-capital law of motion and the public-versus-private schooling margin follow Bénabou (2002, 2000) and Fernández and Rogerson (1998), and Stiglitz (1974) established that preferences over the public financing of education are generically double-peaked, a property that recurs in Appendix E. The environment here is a two-cohort dynastic version of that structure with a two-parameter tax schedule layered on top. The extension relative to Bénabou (2002) is not the education technology, which is standard, but the separation of the fiscal instrument into a size and a progressivity dimension, which Appendix C.3 shows act on different margins, and the use of the resulting comparative statics as untargeted predictions.

The tax function is Heathcote et al. (2017) and the cross-country progressivity estimates are Holter et al. (2019). Residual-earnings primitives are disciplined from Heathcote et al. (2010) and Krueger et al. (2010), with a direct estimate on the Panel Study of Income Dynamics reported in Section 4.4.4. Badel et al. (2020) and Daruich and Fernández (2024) combine human-capital accumulation with an asset margin, which this environment does not; Section 4.2.5 states what that costs.

Finally, Alesina and Angeletos (2005), Bénabou and Tirole (2006) and Piketty (1995) model beliefs about luck and effort as an equilibrium object with feedback into policy, and Alesina et al. (2018) and Stantcheva (2021) document how survey-measured beliefs relate to policy preferences. This paper does not model belief formation and does not use belief data as a test of its mechanism; Appendix C.4 explains why. On the political-economy side, Carroll et al. (2021) confront the difficulty that arises in Appendix E, a multidimensional policy space in which the median-voter theorem's premise fails, and their use of the uncovered set of McKelvey (1986) and of probabilistic voting is the protocol adopted here. Carroll et al. (2025) show that survey-measured attitudes toward redistribu-

tion track political identity more strongly than economic circumstance, which bounds how literally the mapping from economic state to vote in Appendix E should be read.

3 The Decomposition and the Statistic

The mechanism this paper studies follows from the structure of any calibrated counterfactual, and its magnitude can be computed in advance without ever re-calibrating at a finer grid. This section derives it in general; the models come later.

Let $\theta \in \mathbb{R}^k$ collect the internally calibrated parameters and let $m(\theta; g) \in \mathbb{R}^k$ be the model-implied target moments at discretization g, with data counterparts m^*. A word on g, since a derivative with respect to a node count is not well defined. Read g as a scalar index of a one-parameter family of refinements, a mesh width or a scaling of the number of nodes along a fixed path, so that the derivatives below are the limits of the finite differences actually computed. In the application g moves along a single such path, and the operational content of every $\partial/\partial g$ below is the difference between a coarse and a fine solve at fixed parameters. Equation (2) is then a first-order statement about that difference, which is why Section 6.2 tests how well it survives a large step. Calibration selects $\hat{\theta}(g)$ to solve $m(\hat{\theta}(g); g) = m^*$. Let $W(\theta; g)$ be the reported conclusion: a welfare number, an elasticity, a counterfactual difference.

Differentiating the calibration condition with respect to g and rearranging,

$$\frac{d\hat{\theta}}{dg} = -\left(\frac{\partial m}{\partial \theta}\right)^{-1} \frac{\partial m}{\partial g}, \tag{1}$$

so that the total derivative of the conclusion decomposes as

$$\underbrace{\frac{dW}{dg}}_{\text{what the reader sees}} = \underbrace{\frac{\partial W}{\partial g}}_{\text{solver channel}} \underbrace{- \left(\frac{\partial W}{\partial \theta}\right)' \left(\frac{\partial m}{\partial \theta}\right)^{-1} \frac{\partial m}{\partial g}}_{\text{calibration channel}}. \tag{2}$$

Equation (2) follows directly from the implicit function theorem, but it isolates a channel that conventional convergence diagnostics do not see. The solver channel is what tightening the Bellman tolerance controls and what a value-function convergence check measures. The calibration channel is invisible to both, and it can be large under three conditions that are easy to satisfy without noticing. They are a mechanism rather than a characterization. With more than one parameter the numerator is a sum, and terms in it can cancel, so the conditions are neither necessary nor sufficient taken singly. The conclusion is sensitive to the parameter, $\partial W/\partial \theta$ large. The calibration mapping from moment to parameter is locally flat, $\partial m/\partial \theta$ small. This is a conditioning property of the calibration, not identification in the statistical sense. And the target moment is itself grid-sensitive, $\partial m/\partial g$ non-negligible. Define

$$A \equiv \left| \left(\frac{\partial W}{\partial \theta}\right)' \left(\frac{\partial m}{\partial \theta}\right)^{-1} \frac{\partial m}{\partial g} \right| \Big/ \left| \frac{\partial W}{\partial g} \right|, \tag{3}$$

the ratio of the two channels. When $A \gg 1$, refining the solver at fixed parameters is not the check that matters.

Every object in (3) is cheap, and the protocol has three steps.

1. **Take $\partial m/\partial \theta$ from the calibration you already ran.** A bisection or Newton search produces the Jacobian as a by-product; no new solves are needed.

2. **Buy $\partial W/\partial\theta$ with two solves per calibrated parameter**, at the calibration you have, on the grid you have.
3. **Buy one solve at a finer grid, holding parameters fixed.** It returns $\partial m/\partial g$ and $\partial W/\partial g$ together, the grid-sensitivity of the target and the solver channel, so both terms of (2) are in hand.

Nothing here requires re-calibrating at the finer grid, which is the expensive step the diagnostic exists to let a researcher avoid, and nothing requires the two-by-two of Section 5, which requires it twice over. For the model below the whole calculation is a few dozen stationary solves at the cost established in Section 4.3. Section 6.2 carries it out under exactly this restriction and reports what the resulting number can and cannot be trusted to say.

4 The Fragile Case: An Economy of Education Finance

The first application is the environment of the opening vignette: a heterogeneous-agent dynastic economy of public versus private education finance, used to compare U.S. and European fiscal regimes. It is chosen as the fragile case, for the reasons the diagnostic itself will identify, and this section contains what the case requires: the facts the model is calibrated to and scored on, the model, the solution method, the calibration, and the two equilibrium objects the conclusion is built from — the stationary $\cup$-shape and the transition-based welfare number. Additional properties of the calibrated economy are in Appendix C, and the sensitivity of its substantive conclusions to economic assumptions is in Appendix D.

4.1 The Setting

The model is one of a class used to compare U.S. and European fiscal regimes. This section fixes the facts it is calibrated to and scored on, and Table 1 collects them. The macroeconomic aggregates are pre-COVID averages over 2014–2019, or the most recent pre-COVID year where the source publishes a single snapshot; the persistence ordering is taken from studies with their own samples and vintages. The European comparison group is the thirteen developed economies used by Alesina and Angeletos (2005), except for the education share, which the source reports for twelve of them. Full provenance is in Appendix B.

Table 1: The cross-Atlantic facts the model is asked to reproduce

Indicator	U.S.	Europe-13	Role in what follows
Public share of education expenditure (%)	69.6	87.0	calibration target
Progressivity index τ_p	0.137	0.199	set externally
Annual hours per worker	1,783	1,540	untargeted prediction
Gini, market income	0.517	0.479	untargeted prediction
Gini, disposable income	0.395	0.296	untargeted prediction
GDP per capita, PPP (USD)	62,876	56,339	untargeted prediction
Intergenerational persistence	U.S. > Europe		untargeted prediction

Notes: Education shares are OECD *Education at a Glance* (2020), averaged over the twelve European countries for which the source reports a value. Progressivity is Holter et al. (2019), Table 1. Hours are the OECD productivity database (2015); Ginis the OECD Income Distribution Database (≈2019); GDP per capita the World Bank WDI (2018). Persistence is the ordering in Corak (2013) and Chetty et al. (2014). "Market income" in the OECD series is household labor and capital income before taxes and public cash transfers, equivalized for household size; "disposable income" is the same concept after direct taxes and transfers. The model's income concept is narrower than either, being labor earnings of working-age households with no capital income, so market income is the closer comparator and Appendix C.3 returns to what that implies for levels.

The pattern is familiar. The United States finances a smaller share of education publicly, taxes labor income less progressively, works more hours, and redistributes less, from a market-income distribution only modestly more unequal. Two of these are inputs. The public financing share of education identifies the size-of-government instrument, one region at a time, in Section 4.4.3; progressivity is taken from published estimates. The rest are predictions, and Section 4.5.3 scores them.

Two further cross-Atlantic differences motivate this class of model in the literature and are deliberately not used here. The United States finances 50% of health expenditure publicly against a European average of 78%; the environment of Section 4.2 contains schooling and cash transfers but no health-insurance channel, and no claim in this paper concerns health. And the share of World Values Survey respondents attributing success to luck rather than effort is 22.9% in the United States against a European average of 36.3%. The model produces an objective variance share that is conceptually adjacent to that survey moment, but the two are not commensurable, one being a share of variance and the other a share of respondents, and the model contains no theory of how beliefs form or whether they cause the fiscal system rather than reflect it. Appendix C.4 returns to the point. Neither fact is a moment this paper scores.

4.2 Model

4.2.1 Environment

The economy is populated by a unit measure of households organized in dynastic lineages. Each household has two life stages, young and middle-aged. The young inherit human capital h from their parent's bequest decision and the prevailing public-education provision; they then mature into the middle-aged stage, where they work, consume, and choose a private bequest for their own descendants. There is no aggregate uncertainty and no population growth. I focus on the stationary equilibrium throughout Sections 4.4–4.5, and solve perfect-foresight transitions across stationary equilibria in Section 4.6.

Three ingredients do the work, and each is included because a specific competing account of the cross-Atlantic differences requires it. *First*, an intensive-margin labor supply choice with convex disutility: this is

the Prescott (2004) channel, through which the level of taxation moves hours, and no account of the cross-Atlantic hours gap can dispense with it. *Second*, a discrete public-versus-private education margin embedded in a human-capital law of motion, following Bénabou (2002, 2000) and Fernández and Rogerson (1998): this is what makes the *composition* of social spending endogenous rather than a parameter, and it is the ingredient that distinguishes this environment from the hours literature. *Third*, a two-parameter Heathcote et al. (2017) tax function: separating the size of government from the progressivity of the schedule is necessary because the two instruments turn out to act asymmetrically, one moving the education margin and the other post-tax dispersion, so a one-parameter flat tax cannot represent that asymmetry. What the environment omits, and what that omission costs, is set out in Section 4.2.5.

A middle-aged household is described by the pair (h, ξ), where $h \in \mathbb{R}_{++}$ is the household's own human capital and $\xi \in \mathbb{R}_{++}$ is the inborn cognitive endowment of the descendant. Both are realized at the start of the period. The household then chooses labor supply $l \in [0,1]$ and a private education bequest $e \geq 0$, and a multiplicative earnings shock ε is realized after those choices are made. The timing assumption, that $\{l, e\}$ are measurable with respect to (h, ξ) but not with respect to ε, splits log pre-tax income exactly into a component the household could foresee when it acted and an orthogonal residual. That split is definitional in this environment, and Section 4.2.4 is explicit that it is a statement about the model's information structure rather than a measure of what is deserved.

4.2.2 Household problem

The middle-aged household solves the recursive problem

$$V(h, \xi; \bar{H}, \bar{E}, \bar{T}r, \bar{Y}, \lambda_{HSV}, \tau_p) = \max_{\{l,e\}} \left\{ \mathbb{E}_\varepsilon[u(c)] + v(1-l) + \rho\, \mathbb{E}_{\xi'|\xi}\left[V(h', \xi'; \bar{H}, \bar{E}, \bar{T}r, \bar{Y}, \lambda_{HSV}, \tau_p) \right] \right\} \tag{4}$$

subject to the production, budget, non-negativity, and human-capital law-of-motion constraints

$$y = \Theta\, l^{1-\lambda}\, h^{\lambda} \exp(\varepsilon), \tag{5}$$

$$y_{\text{post}}(y; \lambda_{HSV}, \tau_p, \bar{Y}) = \bar{Y}\, \lambda_{HSV} \left(y / \bar{Y} \right)^{1-\tau_p}, \tag{6}$$

$$c = y_{\text{post}}(y; \lambda_{HSV}, \tau_p, \bar{Y}) + \bar{T}r + b(h) - e, \tag{7}$$

$$e \geq 0, \tag{8}$$

$$h' = \begin{cases} \xi\, \bar{E}^{\alpha}\, h^{(1-\alpha)\gamma}\, \bar{H}^{(1-\alpha)(1-\gamma)} & \text{if } e = 0, \\ \xi\, (e + \nu \bar{E})^{\alpha}\, h^{(1-\alpha)\gamma}\, \bar{H}^{(1-\alpha)(1-\gamma)} & \text{if } e > 0. \end{cases} \tag{9}$$

Equation (6) is the Heathcote et al. (2017) progressive tax function applied to realized labor income. The progressivity parameter $\tau_p \in [0, 1)$ governs the curvature of the after-tax mapping: $\tau_p = 0$ collapses to a flat tax with rate $1 - \lambda_{HSV}$, while $\tau_p \to 1$ approaches a fully redistributive scheme in which after-tax income is independent of pre-tax income. The level parameter λ_{HSV} governs the average tax rate at given τ_p. Because equation (5) makes $y > 0$ with probability one, the tax function is well defined on the whole support without extension, and post-tax income is strictly positive.

Two features of the earnings equation need comment. The exponent $1 - \lambda < 1$ on hours makes earnings concave in hours, following Bénabou (2002): total compensation rises with hours but less than proportionally, so at $\lambda = 0.625$ a household that halves its hours loses 23% of its earnings rather than half. Note what this specification does *not* do. Implied hourly compensation, $y/l \propto l^{-\lambda}$, is decreasing in hours, so the functional form is the

opposite of the part-time wage penalty estimated by Aaronson and French (2004), who find that men moving from full-time to half-time hours lose roughly a quarter of their hourly wage. Representing that penalty would require an hours exponent above one. I retain the Bénabou form because it is the specification of the antecedent literature this paper is extending, and because the curvature of earnings in hours is not an object any result here turns on; Appendix D.4 reports the sensitivity of the cross-Atlantic moments to the Frisch elasticity, which governs the hours response. The exponent λ on human capital governs the return to accumulated human capital in earnings; the return to the education *input* is α in equation (9). The shock enters multiplicatively in levels, or equivalently additively in logs,

$$\log y \;=\; \log\Theta + (1-\lambda)\log l + \lambda \log h + \varepsilon, \tag{10}$$

with $\varepsilon \sim \mathcal{N}(-\sigma_\varepsilon^2/2, \sigma_\varepsilon^2)$ so that $\mathbb{E}[\exp\varepsilon] = 1$ and aggregate income is unaffected by the dispersion of the shock. This is the specification the empirical earnings-dynamics literature uses (Heathcote et al., 2010; Krueger et al., 2010; Guvenen and Smith, 2014), and adopting it matters for three reasons beyond conformity. It makes σ_ε^2 the variance of a log residual, which is the object those papers estimate. It renders the variance decomposition of Section 4.2.4 invariant to the units in which income is measured. A decomposition taken on levels while holding a fixed level-variance shock is not, and its cross-regime comparative statics are unreliable in consequence. And it guarantees $c > 0$ without an artificial consumption floor, since $y_{\text{post}} > 0$ and the household chooses e knowing the distribution of ε.

The shock ε is independent across households and generations and orthogonal to the inborn endowment ξ and to everything in the household's information set when (l, e) are chosen.

The term $b(h)$ in equation (7) is exogenous non-labor income: capital income, private transfers, and public transfers outside the modeled labor-income tax instrument. It is untaxed by the HSV schedule and excluded from aggregate labor income $\bar{Y}$. In the benchmark $b \equiv 0$; Appendix D.1 switches it on, calibrated to its empirical distribution, and reports what changes. The schedule is

$$b(h) \;=\; \bar{B}\left(h/\bar{H}\right)^{\lambda\chi}, \tag{11}$$

so that b is proportional to permanent labor income raised to the power χ, is predetermined when (l, e) are chosen, and integrates to a share ω_b of total household income. This follows the route taken by Birinci et al. (2026): rather than endogenize a savings choice, households are endowed with non-labor income identified from cross-sectional micro data, so that asset- and transfer-income heterogeneity enters budget sets, and with them labor supply, welfare and the distribution of policy preferences, without adding a state variable.

Per-period preferences take the standard form $u(c) = \log c$ and $v(1-l) = -\phi\, l^{1+\omega}/(1+\omega)$, where ϕ is a disutility scale and $1/\omega$ is the Frisch elasticity of labor supply. The parameter ρ is the dynastic discount factor applied to the next generation's expected lifetime utility; under log per-period utility it is the discount factor and not a coefficient of relative risk aversion.

The law of motion in equation (9) embeds two features. First, parental and aggregate human capital enter symmetrically through the Cobb–Douglas combination $h^{(1-\alpha)\gamma}\,\bar{H}^{(1-\alpha)(1-\gamma)}$, with γ governing the share of own versus aggregate human capital; this is the usual neighborhood-effects channel familiar from Bénabou (2002, 2000); Fernández and Rogerson (1998). Second, education input enters as $\bar{E}$ if the household chooses public school ($e = 0$) and as $e + \nu\bar{E}$ if the household chooses private education ($e > 0$); the spillover parameter $\nu \in [0, 1]$ captures the extent to which a privately educated child still benefits from the publicly funded system (for exam-

ple, through public infrastructure that is non-rival in nature). Setting $\nu = 0$ yields a pure-segregation benchmark in which public funds are inputs only to the children of public-school households; setting $\nu = 1$ yields a pure-spillover benchmark in which public funds add to the educational input of private-school households as well. I take an intermediate $\nu = 0.5$ as the baseline and report robustness to the polar cases in Appendix D.

The inborn cognitive endowment $\log \xi$ is governed by an AR(1) process,

$$\log \xi_t = (1-\kappa)\,\mu_\xi + \kappa \log \xi_{t-1} + u_t, \qquad u_t \sim \mathcal{N}\Big(0,\,(1-\kappa^2)\sigma_\xi^2\Big), \tag{12}$$

with persistence parameter $\kappa \in [0,1)$ and unconditional distribution $\log \xi \sim \mathcal{N}(\mu_\xi, \sigma_\xi^2)$ in stationary equilibrium. I take $\kappa = 0$ as the i.i.d. benchmark and report the persistent variant $\kappa = 0.4$ as a robustness exercise in Appendix D.4.

4.2.3 Government and equilibrium

The government's fiscal regime is parameterized by a pair $(\bar{T}/\bar{Y}, \tau_p)$: the size of government as a share of aggregate output and the progressivity of the labor-income tax. Given $(\bar{T}/\bar{Y}, \tau_p)$ and the equilibrium joint distribution, the HSV level parameter λ_{HSV} is pinned in closed form by the requirement that the implied tax revenue match the target,

$$\bar{T} \;=\; \mathbb{E}_{\mu,\varepsilon}\big[y - y_{\text{post}}(y;\,\lambda_{HSV},\,\tau_p,\,\bar{Y})\big] \;=\; (\bar{T}/\bar{Y})\,\bar{Y}, \qquad \bar{Y} \equiv \mathbb{E}_{\mu,\varepsilon}[y] = \mathbb{E}_{\mu}\big[\Theta\left(l^*(h,\xi)\right)^{1-\lambda} h^{\lambda}\big]. \tag{13}$$

Equation (13) delivers $\lambda_{HSV} = (1-\bar{T}/\bar{Y})/M(\tau_p;\,\mu,\,\bar{Y})$, where $M(\tau_p;\,\mu,\,\bar{Y}) \equiv \mathbb{E}_{\mu,\varepsilon}[(y/\bar{Y})^{1-\tau_p}]$. For $\tau_p \in [0,1)$ and $\bar{T}/\bar{Y} < 1$ the shifter satisfies $\lambda_{HSV} > 0$, so post-tax income is strictly increasing in pre-tax income and the schedule is well defined on the whole support. Total revenue is then allocated to public-education provision and per-capita transfers in fixed proportion, $\bar{E} = (1-\psi)\,\bar{T}$ and $\bar{T}r = \psi\,\bar{T}$, where $\psi \in (0,1)$ captures the average composition of public spending observed in the U.S. data and is held fixed across counterfactuals.

Definition 4.1. A stationary equilibrium consists of a value function $V(h,\xi)$, decision rules $l^*(h,\xi)$ and $e^*(h,\xi)$, an implied human-capital transition $h^{*\prime}(h,\xi) \equiv h'(h,\xi,e^*(h,\xi),\bar{E},\bar{H})$ defined by equation (9), a joint distribution μ on $\mathbb{R}^2_{++}$, an endogenous HSV level parameter λ_{HSV}, and aggregates $(\bar{H},\bar{E},\bar{T}r,\bar{Y})$ such that, given fiscal regime $(\bar{T}/\bar{Y},\tau_p)$:

1. *Household optimization:* Given $(\bar{H},\bar{E},\bar{T}r,\bar{Y},\lambda_{HSV},\tau_p)$, V solves the Bellman equation (4) subject to constraints (5)–(9), and (l^*,e^*) are the associated maximizers.

2. *Distribution invariance:* The joint distribution μ is invariant under the policy-induced transition:

$$\mu(A) \;=\; \int \mathbf{1}\big\{(h^{*\prime}(h,\xi),\xi') \in A\big\}\, dF_\xi(\xi' \mid \xi)\, d\mu(h,\xi) \quad \text{for every measurable } A \subset \mathbb{R}^2_{++},$$

   where F_ξ is the conditional distribution of ξ' implied by equation (12).

3. *Aggregate consistency:* $\bar{H} = \mathbb{E}_\mu[h]$ and $\bar{Y} = \mathbb{E}_\mu[y] = \Theta\,\mathbb{E}_\mu[\left(l^*(h,\xi)\right)^{1-\lambda} h^{\lambda}]$, since $\mathbb{E}[\exp\varepsilon] = 1$.

4. *Government budget balance and HSV closure:* λ_{HSV} satisfies equation (13), $\bar{T} \equiv (\bar{T}/\bar{Y})\,\bar{Y}$, $\bar{E} + \bar{T}r = \bar{T}$, with $\bar{E} = (1-\psi)\,\bar{T}$ and $\bar{T}r = \psi\,\bar{T}$.

5. *Aggregate resource constraint:* $\mathbb{E}_\mu[y] = \mathbb{E}_\mu[c] + \bar{E} + \mathbb{E}_\mu[e^*(h,\xi)]$.

6. *Variance accounting:* The cross-sectional variance of *log* pre-tax labor income decomposes additively as $\mathrm{Var}(\log y) = \sigma^2_{\mathrm{det}} + \sigma^2_\varepsilon$, where $\sigma^2_{\mathrm{det}} = \mathrm{Var}_\mu\big(\log\Theta + (1-\lambda)\log l^*(h,\xi) + \lambda\log h\big)$.

I solve the equilibrium by a standard outer-loop fixed point on aggregates wrapping a value-function iteration on $V(h,\xi)$, following the algorithm of Huggett (1993). Computational details are in Appendix A.

4.2.4 The composition of pre-tax inequality

The timing assumption, that (l^*, e^*) are measurable with respect to (h,ξ) but not with respect to ε, splits the cross-section of log pre-tax income into a component the household could foresee when it acted and one it could not. From equation (10),

$$\log y = \underbrace{\log\Theta + (1-\lambda)\log l^*(h,\xi) + \lambda\log h}_{\log y^{\mathrm{det}}(h,\xi)} + \varepsilon, \qquad \iota_l \equiv \frac{\sigma^2_\varepsilon}{\mathrm{Var}_\mu\big(\log y^{\mathrm{det}}\big) + \sigma^2_\varepsilon}. \tag{14}$$

Because $\varepsilon \perp (h,\xi)$, the decomposition is exact. Two properties matter for what follows, and each is a reason to take the decomposition in logs rather than levels; a third — what happens as inborn ability and inherited human capital are themselves reclassified as luck — is developed in Appendix C.

The share is invariant to the tax schedule. Applying equation (6) gives $\log y_{\mathrm{post}} = \log\lambda_{HSV} + \tau_p\log\bar{Y} + (1-\tau_p)\log y$, so $\mathrm{Var}(\log y_{\mathrm{post}}) = (1-\tau_p)^2\,\mathrm{Var}(\log y)$ and the numerator scales by the same factor. The pre-tax and post-tax shares coincide, and no separate post-tax object is needed. A progressive tax compresses the level of log dispersion without altering its composition; everything ι_l does across regimes therefore comes from equilibrium responses in l^* and in the distribution of h, not from the tax function acting mechanically on realized income.

The share is robust to the units of measurement. Rescaling income by a constant shifts $\log y$ by a constant and leaves $\mathrm{Var}(\log y^{\mathrm{det}})$ unchanged. The same statement is false for a decomposition of the variance of income *in levels* against a shock of fixed level-variance: there the foreseeable variance scales with the square of the income unit while the shock variance does not, so the share moves with an arbitrary normalization. Appendix F documents how large that sensitivity is in this model and why it matters for the comparative statics.

4.2.5 What the environment leaves out

Three omissions are worth naming before the quantitative results, because each bounds what the model can be asked to explain.

First, there are no savings in a risk-free asset. Households cannot self-insure against ε; consumption inherits its full dispersion. This overstates the utility cost of earnings risk and, with it, the demand for public insurance, and it forces all cross-country differences in insurance onto the tax-and-transfer system. The literature that studies public insurance in incomplete markets (Aiyagari, 1994; Huggett, 1993) treats precisely the interaction between public insurance and private precautionary saving that is shut down here, and frameworks combining human-capital accumulation with an asset margin exist (Daruich and Fernández, 2024; Badel et al., 2020). Appendix D.1 takes the intermediate route. Non-labor income is calibrated to its empirical cross-sectional distribution so that asset- and transfer-income heterogeneity enters budget sets, without endogenizing the savings

choice. That is a partial answer, and the welfare results in Section 4.6 should be read as describing the labor-income channel rather than the full welfare consequences of a cross-Atlantic fiscal reform.

Second, there is no capital, pensions, health insurance, or unemployment insurance. The fiscal instrument is a labor-income tax financing public education and cash transfers. Pay-as-you-go pensions, capital taxation, social-insurance contributions and publicly delivered health care are outside it. Section 4.1 documents the cross-Atlantic gap in health financing because it is part of the pattern the paper is motivated by, but health is not modeled and no claim is made about it.

Third, there is no belief formation. Section 4.1 documents the cross-Atlantic gap in luck-leaning beliefs, and the model produces an objective variance share that is conceptually adjacent to it. The two are different objects, measured in different units, and the model contains no mechanism by which beliefs form, propagate, or feed back into policy. Appendix C.4 explains why the paper does not use belief data as a test of the mechanism.

4.3 Solving the Model

The stationary equilibrium is an outer fixed point on the aggregate state $(\bar{H}, \bar{E}, \bar{T}r)$ wrapping value-function iteration on $V(h,\xi)$, in the manner of Huggett (1993). Written naively, each Bellman step evaluates

$$V(h_i,\xi_j) \;=\; \max_{a,\,b} \Big\{ u\big(e_a, l_b, h_i\big) \;+\; \rho\,\mathbb{E}\big[V\big(h'(e_a,\xi_j,h_i),\xi'\big) \mid \xi_j\big] \Big\}$$

over an $(N_e \times N_l \times N_\xi \times N_h)$ array of candidates. That is more work than the problem requires, for two reasons that are visible in the primitives rather than in the algorithm.

Next-period human capital does not depend on hours. In equation (9), h' is a function of (e,ξ,h) only; labor supply enters current income and the disutility term but not the law of motion. The continuation value therefore needs to be interpolated on an $(N_e \times N_\xi \times N_h)$ grid, not an $(N_e \times N_l \times N_\xi \times N_h)$ one. Interpolation is the dominant cost of the inner loop, so this alone removes a factor of N_l from it.

Current utility does not depend on inborn ability. The flow payoff $u(e_a, l_b, h_i)$ is independent of ξ_j, which enters only through the continuation. Writing $W[a,j,i]$ for the interpolated continuation value, the maximization separates exactly. Both reductions are instances of nested maximization rather than of separability in the usual sense, since the value function does not decouple. The conditions generalize, and are stated that way here because they are not special to this environment.

Proposition 1 (Nested intra-period maximization)**.** *Consider a Bellman equation of the form*

$$V(s) \;=\; \max_{a\in\mathscr{A},\; b\in\mathscr{B}(a,s)} \Big\{ u(a,b,s) \;+\; \rho\,\mathbb{E}\big[V(s') \mid a,s\big] \Big\}, \qquad s=(s_1,s_2).$$

(i) Conditional maximization: *If the transition depends on the first choice but not the second, $s' = s'(a,s)$, then*

$$V(s) \;=\; \max_{a\in\mathscr{A}} \Big\{ \rho\,\mathbb{E}[V(s'(a,s)) \mid a,s] \;+\; \max_{b\in\mathscr{B}(a,s)} u(a,b,s) \Big\},$$

and the maximizers coincide with those of the joint problem. The continuation value therefore need only be evaluated on the (a,s) grid rather than the (a,b,s) grid.

(ii) Reuse across a state sub-vector: *If in addition the flow payoff does not depend on s_2, so that $u(a,b,s) = u(a,b,s_1)$ while the continuation does depend on s_2, then the inner maximization is independent of s_2 and can be performed once and reused across it.*

Both parts are immediate. In (i) the continuation is constant in b, so the maxima may be taken in either order; in (ii) the inner problem does not contain s_2. Neither part alters the Bellman operator. Its right-hand side is the same function of V as before, evaluated in a different order, so the fixed point and the convergence properties of the iteration are untouched; what changes is the number of candidate points at which the objective is formed. The content is in the conditions, and it is worth being precise about what they do and do not require. Part (i) needs *only* that the continuation not depend on b. It does *not* require the flow payoff to be additively separable in (a, b), and here it is not, because education and hours both enter consumption through the budget constraint (7), so $\mathbb{E}_\varepsilon[u(c)]$ is a non-separable function of the two. What makes the reformulation exact here is that the inner maximization over hours is taken *conditional on* education, one a at a time, not that the two choices decouple.

In the application a is education, b is hours, $s_1 = h$, and $s_2 = \xi$. Condition (i) holds because next-period human capital in equation (9) does not depend on l. Condition (ii) holds because the flow payoff depends on (e, l, h) but not on the descendant's inborn ability ξ, which enters only through the continuation $\mathbb{E}_{\xi'|\xi}[V(h', \xi')]$. The same structure appears in any dynastic or life-cycle problem in which an intra-period labor choice affects only the current payoff while an investment choice, whether education, saving or health, moves the continuation state. It fails, and the reformulation with it, if hours enter human-capital accumulation directly, as under learning-by-doing, or if the flow payoff depends on ξ. Applying (15) in this environment:

$$\max_{a,b}\Big\{u[a,b,i] + \rho\, W[a,j,i]\Big\} \;=\; \max_{a}\Big\{\rho\, W[a,j,i] \;+\; \max_{b} u[a,b,i]\Big\}. \tag{15}$$

The inner maximization over hours can be performed once, off the ξ dimension, and reused. Equation (15) is an identity rather than an approximation. The optimal policies and the value function are those of the original problem, and I verify as much numerically. Over twenty-five random draws of the continuation value at each of the two discretizations, the maximum discrepancy in the value function is exactly zero and the two formulations disagree on neither the education nor the labor policy at any state.

The two together reduce the candidate array from $(N_e \times N_l \times N_\xi \times N_h)$ to $(N_e \times N_\xi \times N_h)$: at the production discretization, from 10,944,000 candidates to 273,600, a factor of forty. Measured on the machine and software stack recorded in the replication archive, this cuts a single Bellman step from 199 to 5.4 milliseconds, a factor of 37, and a full stationary solve to an outer tolerance of 3×10^{-5} from 137 to 6.9 seconds, a factor of 20. The gap between the two ratios is the fixed cost of the distribution iteration and the outer loop, which the reformulation does not touch. That matters directly for the argument here, because the convergence study in Section 5 and Appendix A requires solving and re-calibrating the model at fourteen discretizations, which is a routine exercise at the reformulated cost and an unattractive one at the naive cost. The reduction in equation (15) is not special to this environment. It holds whenever one choice variable affects only the current payoff and another affects only the continuation, which is a common structure in models with an intra-period labor choice and an inter-period investment choice.

The human-capital grid spans $[0.01, 6.0]$, while the stationary distribution has $\bar{H} \approx 0.3$ and concentrates well below one. A uniform grid over that interval places the overwhelming majority of its nodes where the model puts almost no mass. Placing nodes where the approximation error is largest rather than spreading them evenly is long-standing advice (Judd, 1998; Maliar and Maliar, 2014). What Section 5 adds is that the cost of ignoring it is

paid mostly in the calibration rather than in the solution. The production grid is therefore power-spaced,

$$h_j = h_{\min} + (h_{\max} - h_{\min}) \left(\frac{j}{N_h - 1} \right)^2, \qquad j = 0, \ldots, N_h - 1, \tag{16}$$

which concentrates nodes at low h. Section 5 shows this is not a cosmetic choice. At a given number of nodes it is the difference between a solution that reproduces the cross-Atlantic output ranking and one that does not.

4.4 Calibration

The calibration has three parts. A first group of parameters is set externally, at values standard in the macro-labor and human-capital literatures. A second group is calibrated internally, and it has four members rather than the two the exposition might suggest. Two preference and endowment parameters are set to two laissez-faire targets, and the two fiscal instruments, one per region, are set to each region's observed public financing share of education. Progressivity is taken directly from cross-country estimates and is not calibrated. All four internally calibrated parameters enter the diagnostic of Section 6, which is a distinction that turns out to matter. Table 2 lists every parameter with its source, and Table 3 reports each targeted moment against its model counterpart.

4.4.1 Externally set parameters

The output exponent on human capital, $\lambda = 0.625$, follows Bénabou (2002); it also sets the curvature of earnings in hours discussed in Section 4.2.2. The intergenerational human-capital weight $\gamma = 0.20$ follows the cross-country evidence on the pass-through of cognitive outcomes in Schütz et al. (2008). The human-capital elasticity to the education input is $\alpha = 0.30$, in the range used in the human-capital production literature (Caucutt and Lochner, 2020). The dynastic discount factor $\rho = 0.80$ corresponds to a twenty-five-year generation at an annual discount rate of 0.99. The curvature of labor disutility is $\omega = 0.5$, a Frisch elasticity of two. This is the value Torul (2020) uses, and it lies at the high end of the range. Microeconometric estimates on the intensive margin are far smaller, while elasticities of this order are what Prescott (2004) requires to attribute the cross-Atlantic hours gap to taxes alone. Appendix D reports $\omega \in \{0.25, 1.0\}$, or Frisch elasticities of four and one. The public-private education spillover is $\nu = 0.5$, with the polar cases $\nu \in \{0, 1\}$ reported as robustness. The fiscal allocation $\psi = 0.370$ splits revenue between public education and per-capita transfers at the U.S. composition of social-investment expenditure in the OECD Social Expenditure Database (Organisation for Economic Co-operation and Development, 2019). The TFP scale $\Theta = 3.345$ is a units normalization. Taking the decomposition in equation (14) in logs removes the mechanical dependence of the luck share on this constant, though not all dependence: Θ is not a pure scaling here, because education spending and consumption share units and the human-capital law of motion is not homogeneous of degree one, so the equilibrium itself responds. Appendix F quantifies both channels.

4.4.2 Internally calibrated parameters

Two parameters are calibrated jointly, to two laissez-faire targets:

$$(\mu_\xi^*, \phi^*) \text{ such that } \bar{H}(\bar{T}/\bar{Y} = 0, \tau_p = 0) = 0.793, \qquad L(\bar{T}/\bar{Y} = 0, \tau_p = 0) = 0.338.$$

The disutility scale ϕ is identified primarily off the level of labor supply, and the location μ_ξ of the inborn-ability distribution off the level of aggregate human capital; equilibrium L depends weakly on μ_ξ through the bequest motive, so the two are solved jointly, by alternating bisection on each parameter until both targets are hit, rather than one at a time. Table 3 reports the fit.

These two targets need the same disclosure as σ_ξ^2 below, and more urgently. They are not data. No economy is observed with neither government nor progressivity, so there is no empirical moment to match; the two numbers are the corresponding row of Torul (2020), carried over so that this environment and its antecedent are on the same scale. Two consequences follow. First, they are not unit normalizations, whatever the motive for adopting them. The human-capital law of motion (9) is not homogeneous of degree one and education spending shares units with consumption, so the equilibrium responds to their level rather than merely rescaling, which is the point Section 4.4.1 makes about Θ. Second, and more pointedly for this paper, they are the output of a model solved on a particular grid, so the two parameters they pin inherit whatever discretization error that solution carried. The mechanism this paper formalizes is therefore operating on its own calibration and not only on the counterfactuals it reports. Section 6.2 measures how much of the calibration channel these two parameters account for, and the answer is a third of it.

4.4.3 The fiscal instruments

Size of government: The instrument $\bar{T}/\bar{Y}$ is the share of aggregate labor income collected by the labor-income tax and spent on the two channels the model represents, public education and per-capita transfers, in the fixed proportion ψ. It is *not* the OECD total tax-to-GDP ratio, and it should not be read as one. The OECD aggregate of 24.3% for the United States and 38.4% for the European comparison group finances pay-as-you-go pensions, publicly delivered health care, unemployment and disability insurance, and is raised in part from capital income, none of which appears in this environment. Imposing those numbers on an instrument that funds only schooling and transfers would misstate what the instrument does.

Instead, $\bar{T}/\bar{Y}_c$ is calibrated so that the model reproduces country c's observed public share of total education expenditure (OECD *Education at a Glance*: 69.6% for the United States, 87.0% for the European average), holding $\tau_{p,c}$ at its estimated value. This is an internal calibration and is reported as such: *the public financing share of education is a calibration target, not a prediction.* Hours worked, the level and cross-Atlantic ordering of pre-tax and disposable inequality, intergenerational persistence, and the composition of pre-tax income variation are not targeted, and it is those moments that Section 4.5.3 scores.

Progressivity: The HSV index τ_p is taken from Holter et al. (2019), who estimate equation (6) on OECD tax-and-benefit calculator output for 2000–2007. Their Table 1 gives $\tau_p^{\mathrm{US}} = 0.137$; the unweighted average over the twelve European countries they cover is $\tau_p^{\mathrm{EU}} = 0.199$.[1] The choice of source matters, because the two leading estimates are not taken on the same income concept. Holter et al. (2019) estimate the schedule on *labor earnings measured relative to median earnings*; Heathcote et al. (2017) estimate it on a broader notion of *total income inclusive of capital income*, and report a higher U.S. value, $\tau_p^{\mathrm{US}} = 0.181$. The model's tax base is pre-tax labor earnings of working-age households, normalized by aggregate labor income $\bar{Y}$. That is the Holter et al. concept rather than the Heathcote et al. one, so their estimates are the coherent choice, and they have the further advantage of being produced on a single harmonized sample across all the countries compared here. Appendix D.6 reports the Heathcote et al. (2017) anchor as a sensitivity.

[1] Belgium is absent from their Table 1; imputing it at the EU-12 mean leaves τ_p^{EU} unchanged to three decimals.

4.4.4 Earnings primitives

The variance of the log earnings shock is set to $\sigma_\varepsilon^2 = 0.259$, the residual variance of log U.S. wages in the cohort-based decomposition of Heathcote et al. (2010). Under equation (10) this parameter has the same units as the object that literature estimates, namely the variance of a log residual, which the additive specification of the antecedent literature does not. It is not the same object, and the difference should be stated. A model period here is a generation, so ε is variation in lifetime earnings that a household could not foresee when it chose hours and schooling, while the source moment is a cross-sectional residual from an annual wage regression. The two coincide only under assumptions this environment does not impose. No result about A depends on the value; Appendix C.4 explains why the level of ι_l should be read loosely in consequence. The variance of the inborn cognitive endowment is $\sigma_\xi^2 = 0.769$, and the generational persistence is $\kappa = 0$ in the benchmark, with $\kappa = 0.4$ reported as robustness. The provenance of σ_ξ^2 needs stating, because it governs how much foreseeable dispersion the model generates. It is inherited. In the additive-shock predecessor of this environment it was set to reproduce a target for the U.S. luck share. Under the log specification of equation (10) that target no longer applies, and the value is carried over rather than re-disciplined. Two consequences follow. First, the luck share ι_l is not an independent prediction of the model but a ratio of one externally set variance to a total that σ_ξ^2 largely determines, which is the observation a reader would make on inspecting equation (14). It is accordingly not scored in the model–data scorecard of Section 4.5.3. Second, the parameter can still be checked against data it was not set to match. The model's implied total variance of log pre-tax earnings is 0.612 at the U.S. regime and 0.591 at the European one, against 0.754 estimated on the PSID in Appendix B, an undershoot of about a fifth at the U.S. regime. Re-anchoring σ_ξ^2 to that variance directly would tighten the fit and is the natural next step; it would also make ι_l mechanically pinned by two data moments, which is a reason to stop treating the composition of pre-tax inequality as a prediction at all.

Table 2: Externally set parameters

Parameter	Symbol	Value	Source
Output exponent on human capital	λ	0.625	Bénabou (2002)
Intergenerational human-capital weight	γ	0.20	Schütz et al. (2008)
Human-capital elasticity to education	α	0.30	Caucutt and Lochner (2020)
Dynastic discount factor	ρ	0.80	25-year generation
Curvature of labor disutility (Frisch = $1/\omega = 2$)	ω	0.50	Torul (2020); polar cases in robustness
Public–private education spillover	ν	0.50	intermediate; polar cases in robustness
Revenue split, transfers vs. education	ψ	0.370	U.S. social-investment composition
Variance of log earnings shock	σ_ε^2	0.259	Heathcote et al. (2010)
Variance of log inborn ability	σ_ξ^2	0.769	inherited; provenance in Section 4.4.4
Generational persistence of ability	κ	0	i.i.d. benchmark; 0.4 in robustness
TFP scale (units normalization)	Θ	3.345	normalization

Notes: Parameters set outside the model. The two internally calibrated parameters and the two fiscal instruments are reported in Table 3.

Table 3: Internally calibrated parameters: targets and fit

Target moment	Parameter	Value	Target	Model
Aggregate human capital, $\bar{H}(\bar{T}/\bar{Y}=0)$	μ_ξ	+0.0818	0.793	0.7942
Labor supply, $L(\bar{T}/\bar{Y}=0)$	ϕ	2.2691	0.338	0.3400
Public share of education expenditure, U.S.	$\bar{T}/\bar{Y}^{\text{US}}$	0.0974	0.696	0.697
Public share of education expenditure, Europe	$\bar{T}/\bar{Y}^{\text{EU}}$	0.1166	0.870	0.871
Set externally, not calibrated				
Progressivity, U.S.	τ_p^{US}	0.137	Holter et al. (2019) Table 1	
Progressivity, Europe	τ_p^{EU}	0.199	Holter et al. (2019) Table 1	

Notes: The column is headed *Target* rather than *Data* because its entries have two provenances. The two laissez-faire targets are not empirical moments: no economy is observed at $\bar{T}/\bar{Y}=\tau_p=0$, and they are taken from the corresponding row of Torul (2020), and they are not unit normalizations either; Section 4.4.2 sets out what that inheritance costs. The two education shares are data (OECD *Education at a Glance*, 2020). The first two rows are calibrated jointly, by alternating bisection on each until both targets are hit; the next two set the size-of-government instrument one region at a time. Every moment scored in Table 5 other than the education-financing row is untargeted. Source: author's calculations.

4.5 Stationary Equilibrium

This section characterizes the calibrated stationary equilibrium. Section 4.5.1 documents the central comparative static, that output, human capital and the composition of earnings inequality are all ∪-shaped in the size of government. Section 4.5.3 then scores the model's untargeted predictions against the data. That the calibration places the United States at the trough of the ∪ is what makes the results of Section 5 possible. The converged anchor of 0.097 is the minimizer of output on this grid, so the economy sits at the turning point rather than merely near it, and the coarse anchor of 0.108 sits past it on the rising arm. A change in the anchor of that size therefore does not shift the economy along a monotone response; it moves it from one arm to the other. Figure 1 plots the aggregate response and marks where the two calibrated regimes sit on it.

4.5.1 Aggregates and the ∪-shape

Table 4 reports the stationary equilibrium across the size-of-government grid, holding progressivity at the U.S. value. Output, consumption, and aggregate human capital fall as $\bar{T}/\bar{Y}$ rises from zero, reach a minimum at the calibrated U.S. value of $\bar{T}/\bar{Y}=0.097$, and rise thereafter; labor supply declines monotonically throughout.

The first two rows of Table 4 are different economies, and the difference is worth pausing on because it is easy to misread. The laissez-faire row sets both instruments to zero and reproduces the calibration targets, $\bar{H}=0.794$ and $L=0.340$. The row beneath it sets revenue to zero but holds progressivity at the U.S. value, as the rest of the sweep does, and gives 0.684 and 0.304. A schedule with $\tau_p=0.137$ redistributes whatever its net revenue, because equation (6) compresses after-tax income whenever $\tau_p>0$. Zero revenue and zero progressivity are separate conditions, and only the first row imposes both.

Table 4: Stationary-equilibrium aggregates across the size-of-government grid

$\bar{T}/\bar{Y}$	Y	C	L	$\bar{H}$	$\bar{E}_{\text{tot}}$	Pop. pub.	ι_l
0.0000‡	1.75	1.54	0.340	0.794	0.213	0.0%	0.437
0.0000	1.53	1.35	0.304	0.684	0.178	0.0%	0.439
0.0250	1.47	1.29	0.297	0.644	0.166	1.0%	0.431
0.0500	1.25	1.10	0.286	0.508	0.131	21.3%	0.413
0.0750	1.00	0.89	0.275	0.364	0.093	55.2%	0.412
0.0974*	0.91	0.83	0.268	0.320	0.080	77.2%	0.423
0.1166†	0.93	0.85	0.263	0.333	0.083	87.8%	0.433
0.1500	1.12	1.01	0.257	0.452	0.112	96.5%	0.442
0.1750	1.35	1.20	0.254	0.613	0.152	98.7%	0.444
0.2000	1.60	1.40	0.251	0.807	0.203	99.6%	0.445

Notes: The first row, marked ‡, is the laissez-faire economy with both instruments at zero, and is the economy the two internally calibrated parameters are set to match. Every other row holds progressivity at the U.S. value $\tau_p = 0.137$, and a progressive schedule redistributes even at zero net revenue, which is why the $\bar{T}/\bar{Y} = 0$ row below it is a different economy. The joint U.S.-versus-Europe comparison, which varies both instruments, is in Table 5. "Pop. pub." is the share of households whose children attend public school. ι_l is the luck share of equation (14). * marks the calibrated U.S. size of government and † the European one. Source: author's calculations.

The ∪-shape is a horse race between two effects of taxation on the human-capital law of motion. Private education is purchased out of after-tax income, so its user cost rises with $\bar{T}/\bar{Y}$ and private bequests contract. Against that, revenue is allocated in fixed proportion $1-\psi$ to public provision $\bar{E}$, which rises with $\bar{T}/\bar{Y}$ and raises the schooling input of every household whose private bequest would not have exceeded $(1-\nu)\bar{E}$. At low fiscal size the first effect dominates, because the public-school option is small and unattractive, few households take it, and aggregate human capital falls. At high fiscal size the second dominates, as public provision captures nearly the whole population and $\bar{H}$ rises. Between the two, output bottoms out.

Two features of the rising arm bear directly on how much weight the results at high fiscal size can carry. First, it is not an artifact of the resource accounting. In the law of motion (9) a public-school child receives the aggregate public budget $\bar{E}$ rather than $\bar{E}$ divided among public-school children, which is a non-rivalry assumption about publicly financed educational infrastructure. Re-solving the model with public spending made fully rival leaves the ∪-shape intact and moves the trough by less than one percentage point of $\bar{T}/\bar{Y}$. The rising arm comes from the human-capital technology itself: $\bar{E}$ enters h' with elasticity α and scales with $\bar{T}/\bar{Y}\cdot\bar{Y}$, which feeds back into $\bar{H}$ and hence $\bar{Y}$. Second, and consequently, the model at high fiscal size predicts an economy in which nearly all children are publicly schooled and output climbs back past its laissez-faire level. This is a strong prediction, and it rests on the human-capital elasticity $\alpha = 0.30$ and on the assumption that public and private education dollars are equally productive, neither of which is estimated here. Both bear on where the trough falls, and the trough matters, because the calibrated U.S. economy lies close to it, so small changes in its location move the sign of the welfare comparison. That is the thread running through Sections 5 and D, first for the numerical determinants of the trough's location and then for the economic ones.

4.5.2 Intergenerational persistence

The mechanism that moves ι_l also moves intergenerational mobility. Substituting public for private inputs into h' reduces the cross-section's reliance on parental human capital, so the intergenerational elasticity $\beta_h \equiv \text{Cov}_\mu(\log h', \log h)/\text{Var}_\mu(\log h)$ falls with fiscal size. At the calibrated regimes the model gives $\beta_h^{\text{US}} = 0.250$ against

$\beta_h^{EU} = 0.194$, preserving the cross-Atlantic ordering documented by Corak (2013) and Chetty et al. (2014). This prediction is untargeted and is not mechanically implied by the moments the calibration matches.

4.5.3 A model–data scorecard

Table 5 collects the model's cross-Atlantic predictions. Four moments are targeted: aggregate human capital and labor supply at laissez-faire, and the two regions' public shares of education expenditure. Progressivity is set externally. Everything else is a prediction.

Table 5: Model versus data on the cross-Atlantic gaps

	Model		Data		Sign	Gap
Dimension	U.S.	Europe	U.S.	Europe	match	captured
Targeted						
Public share of education expenditure	0.697	0.871	0.696	0.870	—	—
Set externally						
Progressivity τ_p	0.137	0.199	—	—	—	—
Predictions (untargeted)						
Hours L (normalized)	0.268	0.250	0.286	0.247	yes	46%
Pre-tax inequality $G(y)$	0.422	0.415	0.517	0.479	yes	18%
Disposable inequality $G(\tilde{y})$	0.354	0.322	0.395	0.296	yes	33%
Persistence β_h	0.250	0.194	U.S. > Europe		yes	—
Output per capita Y	0.913	0.819	U.S. > Europe		yes	—
Top 10% share, pre-tax	0.310	0.306	not targeted		—	—

Notes: Income Ginis count the earnings shock, the convention consistent both with the luck share of equation (14) and with the annual survey income underlying the OECD series; Table C.1 reports the shock-marginalized convention, which is roughly ten Gini points lower. Data columns are the empirical anchors of Table 1; pre-tax inequality uses the market-income Gini. "Gap captured" is the model gap as a share of the empirical gap, reported where model and data objects are commensurable. The output row compares the model's ordering against GDP per capita at PPP. The top-share row is reported to locate the residual inequality shortfall: the model has no Pareto upper tail, so its top-end concentration is low by construction. Source: author's calculations.

The model matches the sign of the hours, inequality, and persistence gaps and between 18 and 46 percent of their magnitude. Levels are closer than the shock-marginalized convention suggests. Counting the earnings shock, the U.S. disposable-income Gini is 0.354 against 0.395 in the data and the pre-tax Gini is 0.422 against 0.517. The residual shortfall falls where one would expect. The model's top decile takes 31% of pre-tax income, and with no Pareto upper tail in the earnings process it cannot generate the concentration observed at the very top. The model also reproduces the cross-Atlantic output ranking, which it does *not* do when solved on a coarse grid. That is the subject of Section 5.

Every untargeted moment in Table 5 for which the model and the data measure the same object is matched in sign, including output per capita. The top-share row is the exception and is reported without a sign comparison, since a model with no Pareto tail cannot be scored on tail concentration. The coarse solution of the same model fails the output ranking; Section 5 takes up the difference.

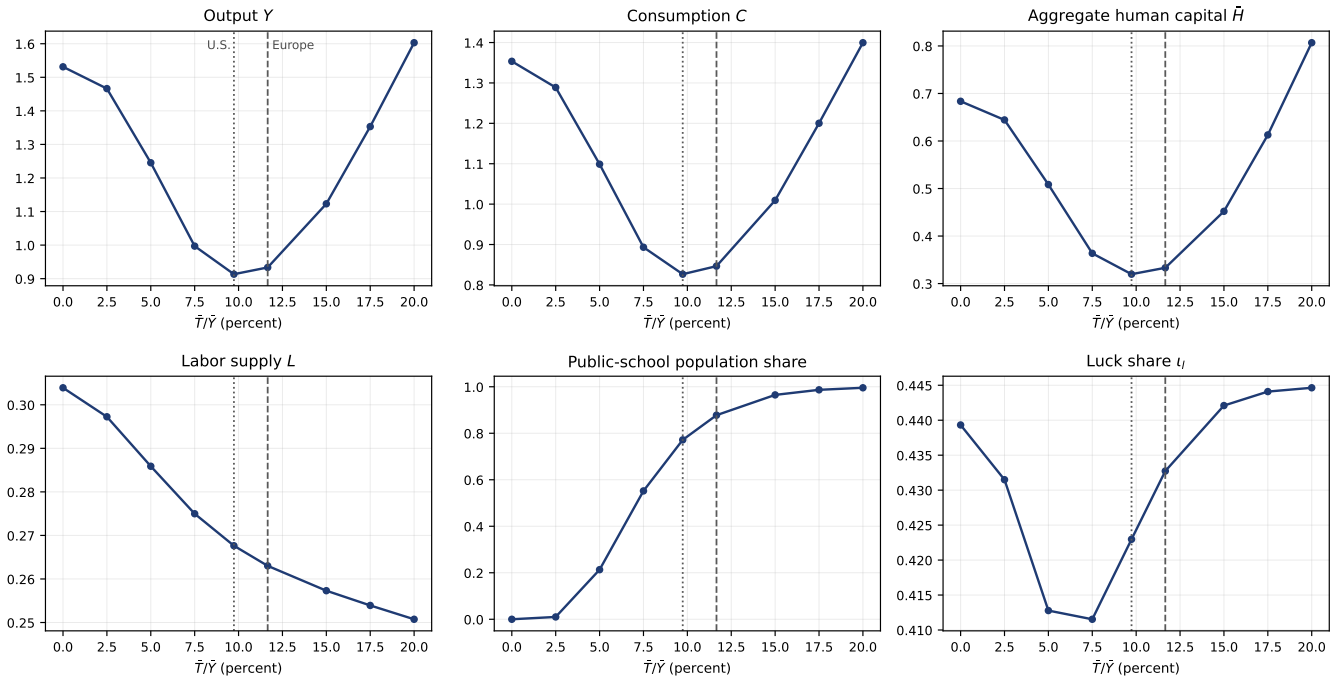

Figure 1: Aggregate variables across the size-of-government grid, holding progressivity at $\tau_p = \tau_p^{\mathrm{US}}$. The dotted vertical line marks the calibrated U.S. size of government and the dashed line the European one; both sit close to the trough of the response, which is the placement Section 5 shows the discretization moves. Panels correspond to the columns of Table 4. Source: author's calculations.

4.6 Transitions, Welfare, and Political Economy

Comparing stationary equilibria is silent about the households alive when a policy changes. This section solves perfect-foresight transitions between the calibrated regimes, and computes consumption-equivalent welfare cell by cell over the initial cross-section. The political-economy tournament built on these valuations is reported in Appendix E.

4.6.1 Setup

At date $t = 0$ the government announces a deterministic path $\{(\bar{T}/\bar{Y}_t, \tau_{p,t})\}$ converging to a new regime; households learn the whole path and re-optimize. The path is solved over $T = 25$ generations with terminal condition $V_T = V(\text{terminal regime})$, by backward induction on V coupled to forward iteration on μ_t from the status-quo distribution.[2] For a household at (h, ξ) at $t = 0$, consumption-equivalent welfare is

$$\zeta(h,\xi) \;=\; \exp\big[(1-\rho)\big(V_{\mathrm{path}}(h,\xi) - V_{\mathrm{SQ}}(h,\xi)\big)\big] - 1, \tag{17}$$

the constant proportional consumption change in every period and state that would make the household indifferent between the reform and the status quo. I report an aggregate ζ_{agg}, the μ-weighted median, the support share, and the difference between the bottom and top deciles of the initial human-capital distribution. The aggregate is the consumption equivalent that equalizes μ-weighted expected lifetime utility, which under log utility

[2] At $T = 15$ the aggregate path still shows a residual gap from the terminal steady state in the Europe-to-U.S. direction, so $T = 25$ is used throughout; the convergence study of Section 5.2 recomputes the welfare number at every discretization on that horizon.

is the geometric rather than the arithmetic mean of the cell-by-cell consumption ratios $1+\zeta$, less one; it is therefore the $\varepsilon_A = 1$ entry of the Atkinson comparison reported in Appendix C (Table C.2), and the arithmetic mean is the $\varepsilon_A = 0$ entry. The two differ by less than half a hundredth of a percentage point in every counterfactual reported here.

4.6.2 Welfare consequences

Table 6: Consumption-equivalent welfare across counterfactuals

Counterfactual	ζ_{agg}	ζ_{med}	Support share	D1 − D10 spread (pp)
U.S. → E.U. one-shot	+0.40%	+0.51%	75.7%	+0.76
U.S. → E.U. phased (5 gen.)	+0.13%	+0.17%	77.1%	+0.20
E.U. → U.S. one-shot	−0.49%	−0.60%	22.0%	−0.69
E.U. → U.S. phased (5 gen.)	−0.31%	−0.33%	9.2%	−0.15

Notes: Percent of consumption per period for life, along perfect-foresight transitions solved over $T = 25$ generations. The D1−D10 spread is the difference in ζ between the bottom and top deciles of the initial human-capital distribution; a positive value means the bottom decile gains relative to the top. Source: author's calculations.

Table 6 reports the transition results. At the benchmark calibration the U.S.-to-European reform is mildly welfare-improving: $\zeta_{\text{agg}} = +0.40\%$ of lifetime consumption, with 76% of the initial cross-section in favor and the remaining quarter opposed. The reverse reform is welfare-decreasing. The magnitude is small because the two instruments work against each other, as the decomposition below makes explicit.

Decomposing the reform between instruments, moving the size of government alone is worth +1.03%, moving progressivity alone −1.08%, and their interaction +0.45%. The two instruments pull in opposite directions and very nearly cancel, which is why the net figure is small. It is also why a one-instrument version of this model cannot represent the comparison, since the scale channel on its own would make the European regime unambiguously attractive, and it is the progressivity channel that offsets it.

Cross-decile dispersion is the second-order object here. The bottom decile of the initial human-capital distribution gains 0.76 percentage points more than the top in the U.S.-to-European direction. Phasing the reform in over more generations shrinks that dispersion substantially, because the dynastic discount factor $\rho = 0.80$ concentrates the welfare integral on the early generations, which a phased reform leaves close to the status quo.

The welfare ranking is invariant to inequality aversion: Atkinson-aggregated versions of Table 6 at $\varepsilon_A \in \{0, 0.5, 1, 2\}$ move each entry by at most a hundredth of a percentage point (Appendix C, Table C.2).

5 A Case: Two Resolutions, Two Conclusions

I solve and calibrate the model twice, at two discretizations, and then decompose the difference.

The coarse configuration is a uniform human-capital grid of sixty nodes, twenty-five education nodes, twenty-five labor nodes, eleven ability nodes, and a seven-node shock. The converged configuration is the production grid of Section 4.3. Within each I re-run the full calibration: the two laissez-faire parameters by alternating bisection until both targets are hit, and each fiscal anchor by bisection on its region's observed public share of education spending. Both columns of Table 7 are complete, internally consistent exercises that hit their targets.

Table 7: The same quantitative exercise at two solution accuracies

	Coarse grid	Converged grid	Data
Discretization			
N_h (spacing)	60 (uniform)	240 (power 2)	—
$N_e \times N_l \times N_\xi \times N_\varepsilon$	25×25×11×7	60×40×19×11	—
Calibration (re-done within each grid)			
μ_ξ	+0.0693	+0.0818	—
ϕ	2.3640	2.2691	—
$\bar{T}/\bar{Y}^{\mathrm{US}}$	0.1084	0.0974	—
$\bar{T}/\bar{Y}^{\mathrm{EU}}$	0.1401	0.1166	—
Targeted moment			
Public share of education, U.S.	0.696	0.697	0.696
Untargeted predictions			
Output ranking ΔY	+0.031	−0.094	< 0
ordering	**wrong**	correct	—
Hours gap ΔL	−0.021	−0.018	−0.039
Disposable Gini gap	−0.038	−0.032	−0.099
Luck share ι_l^{US}	0.303	0.423	—
Persistence β_h^{US}	0.207	0.250	—
Policy conclusion			
ζ_{agg}, U.S.→Europe	+3.76%	+0.40%	—
Support share	100.0%	75.7%	—

Notes: Both columns are complete, internally consistent quantitative exercises: within each discretization the two laissez-faire parameters and both fiscal anchors are re-calibrated to the same targets, so the comparison is between two calibrated models rather than between a calibrated and an uncalibrated one. On what the antecedent literature actually reports, see Section 7.4. Source: author's calculations.

The coarse solution gets an untargeted moment wrong and the converged one does not. Both reproduce the U.S. public financing share, because both are calibrated to it. They disagree about output. The coarse solution predicts Europe produces more per capita than the United States, and the converged solution gets the ranking right. And the policy conclusion differs by a factor of nine, from a large and unanimously supported welfare gain to a small and contested one.

5.1 Which channel: the solver or the calibration?

The natural reading of Table 7 is that the coarse solver computes welfare inaccurately. That reading is wrong, and separating the two candidate channels requires breaking the comparison apart. Table 8 puts the discretization on one axis and the calibrated parameter vector on the other, and evaluates all four combinations. The off-diagonal cells are not internally consistent, since each solves at one resolution using parameters calibrated at the other and therefore misses the education-financing target, as Panel B records. That is what makes them informative.

Table 8: Separating numerical error from re-calibration

	Calibrated parameter vector	
Discretization	Coarse ($\bar{T}/\bar{Y}$: 0.1084 → 0.1401)	Fine (0.0974 → 0.1166)
Panel A: welfare effect and output ranking		
Coarse grid	+3.76%, **wrong**	+0.59%, correct
Fine grid	+3.65%, **wrong**	+0.40%, correct
Panel B: public share of education expenditure, U.S. (target 0.696*)*		
Coarse grid	0.696	0.613
Fine grid	0.779	0.697
Panel C: decomposition of the total change in ζ_{agg}		
Total (coarse/coarse → fine/fine)	−3.36 pp	
Solver alone (calibration held fixed)	−0.11 pp	
Re-calibration alone (grid held fixed)	−3.17 pp	
Interaction	−0.08 pp	

Notes: Rows vary the discretization; columns vary the calibrated vector ($\mu_\xi, \phi, \bar{T}/\bar{Y}^{\text{US}}, \bar{T}/\bar{Y}^{\text{EU}}$). Only the two diagonal cells are complete calibration exercises: each is solved and calibrated at its own discretization and reproduces the education-financing target. The off-diagonal cells are valid solutions of the model at a parameter vector calibrated elsewhere, so they miss that target, as Panel B records, and they are reported because that is what isolates the two channels. Panel C decomposes the total change in the welfare effect. Source: author's calculations.

Panel C is the result. Holding the calibrated parameters fixed at the coarse values and refining only the grid moves the welfare effect by −0.11 percentage points. Holding the grid fixed at the coarse configuration and swapping in the converged calibration moves it by −3.17. The interaction is −0.08. Under this ordering the calibration accounts for 3.17 of the 3.36 percentage-point total, the solver for 0.11, and their interaction for 0.08.

So the household problem is solved accurately enough at the coarse grid. At fixed parameters, coarse and fine solutions agree on welfare to within a tenth of a percentage point, and they agree on the sign of the output ranking. With the coarse parameter vector both grids get the ranking wrong; with the converged vector both get it right. Reading down the columns of Panel A rather than across the rows shows it plainly. The output ranking tracks the *parameters*, not the grid.

What is resolution-dependent is the calibration. The moment that identifies the scale of government is each region's public share of education expenditure. That mapping is not neutral with respect to how the model is solved. The target 0.696 selects $\bar{T}/\bar{Y}^{\text{US}} = 0.1084$ when the model is solved coarsely and 0.0974 when it is solved accurately. Panel B shows the same thing from the other side. Transplanting either calibration onto the other grid misses the target by 8.3 percentage points, in one direction from below and in the other from above. The reform being evaluated is a different one. The coarse calibration raises $\bar{T}/\bar{Y}$ by 0.0318, the converged one by 0.0192, a policy experiment thirty-nine percent smaller.

A ten percent difference in the anchor does this much because the aggregate response is ∪-shaped and the calibrated economies lie near its trough. Section 4.5.1 locates the minimum of output at $\bar{T}/\bar{Y} \approx 0.097$, which is the converged U.S. anchor almost exactly. On the descending arm a larger government lowers output; past the trough it raises it. The coarse calibration places the United States beyond the turning point, so the European regime, further right still, comes out richer and a move toward it looks unambiguously good. The converged calibration places the United States at the turning point, where the progressivity channel, which runs the other way, is enough to reverse the ordering. Small differences in placement have large consequences only because of where the calibration lands, and where it lands depends on the resolution.

The usual diagnostics miss all of this. At both resolutions the inner value-function iteration converges, the outer fixed point on aggregates converges, and the calibration hits its targets. Nothing in that battery is informative about the object that actually moves, which is the mapping from a targeted moment to the parameter it identifies. A model can be solved accurately enough for the household problem and not accurately enough for the calibration, and it is the second that determines what the paper concludes. This is why the convergence study below is run over the calibrated anchors and the untargeted moments rather than over the value function.

5.2 Is the converged solution converged?

Table 9: Numerical convergence at the production calibration

Discretization	$\bar{H}^{\mathrm{US}}$	L^{US}	ι_l^{US}	Public share	ΔY	ζ_{agg}
Production: N_h=240, N_e=60, N_l=40, N_ξ=19, N_ε=11	0.3197	0.2676	0.4230	0.6970	−0.0944	+0.40%
$N_h = 120$	0.3208	0.2678	0.4223	0.6954	−0.0945	+0.35%
$N_h = 360$	0.3206	0.2676	0.4230	0.6967	−0.0960	+0.38%
$N_e = 40$	0.3177	0.2675	0.4235	0.6971	−0.0905	+0.46%
$N_e = 100$	0.3207	0.2678	0.4229	0.6967	−0.0957	+0.39%
$N_l = 30$	0.3196	0.2681	0.4234	0.6973	−0.0940	+0.40%
$N_l = 60$	0.3192	0.2671	0.4240	0.6960	−0.0924	+0.46%
$N_\xi = 15$	0.3259	0.2672	0.4210	0.6982	−0.0943	+0.46%
$N_\xi = 25$	0.3145	0.2676	0.4240	0.6978	−0.0918	+0.49%
$N_\varepsilon = 7$	0.3197	0.2676	0.4230	0.6970	−0.0943	+0.41%
$N_\varepsilon = 15$	0.3198	0.2676	0.4230	0.6970	−0.0944	+0.40%
uniform h-grid, $N_h = 240$	0.3230	0.2674	0.4145	0.6896	−0.1001	+0.30%
uniform h-grid, $N_h = 480$	0.3208	0.2679	0.4212	0.6951	−0.0957	+0.36%
all dimensions refined	0.3152	0.2671	0.4252	0.6965	−0.0899	+0.52%

Notes: Each row changes one discretization dimension from the production grid, holding the calibration fixed. Across all 14 configurations the welfare effect spans 0.22 percentage points ([+0.30%, +0.52%]), the luck share spans 0.0107, and the cross-Atlantic output gap is negative throughout. The production human-capital grid is power-spaced, $h_j = h_{\min} + (h_{\max} - h_{\min})(j/(N_h - 1))^2$; the two uniform-grid rows show what a linear grid delivers at the same and at twice the number of nodes. Source: author's calculations.

Table 9 varies each discretization dimension in turn around the production grid, holding the calibration fixed, and reports the moments the conclusions rest on. Across all fourteen configurations the welfare effect spans 0.22 percentage points and never changes sign, the U.S. luck share spans about one percentage point, and the cross-Atlantic output gap is negative in every case. Two rows are diagnostic rather than robustness checks, namely a uniform h-grid at the production node count, and even at twice that count, still has not recovered what the power-spaced grid delivers. The shock discretization is immaterial, changing the welfare number by roughly a thousandth of a percentage point between seven and fifteen nodes. That is worth recording, because it is the dimension a reader might expect to matter most in a model whose central object is an earnings shock.

Two caveats bound what this establishes. First, 0.22 percentage points is small in absolute terms but is half the size of the 0.40 point estimate. The welfare effect is sign-stable across all fourteen configurations, and its magnitude is pinned only to within a factor of about 1.6, not to three digits. Second, the table holds the calibration fixed, which is the appropriate design for isolating solver accuracy but is not a convergence study of the calibration itself; Table 8 is the exercise that speaks to that, and it shows the calibrated anchors still moving between the coarse and converged grids. What the two together support is that the production grid is on the flat

part of that mapping and the coarse grid is not.

5.3 What this does and does not imply

Two readings should be resisted. The exercise does not imply that this environment is unusually delicate, and it is not an argument that the antecedent literature's conclusions are wrong. On the second point the paper can be exact rather than insinuating, because the replication archive for Torul (2020) documents that paper's grids: Section 7.4 solves this environment on them and finds it lands close to the converged answer. The coarse configuration used here is materially coarser than that, and is a device for making a mechanism visible rather than a reconstruction of anyone's practice.

What the exercise does imply is narrower and practical. When a model's aggregate response to the policy instrument is non-monotone, a calibration target can place the economies near the turning point, and then the identified parameter, and not the computed value function alone, becomes a function of the discretization. The safeguard is correspondingly specific. Report the convergence of the *calibrated parameters* and of the untargeted moments, not only of the value function and the aggregate fixed point, and report where the calibrated economies lie relative to the turning point of the model's own aggregate response.

6 The Statistic in the Fragile Case

Section 5 measured both channels by brute force, with a two-by-two that requires calibrating twice. This section computes the statistic of Section 3 in the same model, first with the benefit of hindsight and then under the informational restriction an actual user faces.

6.1 Computing it for this model

Here θ is the pair of fiscal anchors, m each region's public share of education expenditure, and W the aggregate consumption-equivalent effect of the U.S.-to-European reform. Because each anchor is calibrated to its own region's moment, $\partial m/\partial\theta$ is diagonal to a close approximation and (2) evaluates directly.

Table 10: The calibration-amplification diagnostic, computed and validated

Object	Meaning	U.S. anchor	European anchor
Ingredients, from local derivatives at the calibrated point			
$\partial W/\partial\theta$	conclusion's sensitivity to the parameter	−15.3	106.9
$\partial m/\partial\theta$	how sharply the target identifies it	7.67	3.42
$\partial m/\partial g$	grid-sensitivity of the target	+0.1028	+0.0869
Implied shift in the calibrated parameter			
predicted $\Delta\hat{\theta}$	$-(\partial m/\partial\theta)^{-1}(\partial m/\partial g)$	−0.0134	−0.0254
actual $\Delta\hat{\theta}$	coarse → converged calibration	−0.0110	−0.0235
Channels, in percentage points of ζ_{agg}			
predicted calibration channel	from the formula	−2.51	
measured calibration channel	Table 8	−3.17	
measured solver channel	Table 8	−0.11	
amplification factor A (ex ante)	|predicted calib.|/|solver|	23	
realized ratio (ex post)	|measured calib.|/|solver|	30	

Notes: θ is the pair of fiscal anchors, m each region's public share of education expenditure, W the aggregate consumption-equivalent effect of the U.S.-to-Europe reform, and g the discretization. Derivatives in θ are central differences with step 0.004, evaluated midway between the coarse and converged calibrations because the step between them is large and the response is convex near the turning point. A is formed from the *predicted* calibration channel, so it is computable without ever building the two-by-two; the realized ratio, which needs the two-by-two, is larger. $\partial m/\partial g$ is the change in the target moment when the grid is refined at fixed parameters, which costs one extra solve. Source: author's calculations.

Three features of Table 10 stand out.

The prediction is good at the midpoint, and the midpoint is not the test. Using local derivatives and a single coarse solve, equation (2) predicts a calibration channel of −2.51 percentage points against the −3.17 measured by the full two-by-two, recovering about four-fifths of a channel thirty times the solver channel, and it gets the implied movement in the anchors close: −0.0134 and −0.0254 predicted against −0.0110 and −0.0235 actual. But those derivatives are taken midway between the coarse and the converged calibrations, and the converged calibration is the object the diagnostic exists to let a researcher skip. Section 6.2 runs the calculation the way a researcher actually could, and the answer is more mixed than this paragraph suggests. The residual is what one expects of a first-order approximation over a step of this size, and it errs toward understatement because the response is convex near the turning point. Understatement is the safer direction for a screen, so A should be read as a lower bound on the calibration channel, which is how it is reported throughout.

The amplification is concentrated where the calibration mapping is flattest. Read the two columns against each other. The welfare conclusion barely responds to the U.S. anchor ($\partial W/\partial\theta = -15.3$) and responds strongly to the European one (+106.9). At the same time the European moment identifies its anchor *less* sharply than the U.S. moment does (3.42 against 7.67), because at a public financing share near 87% the model is close to saturation and the share responds sluggishly to further fiscal scale. The parameter the conclusion depends on most is the one the calibration pins down least, and the ratio $|\partial W/\partial\theta|/|\partial m/\partial\theta|$ is more than fifteen times larger for the European anchor than for the U.S. one.

That pairing, rather than the identity of the anchor, is the mechanism, and Section 6.2 supplies the check. Evaluated at the coarse calibration instead, the two anchors trade places. The conclusion is then far more sensitive to the U.S. instrument (−190.2 against +105.7), and it is the U.S. moment that identifies its anchor less sharply (4.75 against 5.77). The same sentence remains true with the labels swapped, and the ratio is again

largest for whichever anchor is simultaneously the most consequential and the least sharply pinned. The saturation story explains why the European moment is flat at the converged geometry; it is not what makes A large, and a reader should not take the mechanism to be a fact about European education spending. What makes A large is the product in equation (3), and a target moment near a boundary of its range is only one of the ways to be flat in the parameter.

Which values of A are benign follows from what it compares. Because it is a ratio of two channels, so $A \ll 1$ says the solver channel dominates and a conventional convergence study is the right check. $A \approx 1$ says both matter. $A \gg 1$, as here, says a value-function convergence study is close to uninformative about the reliability of the reported conclusion. The diagnostic does not require knowing the true parameter or the converged answer; it requires only the local geometry at whatever calibration a paper has, plus one solve at a finer grid. Section 7.1 runs it in a model that lands at the other end of that scale.

6.2 Running the diagnostic the way it is advertised

The claim that A costs a handful of solves at an existing calibration is only worth making if the calculation can be done by a researcher who has the coarse calibration and nothing else. Table 10 does not meet that standard, because its derivatives are evaluated midway between the two calibrations, which requires knowing where the converged one lands. This section redoes everything at the coarse calibration alone.

The protocol is exact about what is available. The Jacobian $\partial m/\partial\theta$ is a by-product of the calibration already run. $\partial W/\partial\theta$ costs two coarse-grid solves per parameter. One solve at the fine grid, holding parameters at their coarse values, supplies $\partial m/\partial g$ and $\partial W/\partial g$ together, so the solver channel is measured from that same solve rather than imported from the two-by-two, and the two-by-two is never built.

Table 11: The diagnostic run ex ante, and under both decomposition orderings

	Ex ante (coarse calibration only)	Midpoint (both endpoints)	Measured (two-by-two)
Panel A: ingredients and prediction			
$\partial W/\partial\theta$, U.S. anchor	−190.2	−15.3	—
$\partial W/\partial\theta$, European anchor	105.7	106.9	—
$\partial m/\partial\theta$, U.S. / European	4.75 / 5.77	7.67 / 3.42	—
predicted calibration channel (pp)	**+1.74**	−2.51	−3.17
solver channel (pp)	−0.11	−0.11	−0.11
amplification A	**16.2**	23	30
Panel B: both orderings of the two-by-two, in pp			
solver channel	−0.11 (at coarse)	−0.18 (at fine)	−0.15 (Shapley)
calibration channel	−3.17	−3.25	−3.21
implied A (midpoint numerator)	23	14	17
Panel C: which parameters carry the calibration channel, in pp			
two fiscal anchors alone		−3.35	
two laissez-faire parameters alone		−1.19	
all four together		−3.17	
interaction		+1.37	

Notes: Panel A. The ex-ante column uses only the coarse calibration and one solve at the fine grid, which supplies $\partial m/\partial g$ and $\partial W/\partial g$ together; it never touches the converged calibration. The midpoint column is Table 10. A compares the predicted calibration channel with the solver channel in the same column. Panel B. Neither ordering of a two-way decomposition is canonical, so both are reported along with their Shapley average, which is exact and additive. Panel C. The calibration channel is decomposed between the two fiscal anchors and the two laissez-faire parameters, each moved from its coarse to its converged value on the coarse grid; the two partial effects overlap, hence the large interaction. Source: author's calculations.

The statistic identifies the problem without predicting its size. Restricted to the two fiscal instruments, the ex-ante statistic is $A = 16$. It says correctly, on information a researcher would have, that the calibration channel is more than an order of magnitude larger than the solver channel, so a convergence study at fixed parameters is close to uninformative. That is the question A is asked and it answers it. The *signed* prediction fails. This version of the formula returns +1.74 percentage points where the truth is −3.17, getting the direction wrong.

The reason is visible in Panel A and is the same convexity the midpoint evaluation was introduced to dodge. At the coarse anchors the conclusion is violently sensitive to the U.S. instrument, $\partial W/\partial\theta = -190$, against −15 at the midpoint; the U.S. term then dominates and carries the wrong sign over a step that crosses the turning point. A first-order extrapolation across a turning point is not reliable, and no amount of care in computing the derivatives repairs that.

This costs the paper a claim and leaves the one it needs. It costs the claim that A forecasts the calibration channel. Over a large step that is false, and Table 11 is the evidence. What survives is the claim the paper actually needs: A is a *screen*. Under every evaluation point and every decomposition ordering, whether ex ante, at the midpoint, at either endpoint of the two-by-two, or under the Shapley average, A lies between 14 and 30. The reading is never in doubt, and it is that reading, not the point estimate, that tells a researcher what to do next. Read that way the diagnostic is more useful than a forecast would be, because its recommendation is unambiguous. When A is large, re-calibrate at the finer grid rather than trusting an extrapolation. A tells you that you must; it does not tell you what you will find.

The Aiyagari economy of Section 7.1 completes the picture. There the calibrated parameter barely moves,

the linearization is nearly exact, and the formula predicts the channel to eight percent. The signed prediction is accurate exactly where the calibration channel is small and inaccurate where it is large. The formula is worse off for that; the screen is not.

Neither ordering of the decomposition is canonical, so Panel B reports both. Measuring the solver channel at the coarse calibration gives −0.11 and the calibration channel −3.17; measuring at the converged calibration gives −0.18 and −3.25. Their average is the Shapley (1953) value of the two-player game whose players are the grid and the calibration. It is exact and additive, and gives −0.15 and −3.21. The paper quotes the first ordering because it is the one an ex-ante user occupies, standing at the coarse calibration, but nothing turns on the choice, since the calibration channel is between twenty and thirty times the solver channel under all three.

The statistic should be computed over the whole calibrated vector rather than the anchors alone. Section 3 defines θ as every internally calibrated parameter, and everything above applies the formula to the two fiscal anchors alone. That is narrower than the theory it illustrates, and Panel C is the reason it matters, because the laissez-faire parameters are not passengers. Table 12 therefore redoes the calculation over $\theta = (\mu_\xi, \phi, \bar{T}/\bar{Y}^{\text{US}}, \bar{T}/\bar{Y}^{\text{EU}})$ and the four moments that pin them, with the full 4×4 Jacobian, on the same ex-ante information set.

Table 12: The diagnostic over the full internally calibrated vector

Panel A: the Jacobian $\partial m/\partial\theta$, rows are moments and columns parameters	μ_ξ	ϕ	$\bar{T}/\bar{Y}^{\text{US}}$	$\bar{T}/\bar{Y}^{\text{EU}}$
$\bar{H}(0,0)$	+3.682	−0.212	+0.000	+0.000
$L(0,0)$	−0.037	−0.071	+0.000	+0.000
m^{US}	−0.940	−0.008	+4.748	+0.000
m^{EU}	−2.116	+0.650	+0.000	+5.772
Panel B: the remaining ingredients				
$\partial W/\partial\theta$	−263.1	+9.9	−190.2	+105.7
$\partial m/\partial g$	−0.0826	−0.0103	+0.0825	+0.0857
implied $\Delta\hat{\theta}$	+0.0137	−0.1518	−0.0149	+0.0073
actual $\Delta\hat{\theta}$	+0.0125	−0.0949	−0.0110	−0.0235
Panel C: channels, in percentage points				
predicted calibration channel, full θ				**−1.51**
the same over the two fiscal anchors only				+1.74
measured, from the two-by-two				−3.17
solver channel				−0.11
amplification A_{full}				**14.0**
A over the fiscal anchors only				16.2
condition number of $\partial m/\partial\theta$				88

Notes: θ is every internally calibrated parameter and m the four moments that pin them: the two laissez-faire targets and the two education-financing shares. All derivatives are central differences at the coarse calibration on the coarse grid; $\partial m/\partial g$ and $\partial W/\partial g$ come from one solve at the fine grid, so the information set is the ex-ante one of Table 11. The top-right 2×2 block of the Jacobian is exactly zero because the laissez-faire moments do not depend on the fiscal anchors, so the calibration is recursive; the bottom-left block is not, because the laissez-faire parameters move the financing shares. Source: author's calculations.

The structure of the Jacobian is generic to a calibration done in stages, and repays a sentence. The top-right 2×2 block is exactly zero, because the laissez-faire moments are computed at $\bar{T}/\bar{Y} = \tau_p = 0$ and cannot depend on the fiscal anchors. The bottom-left block is not zero, because μ_ξ and ϕ move the whole economy and with it the education-financing shares. The calibration is therefore recursive rather than separable, which is why

solving the two blocks in sequence is legitimate but treating the second block in isolation, as Section 6.1 does, is not. The matrix is well conditioned, with a condition number of 88, so the inversion in equation (2) is innocuous here, which is not something to assume without looking.

The result cuts two ways. The recommendation does not change. $A_{\text{full}} = 14$ against 16 over the anchors alone, both an order of magnitude above one, so a reader who ran the narrow version would have reached the same decision. But the wider version is the better calculation, and not only on grounds of consistency with the theory. It gets the *sign* of the calibration channel right, -1.51 percentage points against the -3.17 measured, where the anchors-only version returned $+1.74$ and pointed the wrong way. The reason is visible in Panel B: the conclusion is enormously sensitive to μ_ξ ($\partial W/\partial\theta = -263$), and the grid moves the laissez-faire human-capital moment by -0.083, so omitting that parameter discards a term large enough to flip the sum. The magnitude is still understated by half, which is the same first-order limitation Table 11 documents and which no widening of θ repairs.

The practical reading is that θ should be the full vector whenever the calibration is done in stages, which is to say almost always. Restricting attention to the parameters a paper happens to find interesting is the mistake this section was written to catch, and it cost a sign here.

Panel C splits the calibration channel between the two fiscal anchors and the two laissez-faire parameters (μ_ξ, ϕ). Moving the anchors alone accounts for -3.35 percentage points and moving (μ_ξ, ϕ) alone for -1.19, against -3.17 jointly. The two overlap heavily, and the laissez-faire parameters are not passengers. That matters because, as Section 4.4.2 states, their targets are not data. A reader who wants to attribute the whole channel to the education-financing moment cannot; the discretization moves every internally calibrated parameter, and the conclusion responds to all of them.

7 The Statistic in the Workhorse Model

7.1 A second application: a model where the diagnostic says do not worry

A diagnostic that signals alarm everywhere has no discriminatory power. The question is whether A separates environments in which the calibration channel dominates from those in which it does not, and answering it requires a second model chosen for being canonical rather than for being convenient. I use the workhorse of this literature, and none of the code above.

The environment is the incomplete-markets economy of Aiyagari (1994): households with constant relative risk aversion facing persistent idiosyncratic labor-productivity risk, a single risk-free asset, a no-borrowing constraint, and a competitive representative firm. The calibrated parameter is the discount factor β; the target moment is the capital–output ratio, which is what that literature calibrates β to; the discretization is the asset grid; and the conclusion is the consumption-equivalent effect of a proportional labor-income tax rebated lump-sum, evaluated at the status-quo cross-section. The exercise is the one Section 5 ran on the education model. Solve and calibrate at a coarse grid and at a converged one, measure both channels with a two-by-two, and compare that against what equation (2) predicts from local derivatives alone.

Table 13: The same diagnostic in the workhorse model: Aiyagari (1994)

Object	Meaning	Education finance (Section 6.1)	Aiyagari (this table)
What is being calibrated			
θ	the calibrated parameter	fiscal anchors	discount factor β
m	the target moment	public education share	capital–output ratio
W	the reported conclusion	consumption-equivalent effect of a fiscal reform	
Ingredients			
$\partial m/\partial\theta$	how sharply the target identifies θ	3.42–7.67	22.5
$\partial m/\partial g$	grid-sensitivity of the target	+0.09–+0.10	−0.114
$\Delta\hat{\theta}$ predicted	implied parameter shift	−0.013, −0.025	+0.0051
$\Delta\hat{\theta}$ actual	coarse → fine calibration	−0.011, −0.024	+0.0051
Channels, in percentage points of the conclusion			
solver channel	measured	−0.11	−0.79
calibration channel	predicted by (2)	−2.51	−0.12
calibration channel	measured	−3.17	−0.13
amplification A (ex ante)	|predicted calib.|/|solver|	**23**	**0.15**
realized ratio (ex post)	|measured calib.|/|solver|	30	0.16
Verdict		**check the calibration**	**check the solver**

Notes: The right-hand column applies the diagnostic of equation (3) to a standard incomplete-markets economy: CRRA utility with $\sigma = 2$, a persistent idiosyncratic labor-productivity process ($\rho = 0.90$, $\sigma_\epsilon = 0.20$, seven Tauchen nodes), Cobb–Douglas production with capital share 0.36 and depreciation 0.08, and a no-borrowing constraint. The discount factor is calibrated to a capital–output ratio of 3; the conclusion W is the consumption-equivalent effect, evaluated at the status-quo cross-section, of a 20% proportional labor-income tax rebated lump-sum. The coarse grid is 40 uniformly spaced asset nodes and the fine grid 600 power-spaced nodes on the same interval; refining further moves the conclusion by under 0.01 percentage points. Derivatives in this column are evaluated at the converged calibration; evaluated at the coarse calibration alone, the strictly ex-ante information set, the ingredients are $\partial m/\partial\theta = 21.8$, $\partial m/\partial g = -0.111$, and $A = 0.155$, with the predicted channel −0.123. Left-hand column reproduced from Table 10. Source: author's calculations.

Table 13 sets the two economies side by side. The verdict reverses, and it reverses for the reason the formula names. Three implications follow.

As a screen the diagnostic is right in both models; as a forecast only here. In the Aiyagari economy equation (2) predicts a calibration channel of −0.12 percentage points against −0.13 measured, an error of eight percent, and it predicts the shift in the calibrated discount factor to four decimal places: +0.0051 against +0.0051 actual. Those derivatives are evaluated at the converged calibration; re-running the calculation strictly ex ante, at the coarse calibration alone, returns $A = 0.155$ with a predicted channel of −0.123 against the same −0.127 measured, so both the verdict and the accuracy survive the informational restriction of Section 6.2. In the education model the same formula recovers four-fifths of the channel when evaluated at the midpoint; run the way a researcher could, as Section 6.2 shows, it gets the magnitude wrong by half. It is more accurate here because the calibrated parameter barely moves, so a first-order approximation to its movement is nearly exact. That is the ordinary case, and the diagnostic is at its best there, which is where most researchers will find themselves.

The two verdicts differ where the formula says they should. The raw derivatives are not comparable across the two models, since a ratio near three and a share near 0.7 are not measured in the same units. The comparison has to be made in elasticities, and both factors run the same way. Refining the grid moves the capital–output ratio by 3.8% of its own value and the two education shares by 10% and 15% of theirs, so the education model is the more grid-sensitive of the two rather than the less. And the capital–output ratio is a far sharper instrument for

the discount factor than either education share is for its anchor: a one percent change in β moves the target by 7.0 percent, against 1.1 and 0.5 percent for the U.S. and European anchors. The two effects compound. A larger relative displacement is divided by a slope six to fifteen times flatter, and the product is what equation (3) punishes. Nothing about the absolute size of the numerical error distinguishes the two models; what distinguishes them is that displacement and flatness reinforce one another in one and not the other.

The practical readings are opposite, and both are actionable. In the Aiyagari economy the solver channel is −0.79 percentage points of a total −0.91: refining the asset grid at fixed parameters recovers most of the discretization error, and conventional practice, a convergence study over the household problem, is the right check and a sufficient one. In the education model the same check recovers three percent of the error and is close to no evidence at all. A researcher who ran A in the first model would rightly stop worrying; one who ran it in the second would learn that the check they were about to perform was not the one their conclusion needed. That is what a diagnostic is for, and it is why the interesting output of equation (3) is not the alarm but the all-clear.

A is a ratio, and ratios need a scale. A referee's natural question is whether $A = 0.15$ is an artifact of how coarse the coarse grid was. It is not, but the answer is more interesting than a yes or no. Table 14 recomputes A in the Aiyagari economy against progressively better coarse grids.

Table 14: A is informative while the discretization error is material

Coarse grid	$\partial m/\partial g$	Solver channel (pp)	Calibration channel (pp, predicted)	A
40 uniform	−0.1139	−0.781	−0.117	0.15
100 uniform	−0.0465	−0.286	−0.048	0.17
200 uniform	−0.0207	−0.090	−0.021	0.24
100 power 3	−0.0028	−0.006	−0.003	*0.52*
200 power 3	−0.0006	+0.000	−0.001	*3.76*

Notes: Aiyagari economy. Each row pairs the same converged grid (600 power-spaced nodes) with a different coarse grid, holding the discount factor at its converged calibration. Both channels shrink as the coarse grid improves, but the solver channel shrinks faster, so A rises. The last two rows, set in italics, compare two grids that both already resolve the problem: there the two channels are of order a thousandth of a percentage point and A is a ratio of negligible quantities. A should be read alongside the magnitude of the total effect it decomposes. Source: author's calculations.

Both channels shrink as the coarse grid improves, but the solver channel shrinks faster, so A rises from 0.15 to 3.76. The last figure is not evidence that the calibration channel matters after all. It is a ratio of two quantities that are by then of order a thousandth of a percentage point. At 200 power-spaced nodes the discretization has no material effect on the conclusion through either route, and asking which of two negligible channels is larger has no content. Over the range where the discretization does move the answer, the first three rows, where the solver channel runs from 0.78 to 0.09 percentage points, A stays between 0.15 and 0.24 and the conclusion is stable.

The protocol this implies is short, and it applies to the education model too, where the total discretization effect is 3.36 percentage points and squarely material. Report A next to $|dW/dg|$, the size of the effect it decomposes. If that magnitude is negligible, the discretization is not a problem and A is not worth interpreting. If it is material, A says which of the two routes to spend effort on, and that is precisely when the answer is not obvious, because the visible diagnostics speak only to one of them.

Two further limits bound it. The Aiyagari conclusion is a steady-state comparison rather than a transition, so W there is a simpler object than the transition-based welfare number of Section 4.6.2; the diagnostic is indifferent

to which, since it treats W as a black box. And two models are two models. What the pair establishes is that A varies by about a factor of a hundred across ordinary quantitative environments, that the variation is driven by the flatness of the calibration mapping rather than by solver quality, and that a formula costing a handful of solves gets the answer right in both.

7.2 What makes A large: a controlled experiment

Two environments with opposite verdicts show that A *can* discriminate. They do not show that it discriminates because of the identification slope, because the two differ in every other respect. One has a non-monotone aggregate response, a transition-based welfare object and a target near a boundary; the other has none of these. A reader is entitled to suspect that A is tracking the non-monotonicity rather than the slope.

The suspicion is testable, and cheaply, because the Aiyagari economy admits several natural targets for the same parameter. Table 15 holds the environment, the discretization, the pair of grids, the conclusion W and the true discount factor all fixed, and varies only the moment β is calibrated to.

Table 15: One environment, six target moments: A spans two orders of magnitude

Target moment for β	m^*	$\partial m/\partial\beta$	$\partial m/\partial g$	implied $\Delta\hat{\beta}$	A
wealth-to-income	3.0000	+22.462	−0.1139	+0.0051	0.15
capital-output ratio	3.0000	+22.462	−0.1139	+0.0051	0.15
aggregate capital	5.5655	+65.115	−0.3336	+0.0051	0.15
wealth Gini	0.5467	−0.423	+0.0182	+0.0429	1.27
top 10% wealth share	0.3565	+0.487	−0.0902	+0.1853	5.48†
fraction at the constraint	0.0248	−0.160	−0.1990	−1.2420	36.76†

Notes: The Aiyagari economy of Table 13, at the same converged calibration, the same pair of grids, and the same conclusion W. Only the moment the discount factor is calibrated to changes. Because $\partial W/\partial\beta$ and $\partial W/\partial g$ are therefore common to every row, A is proportional to the implied shift in $\hat{\beta}$ by construction; what the table shows is not that proportionality but the range it produces. The capital–output ratio and its two rescalings are the conventional choice. The fraction of households at the borrowing constraint combines a flat slope in β with a large grid sensitivity, which is the configuration equation (3) punishes. Source: author's calculations.

A runs from 0.15 to 36.8 inside a single model with a monotone, well-behaved aggregate response and no turning point anywhere near the calibration. The economic environment is identical in every row. What changed is which statistic of the same stationary distribution the calibration was asked to match.

The ordering is exactly what equation (3) says it should be. The capital–output ratio, the conventional choice, has a steep slope in β (22.5) and gives $A = 0.15$; so do its two rescalings, which is a useful check that the diagnostic is invariant to the units of the target. The wealth Gini is nearly flat in β (-0.42) and gives $A = 1.3$. The top-decile wealth share is flatter still relative to its grid sensitivity and gives $A = 5.5$. The fraction of households at the borrowing constraint, a boundary statistic that is flat in β and badly resolved on a coarse asset grid, gives $A = 37$ in a model where nothing else is fragile at all.

Three qualifications bound the reading. Since W and the pair of grids are held fixed, $\partial W/\partial\beta$ and $\partial W/\partial g$ are the same in every row, so A is proportional to the implied shift in $\hat{\beta}$ by construction. That proportionality is arithmetic rather than evidence. The substantive result is the range, which is a property of the environment and not of the formula.

These are also the diagnostic's predictions, not measured two-by-twos. Verifying them would mean re-

calibrating to each target at both grids, which is the expensive exercise A exists to triage.

The third qualification matters most for how the two large entries should be read. For the top-decile share and the constrained fraction, the implied shift would carry the discount factor to 1.12 and -0.31, outside the range in which it is even defined. The linear extrapolation has broken down there. That is not a failure of the screen so much as an unambiguous reading from it, since a first-order step that leaves the parameter space is itself the signal that the target is unusable for calibration. The reported 5.5 and 37 therefore carry no information about the size of the calibration channel, only about the unsuitability of those two targets. The three admissible rows are the ones whose magnitudes carry information, and they already span an order of magnitude.

What the experiment establishes is the claim Section 7.1 could only assert. The calibration channel is not a property of a model. It is a property of the pairing between a model, a conclusion, and a target moment, and a researcher chooses the last of those. That makes equation (3) a design tool as much as a diagnostic. Among moments that identify a parameter, the ones with steep slopes and low grid sensitivity are the ones that keep discretization error out of the answer, and A prices that choice before the calibration is run.

Nor are the fragile rows of Table 15 hypothetical choices: the wealth Gini and the top wealth shares are exactly the moments the quantitative wealth-inequality literature asks its calibrations to match (Castañeda et al., 2003; De Nardi, 2004). The next subsection tests the claim directly, by running the screen in further environments under both kinds of target.

7.3 The screen in further environments

The controlled experiment varies the target inside one model. This subsection varies the model, running the identical protocol — the coarse calibration, its Jacobian, two solves per parameter, one finer solve — in three further environments. Two are implemented from their published specifications, as the Aiyagari application is: an exchange economy in the manner of Huggett (1993), where the discount factor is calibrated to the equilibrium risk-free rate and the conclusion is the welfare effect of tightening the credit limit, and a life-cycle economy in the tradition of Huggett (1996), where the discount factor is calibrated to the capital–output ratio and the conclusion is the welfare effect of halving the social-security replacement rate. The third is not an implementation but the actual replication archive of Conesa et al. (2009), compiled unmodified apart from a root-finder shim and output hooks, and screened *at its published discretization*: the archive's discount factor reproduces the paper's capital–output target of 2.7 on its 101-node asset grid, and the conclusion is the newborn consumption-equivalent gain of the tax reform the paper reports as optimal, which evaluates to 1.33% there, against the roughly 1.3% the paper reports. Table 16 collects the results.

Panel A is the conventional-target row of each environment, and the pattern of Section 7.1 repeats. Where the discount factor is pinned by the risk-free rate or the capital–output ratio, the slope $\partial m/\partial\theta$ is steep, A is between 0.01 and 0.15, and the screen's signed prediction of the calibration channel is validated by the measured two-by-two to within a few thousandths of a percentage point. The life-cycle row makes the sharpest version of the point: there the coarse grid moves the conclusion by 1.4 percentage points — a large discretization effect — and essentially all of it travels through the solver, so the fixed-parameter convergence study that quantitative papers already run is exactly the right check and would catch it. The Conesa et al. (2009) row is the screen doing its everyday job on a published configuration: both channels are below two hundredths of a percentage point on a conclusion of 1.33%, so the published grid resolves the published conclusion, and the A of 2.2 is a ratio of negligible quantities, which is precisely the case the protocol of Section 7.1 instructs a reader to classify as

an all-clear by reading A next to $|dW/dg|$. The antecedent of the education model supplies a second published configuration, and it passes too. The replication archive of Torul (2020) documents its grids in full — 100 uniform human-capital nodes whose spacing, over the region the stationary distribution occupies, is more than three times finer than the coarse configuration of Section 5, together with denser education, labor and ability grids. Solved and re-calibrated on that documented discretization, this paper's environment recovers fiscal anchors within about one percent of the converged ones, reproduces the cross-Atlantic output ranking, and values the reform at +0.64% against +0.40% on the production grid; the pathology of Section 5 does not appear there. Two published configurations, two passes, and in both cases the reporting that made the check possible cost the authors a node count, a spacing rule, and a tolerance.

Panel B re-screens two of the same environments with the discount factor calibrated to a distributional statistic instead, and the fragility returns in a form starker than Table 15 showed. Calibrated to the fraction of agents at the credit limit, the implied parameter step leaves the admissible range entirely, the same alarm as the two italicized rows of that table. Calibrated to the wealth Gini, the alarm is of a new kind: the coarse grid puts the Gini at 0.49 where the converged value is 0.82, and no discount factor near the calibrated one reproduces the coarse value on the fine grid at all. A calibration to a badly resolved distributional moment can fail not by delivering distorted parameters but by chasing a number the accurately solved model cannot produce. Matching wealth concentration is what Panel B's targets are asked to do in the literature, so this is the configuration in which running the screen earns its keep.

Three limits bound the exercise. The two Huggett-style environments are implementations from published specifications rather than the authors' code, and the annualized parameterization of the exchange economy is stated in the table notes. Every refinement here varies the asset grid alone, along each environment's own spacing rule. And the Conesa et al. (2009) conclusion is evaluated at the reform the paper reports as optimal with the deduction parameter fixed at the midpoint of the archive's search grid, a choice that is held fixed across grids and so does not touch the decomposition. The scripts, the patched archive, and the computed output are part of the replication package.

Table 16: The screen in further environments

Panel A: conventional targets Environment	θ	Target m	A	Calibration channel pred.	meas.	Solver meas.	Verdict
Huggett (1993) exchange econ.	β	risk-free rate	0.15	−0.033	−0.032	0.218	solver
Life cycle (Huggett 1996 trad.)	β	capital–output ratio	0.01	−0.017	−0.017	1.447	solver
Conesa et al. (2009) archive	β	capital–output ratio	2.18	−0.012	−0.019	−0.006	resolved
Panel B: alternative targets (screen only) Environment and target			A	implied $\Delta\hat{\theta}$		reading	
Huggett (1993), $\beta \leftarrow$ fraction at the credit limit	1124.2	30.3	implied step leaves $(0,1)$: target unusable				
Life cycle, $\beta \leftarrow$ wealth Gini	2.8	−0.107	coarse-grid Gini unattainable at the fine grid				

Notes: Each row runs the protocol of Section 3: A and the predicted calibration channel use only the coarse (for the Conesa et al. row, the *published*) calibration and one solve at a finer grid; measured channels come from re-calibrating at the finer grid and building the two-by-two of Section 5. Channels are in percentage points of each environment's conclusion. Huggett (1993): annual parameterization of the published specification; W is the consumption-equivalent effect of tightening the credit limit by half, stationary comparison; grids 30 uniform vs. 600 power-spaced asset nodes. Life cycle: W is the effect of halving the social-security replacement rate, stationary comparison; grids 40 uniform vs. 600 power-spaced nodes; both fine configurations verified converged at 1,200 nodes. Conesa et al. (2009): the authors' replication archive, compiled unmodified except for a root-finder shim and output hooks, screened *at its published discretization* (101 asset nodes) against a refinement to 301 nodes along the archive's own spacing rule; θ is the archive's discount factor (1.00093, which reproduces the paper's capital–output target of 2.7 on the published grid), and W is the newborn consumption-equivalent gain of the tax reform the paper reports as optimal ($\tau_k = 0.36$, $a_0 = 0.23$, $a_1 = 7.0$, the midpoint of the archive's search grid), which evaluates to 1.33% on the published grid, against the paper's reported gain of about 1.3%. In Panel B the same environments are re-screened with β calibrated to a distributional statistic instead; the life-cycle wealth Gini is 0.49 on the coarse grid against a converged 0.82, and no β near the calibrated value reproduces the coarse value on the fine grid. Source: author's calculations.

8 Conclusion

Discretization error reaches a calibrated model's conclusions by two routes, and the profession checks one of them. The direct route runs through the solution of the agent's problem, and that is what tolerances, residuals and convergence studies measure. The indirect route runs through the calibration, because the moment used to pin down a parameter is computed from the same discretized model, so the parameter inherits the error and passes it along to whatever the paper concludes. The implicit function theorem separates the two, and their ratio is a number a researcher can compute at the calibration already in hand, at a cost of two extra solves per calibrated parameter and one solve at a finer grid.

Whether that ratio is large depends on the model, and this paper answers the question in five environments and then inside one of them. In the education-finance model solved and calibrated at two resolutions, the indirect route accounts for ninety-five percent of the difference in the reported welfare effect and the direct route for three, and the screening statistic is $A = 14$. Narrower and more generous versions of the same arithmetic give sixteen, twenty-three and thirty, and none of them changes the instruction. In the incomplete-markets economy of Aiyagari (1994), with the discount factor set to a capital–output ratio, $A = 0.15$: the solver carries seven-eighths of the error, refining the asset grid at fixed parameters is the right check, and the formula predicts the small residual to within eight percent. Calibrating that same discount factor to the share of households at the borrowing constraint instead raises the statistic by two orders of magnitude, in the same economy on the same grids, so the channel belongs to the pairing of a model, a conclusion and a target rather than to the model. The same protocol, run in the exchange economy of Huggett (1993), a life-cycle economy in the tradition of Huggett (1996), and the replication archive of Conesa et al. (2009) at its published discretization, reproduces the division: small A and

validated signed predictions under the conventional aggregate targets, an all-clear for the published configuration, and, under distributional targets, either a parameter step that leaves the admissible range or a coarse-grid target value that the accurately solved model cannot produce at any nearby parameter.

The economics behind the contrast is ordinary. Measured in elasticities, refining the grid displaces the education-financing shares by two to four times as much as it displaces the capital–output ratio, and the education shares respond to their anchors six to fifteen times less sharply than the capital–output ratio responds to the discount factor. A larger displacement divided by a flatter mapping is what the formula punishes, and a target moment close to the boundary of its range is generically flat. It is the flatness, not the size of the numerical error, that turns a modest discretization into a large error in the answer.

The implication for practice is a screening rule rather than a correction. A convergence study over the value function is not evidence that a calibrated conclusion is resolution-independent, and at $A = 14$ it is close to no evidence at all; at $A = 0.15$ it is the right evidence. The other half of what makes the statistic worth computing is that its usual output is an all-clear, and an all-clear bought with a handful of solves is worth having. What the statistic does not do is stand in for the re-calibration it flags. Over a step that crosses a turning point the signed prediction fails, and that is precisely the regime a large A identifies. So compute A over the full calibrated vector, not over the parameters that happen to be interesting. If it and the effect it decomposes are both negligible in the units the conclusion is reported in, the discretization is not a problem. If either is material and A is well above one, re-calibrate at the finer grid and report what changes, because no local formula will say in advance what that will be. The cheapest place to avoid the problem is earlier still, in the choice of target moment. Among the statistics that pin a parameter down, prefer those with steep slopes and low grid sensitivity.

What a paper should report follows from the same logic. Convergence of the calibrated parameters and of the untargeted moments, not only of the value function and the aggregate fixed point; and, where the model's aggregate response to the policy instrument is non-monotone, where the calibrated economies sit relative to the turning point. Neither is expensive once the household problem is solved efficiently, and Proposition 1 gives conditions under which it can be, in a class of models considerably wider than this one.

On the substantive question that motivated the exercise, the converged model says a European-style fiscal regime is worth close to nothing to the United States. The number is small enough that the economic sensitivities dominate it. It reverses to -2.5% when non-labor income measured on the Panel Study of Income Dynamics enters household budgets, and it moves with where the fiscal anchors sit, so the interval those extensions span is the better summary. What survives independently of the interval is narrower and harder to dismiss. On a uniform sixty-node grid that hit every calibration target and converged by its own tolerances, the education-share moment selected anchors past the trough of the aggregate response, and the same exercise put the welfare gain at 3.8% with every household in favor. The tolerances were not the check that mattered.

Acknowledgements

This paper began as a study of the cross-Atlantic composition of earnings inequality. It became a paper about numerical accuracy because three anonymous referees and an associate editor at the *Review of Economic Dynamics* pressed hard on the earlier version's identification and its treatment of belief data, and following those objections led to the discovery reported here. I thank them, and I thank conference and seminar participants at EEA-ESEM 2013, PET 2014, ASSET 2016, the New Economic School, the University of Oslo, Cleveland State

University, Özyeğin University, Boğaziçi University, Sabancı University, Bilgi University, Middle East Technical University, TOBB Economics and Technology University, and the Research Department of the Central Bank of Türkiye. I am also grateful to Allan Drazen, Pablo D'Erasmo, Ethan Kaplan, John Shea, and John Haltiwanger for their valuable comments. All remaining errors are my own.

Declaration of the use of AI-assisted technologies

During the preparation of this work the author used Anthropic's Claude to assist with drafting and editing prose, with writing and debugging the Python code that produces the paper's computational results, and with checking the internal consistency of the reported numbers against that code's output. Every model, calibration, derivation and result was specified by the author; every number reported in the paper is emitted directly from executed code rather than transcribed, and the replication archive regenerates all of them. The author reviewed and edited all output and takes full responsibility for the content of the article.

Data and code availability

All empirical inputs are from publicly accessible sources and no proprietary data are used. The replication archive accompanying this submission contains every production script, the processed output behind each table and figure, and the raw OECD and World Values Survey source files, and will be deposited in a public repository with a citable identifier upon publication, in accordance with the Econometric Society's Data and Code Availability Policy. The Panel Study of Income Dynamics microdata used in Appendices B and D.1 are not redistributed; they are available from ICPSR through the PSID-SHELF release (Daumler et al., 2025), and the extraction and estimation scripts are included.

References

Aaronson, D. and French, E. (2004). The effect of part-time work on wages: Evidence from the social security rules. *Journal of Labor Economics*, 22(2):329–352.

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2022). Income and wealth distribution in macroeconomics: A continuous-time approach. *Review of Economic Studies*, 89(1):45–86.

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics*, 109(3):659–684.

Alesina, A. and Angeletos, G.-M. (2005). Fairness and redistribution. *American Economic Review*, 95(4):960–980.

Alesina, A., Stantcheva, S., and Teso, E. (2018). Intergenerational mobility and preferences for redistribution. *American Economic Review*, 108(2):521–554.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *Quarterly Journal of Economics*, 132(4):1553–1592.

Aruoba, S. B., Fernández-Villaverde, J., and Rubio-Ramírez, J. F. (2006). Comparing solution methods for dynamic equilibrium economies. *Journal of Economic Dynamics and Control*, 30(12):2477–2508.

Badel, A., Huggett, M., and Luo, W. (2020). Taxing top earners: A human capital perspective. *Economic Journal*, 130:1200–1225.

Bénabou, R. (2000). Unequal societies: Income distribution and the social contract. *American Economic Review*, 90(1):96–129.

Bénabou, R. (2002). Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica*, 70(2):481–517.

Bénabou, R. and Tirole, J. (2006). Belief in a just world and redistributive politics. *Quarterly Journal of Economics*, 121(2):699–746.

Birinci, S., Karabarbounis, L., and See, K. (2026). Why do americans no longer work so much more than non-americans? Working paper, Federal Reserve Bank of St. Louis.

Carroll, C. D. (2006). The method of endogenous gridpoints for solving dynamic stochastic optimization problems. *Economics Letters*, 91(3):312–320.

Carroll, D. R., Dolmas, J., and Young, E. R. (2021). The politics of flat taxes. *Review of Economic Dynamics*, 39:174–201.

Carroll, D. R., Luduvice, A. V. D., and Young, E. R. (2025). A note on aggregating preferences for redistribution. *Economics Letters*, 256:112535.

Castañeda, A., Díaz-Giménez, J., and Ríos-Rull, J.-V. (2003). Accounting for the U.S. earnings and wealth inequality. *Journal of Political Economy*, 111(4):818–857.

Caucutt, E. M. and Lochner, L. (2020). Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy*, 128(3):1065–1147.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the United States. *Quarterly Journal of Economics*, 129(4):1553–1623.

Conesa, J. C., Kitao, S., and Krueger, D. (2009). Taxing capital? Not a bad idea after all! *American Economic Review*, 99(1):25–48.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, 27(3):79–102.

Daruich, D. and Fernández, R. (2024). Universal basic income: A dynamic assessment. *American Economic Review*, 114(1):38–88.

Daumler, D., Friedman, E., and Pfeffer, F. T. (2025). Psid-shelf user guide and codebook, 1968–2021. Institute for Social Research, University of Michigan; ICPSR Open Data Release 2025-01 (DOI:10.7302/25205). Panel Study of Income Dynamics–Social, Health, and Economic Longitudinal File (PSID-SHELF), 1968–2021 beta.

De Nardi, M. (2004). Wealth inequality and intergenerational links. *Review of Economic Studies*, 71(3):743–768.

Den Haan, W. J. (2010). Comparison of solutions to the incomplete markets model with aggregate uncertainty. *Journal of Economic Dynamics and Control*, 34(1):4–27.

Fernández, R. and Rogerson, R. (1998). Public education and income distribution: A dynamic quantitative evaluation of education-finance reform. *American Economic Review*, 88(4):813–833.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., and Santos, M. S. (2006). Convergence properties of the likelihood of computed dynamic models. *Econometrica*, 74(1):93–119.

Guvenen, F. and Smith, A. A. (2014). Inferring labor income risk and partial insurance from economic choices. *Econometrica*, 82(6):2085–2129.

Heathcote, J., Perri, F., and Violante, G. L. (2010). Unequal we stand: An empirical analysis of economic inequality in the united states, 1967–2006. *Review of Economic Dynamics*, 13(1):15–51.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). Optimal tax progressivity: An analytical framework. *Quarterly Journal of Economics*, 132(4):1693–1754.

Holter, H. A., Krueger, D., and Stepanchuk, S. (2019). How do tax progressivity and household heterogeneity affect Laffer curves? *Quantitative Economics*, 10(4):1317–1356.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, 17(5–6):953–969.

Huggett, M. (1996). Wealth distribution in life-cycle economies. *Journal of Monetary Economics*, 38(3):469–494.

Judd, K. L. (1992). Projection methods for solving aggregate growth models. *Journal of Economic Theory*, 58(2):410–452.

Judd, K. L. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge, MA.

Kopecky, K. A. and Suen, R. M. H. (2010). Finite state Markov-chain approximations to highly persistent processes. *Review of Economic Dynamics*, 13(3):701–714.

Krueger, D., Perri, F., Pistaferri, L., and Violante, G. L. (2010). Cross-sectional facts for macroeconomists. *Review of Economic Dynamics*, 13(1):1–14.

Krusell, P. and Smith, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896.

MacQueen, J. and Porteus, E. L. (1975). Numerical methods and bounds in markovian decision processes. *Journal of Mathematical Analysis and Applications*, 49(2):421–442.

Maliar, L. and Maliar, S. (2014). Numerical methods for large-scale dynamic economic models. In Schmedders, K. and Judd, K. L., editors, *Handbook of Computational Economics*, volume 3, pages 325–477. Elsevier.

McKelvey, R. D. (1986). Covering, dominance, and institution-free properties of social choice. *American Journal of Political Science*, 30(2):283–314.

Organisation for Economic Co-operation and Development (2019). Public social spending by category, as a percentage of GDP. OECD Social Expenditure Database (SOCX), 2019 release.

Piketty, T. (1995). Social mobility and redistributive politics. *Review of Economic Studies*, 62(4):551–584.

Prescott, E. C. (2004). Why do Americans work so much more than Europeans? *Federal Reserve Bank of Minneapolis Quarterly Review*, 28(1):2–13.

Roemer, J. E. (1998). *Equality of Opportunity*. Harvard University Press, Cambridge, MA.

Santos, M. S. (2000). Accuracy of numerical solutions using the Euler equation residuals. *Econometrica*, 68(6):1377–1402.

Schütz, G., Ursprung, H. W., and Wößmann, L. (2008). Education policy and equality of opportunity. *Kyklos*, 61(2):279–308.

Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. W. and Tucker, A. W., editors, *Contributions to the Theory of Games, Volume II*, pages 307–317. Princeton University Press.

Stantcheva, S. (2021). Understanding tax policy: How do people reason? *Quarterly Journal of Economics*, 136(4):2309–2369.

Stiglitz, J. E. (1974). The demand for education in public and private school systems. *Journal of Public Economics*, 3(4):349–385.

Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters*, 20(2):177–181.

Torul, O. (2020). Education in a heterogeneous-agent economy: Revisiting transatlantic differences. *Journal of Human Capital*, 14(2):165–216.

Young, E. R. (2010). Solving the incomplete markets model with aggregate uncertainty using the Krusell–Smith algorithm and non-stochastic simulations. *Journal of Economic Dynamics and Control*, 34(1):36–41.