Online Appendix for "Sources of U.S. Wealth Inequality: Past, Present, and Future"

Joachim Hubmer, Per Krusell, and Anthony A. Smith, Jr.*

May 17, 2020

1 Computational appendix

1.1 Dynamic programming problem

The consumers' dynamic programming problem is solved by value-function iteration using Carroll (2006)'s endogenous grid-point method (EGM) on a grid for cash-on-hand and the persistent idiosyncratic shocks (β, p).

Unlike in the original Aiyagari (1994) model, the support of the ergodic wealth distribution is unbounded in this framework. We use a log-spaced grid with 100 points for cash-on-hand $(x_i)_{i=1}^{100}$ with a very large upper bound (one million times average wealth) to minimize the truncation error.[1] Cubic splines are used to interpolate the value function along the wealth dimension.

The grid for the persistent component of individual productivity $(p_j)_{j=1}^{17}$ is chosen to account for the long right tail in earnings. First, we chose the grid points as the 0.0001, 0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 0.925, 0.95, 0.975, 0.99, 0.999, ..., 0.99999999 quantiles of the unconditional (i.e., cross-sectional) p-distribution (which is a Gaussian). Second, we compute the corresponding grid in actual efficiency units of labor $(\psi(p_1), ..., \psi(p_{17}))$. Third, given that in the current period $p = p_j$ for $j = 1, ..., 17$, we use Gauss-Hermite quadrature to integrate over $p'|p$, the value of idiosyncratic productivity in the next period, when updating the value function. In doing so, we use linear interpolation in $\psi(p)$-space to evaluate the value function off the grid (the value function is much more non-linear in p-space than in $\psi(p)$-space).[2]

Regarding the discount factor, we choose the grid points $(\beta_m)_{m=1}^{15}$ as the Gauss-Hermite quadrature points of the unconditional (i.e., cross-sectional) β-distribution (this will turn out to be useful when integrating over the joint distribution to compute aggregate wealth). Again, when updating the value function, we integrate over $\beta'|\beta$ using Gauss-Hermite quadrature and linear interpolation in β-space.

*The authors' affiliations are, respectively, University of Pennsylvania; Institute for International Economic Studies, NBER, and CEPR; and Yale University and NBER.

[1] Alternatively, given that the Pareto tail has stabilized at some $\bar{x}$, one could in principle also impute the distribution for $x > \bar{x}$. However, this did not turn out to be necessary as the log-spaced grid—which works well as the curvature of the value function is high only close to the borrowing constraint—allows for selecting a very large upper bound while keeping the number of grid points computationally feasible.

[2] Note that these linear interpolation coefficients can be pre-computed, resulting in a 17×17 - matrix w^p, where $w^p_{j,\cdot}$ are the integration weights for evaluating next period's value function on $(p_1, ..., p_{17})$ given that in the current period $p = p_j$.

In addition to the these three state variables, the setup requires numerical integration over the two idiosyncratic i.i.d. shocks to earnings ν' and capital returns η' (as they affect next period's cash-on-hand x'). As both shocks are normally distributed, we use Gauss-Hermite quadrature once again.

1.2 Computing the ergodic distribution

The focus on tiny population groups such as the top 0.01% of the wealth distribution implies that solving for the ergodic distribution directly is more efficient than simulating a large number of agents and applying the ergodic theorem. In doing so, simulation error is eliminated; instead one can directly control the numerical error by updating the distribution until convergence.

Specifically, the EGM entails using a grid for assets $(a_i)_{i=1}^{100}$. Given p_j and β_m, saving a_i is optimal with cash-on-hand $x(a_i; p_j, \beta_m)$ that solves

$$\frac{\partial u(x(a_i; p_j, \beta_m) - a_i)}{\partial c} = \beta_m \mathbb{E}\left[\left(1 + \frac{\partial y'}{\partial a'}\left(1 - \frac{\partial \tau(y')}{\partial y}\right) + \frac{\partial \tilde{y}'}{\partial a'}(1-\tilde{\tau})\right)\frac{\partial V(x', p', \beta')}{\partial x} | p_j, \beta_m\right], \quad (1)$$

$$\text{where } x' = a_i + y' - \tau(y') + (1-\tilde{\tau})\tilde{y}' + T, \quad (2)$$

$$\text{and } \frac{\partial y'}{\partial a'} = \left(\underline{r} + r^X(a_i) + \frac{\partial r^X(a_i)}{\partial a} a_i\right), \quad (3)$$

$$\text{and } \frac{\partial \tilde{y}'}{\partial a'} = \left(\sigma^X(a_i)\eta' + \frac{\partial \sigma^X(a_i)}{\partial a}\eta' a_i\right). \quad (4)$$

While the main advantage of the EGM is efficiency ($x(a_i; p_j, \beta_m)$ can be found without maximizing the right-hand side of the Bellman equation), it is also convenient that the savings function is already inverted. First, for all $p_j, \beta_m, \nu_q, \eta_h$ and for all a_i, $i = 1, ..., 100$, there exists a unique level of asset holdings $a = s^{-1}(a_i; p_j, \beta_m, \nu_q, \eta_h)$ such that saving a_i is optimal.[3] Second, we define a finer grid for asset holdings $(k_i)_{i=1}^{1000}$ and interpolate (using a cubic spline) to find the inverse savings function $s^{-1}(k_i; p_j, \beta_m, \nu_q, \eta_h)$. Note that the borrowing constraint is binding for all $k \leq s^{-1}(k_1; p_j, \beta_m, \nu_q, \eta_h)$. Finally, we can solve for the ergodic distribution $G(k_i; p_j, \beta_m) \equiv Prob(k \leq k_i | p = p_j, \beta = \beta_m)$ at the grid points $(k_i)_{i=1}^{1000}$, $(p_j)_{j=1}^{17}$ and $(\beta_m)_{m=1}^{15}$. To simplify notation, we will denote by $G_{j,m}(k_i)$ this conditional cdf evaluated at grid points (p_j, β_m). This distribution has to satisfy

$$G_{j,m}(k_i) = \int_p \int_\beta \int_\nu \int_\eta G(s^{-1}(k_i; p, \beta, \nu, \eta); p, \beta) d\Gamma_\eta(\eta) d\Gamma_\nu(\nu) d\Gamma_\beta(\beta|\beta_m) d\Gamma_p(p|p_j). \quad (5)$$

Note that p_j and β_m are the realizations of the shock in period $t+1$ and the integration is over the shock values in period t. Nevertheless, e.g., $\Gamma_\beta(\beta|\beta_m)$ is the correct distribution as for any stationary Gaussian AR(1) process z_t the conditional random variables $z_t|z_{t+1}$ and $z_{t+1}|z_t$ have the same distribution.[4] Starting from some initial distribution $G^0_{j,m}(k_i)$ and using the short-hand notation $s^{-1}_{j,m,q,h}(k_i) =$

[3] $s^{-1}(a_i; p_j, \beta_m, \nu_q, \eta_h)$ is defined as the unique a that solves

$$x(a_i; p_j, \beta_m) = a + y - \tau(y) + (1-\tilde{\tau})\tilde{y} + T,$$

where $y = (\underline{r} + r^X(a))a + wl(p_j, \nu_q)$ and $\tilde{y} = \sigma^X(a)\eta_h a$.

[4] That is, the densities satisfy $f_{z_t|z_{t+1}}(x|y) = f_{z_{t+1}|z_t}(x|y)$.

$s^{-1}(k_i; p_j, \beta_m, \nu_q, \eta_h)$, we update until convergence according to

$$G^1_{j',m'}(k_i) = \sum_j w^p_{j',j} \sum_m w^\beta_{m',m} \sum_q w^\nu_q \sum_h w^\eta_h \hat{G}^0_{j,m}(s^{-1}_{j,m,q,h}(k_i)). \tag{6}$$

In (6), w^ν_q and w^η_h are the Gauss-Hermite quadrature weights for the transitory shocks ν and η (normalized to sum to one). The construction of the integration weights for the persistent shocks p and β is based on linear interpolation in $\psi(p)$- and β-space, respectively (see details below). $\hat{G}^0_{j,m}(\cdot)$ linearly interpolates $G^0_{j,m}(k_i)$ off the grid in the k-dimension.

Integration weights $w^p_{j,'j}$ and $w^\beta_{m',m}$. Consider the persistent earnings shock p. Conditional on its value in the next period being $p' = p_{j'}$ for some fixed $j' \in \{1, ..., 17\}$, the integration over the current period value p is with respect to the distribution of p, conditional on p', where $p|p' \sim N(\rho^P p' + (1-\rho^P)\mu^P, \sigma^P)$. Gauss-Hermite quadrature, here with ten sample points, entails evaluating the function of interest $G(s^{-1}(k_i; p, \beta, \nu, \eta); p, \beta)$ at $(\tilde{p}_n)_{n=1}^{10}$, where $\tilde{p}_n = \rho^P p' + (1-\rho^P)\mu^P + \sqrt{2}\sigma^P \tilde{x}_n$ and $(\tilde{x}_n)_{n=1}^{10}$ are the roots of the Hermite polynomial, and approximating the integral using the associated weights $(\tilde{w}_n)_{n=1}^{10}$ as

$$\approx \frac{1}{\sqrt{\pi}} \sum_{n=1}^{10} \tilde{w}_n G(s^{-1}(k_i; \tilde{p}_n, \beta, \nu, \eta); \tilde{p}_n, \beta).$$

Of course, $\tilde{p}_n$ will in general not lie on the p_j-grid, where the function value is known, requiring interpolation. Using linear interpolation, we can pre-compute the integration weights $(w^p_{j',j})_{j=1}^{17}$ we put on evaluating the function of interest at $(G(s^{-1}(k_i; p_j, \beta, \nu, \eta); p_j, \beta))_{j=1}^{17}$ in an efficient manner: for $n = 1, ..., 10$, locate $j(n)$ such that $p_{j(n)} \leq \tilde{p}_n \leq p_{j(n)+1}$ and compute the linear interpolation coefficient in $\psi(p)$-space λ_n as

$$\lambda_n = \frac{\psi(\tilde{p}_n) - \psi(p_{j(n)})}{\psi(p_{j(n)+1}) - \psi(p_{j(n)})}.$$

Then, looping over $n = 1, ..., 10$, add $(1-\lambda_n)\frac{1}{\sqrt{\pi}}\tilde{w}_n$ to $w^p_{j',j(n)}$ and $\lambda_n \frac{1}{\sqrt{\pi}}\tilde{w}_n$ to $w^p_{j',j(n)+1}$. The construction of the integration weights for β is analogous, except that linear interpolation can be performed directly in β-space.

Computing moments of the distribution. For example, aggregate wealth is given by

$$K = \int_p \int_\beta \left(\int_k k dG(k|p, \beta) \right) f_p(p) f_\beta(\beta) dp d\beta, \tag{7}$$

where $f_p(\cdot)$ and $f_\beta(\cdot)$ are the unconditional (i.e., cross-sectional) normal densities of the persistent shocks p and β. We integrate numerically according to

$$\hat{K} = \sum_{j=1}^{17} \bar{w}^p_j \sum_{m=1}^{15} \bar{w}^\beta_m \left(k_1 G_{j,m}(k_1) + \sum_{i=2}^{1000} \frac{k_{i-1} + k_i}{2} \left(G_{j,m}(k_i) - G_{j,m}(k_{i-1}) \right) \right). \tag{8}$$

As the discount factor grid $(\beta_m)_{m=1}^{15}$ was chosen as the Gauss-Hermite sample points, we set $(\bar{w}^\beta_m)_{m=1}^{15}$ to be the associated Gauss-Hermite quadrature weights. Recall that the Pareto tail transformation of the persistent earnings component p prompted us to define a grid $(p_j)_{j=1}^{17}$ with a particular emphasis on the right tail. Hence, we (pre-)compute the integration weights $(\bar{w}_j^p)_{j=1}^{17}$ manually: (i) define a very fine equally spaced grid $(\hat{p}_n)_{n=1}^{N}$ (if, say, $N = 100,000$, this has to be carried out only once) that covers the coarser grid $(p_j)_{j=1}^{17}$; (ii) for all $n = 1, ..., N$, locate $j(n)$ and compute λ_n as above; (iii) looping over $n = 1, ..., N$, add $(1 - \lambda_n) f_p(\hat{p}_n)$ to $\bar{w}_{j(n)}^p$ and $\lambda_n f_p(\hat{p}_n)$ to $\bar{w}_{j(n)+1}^p$ ($f_p(\cdot)$ is the pdf of $p \sim N(\mu^P, \frac{\sigma^P)}{1-\rho^P})$; and (iv) finally, normalize such that $\sum_{j=1}^{17} \bar{w}_j^p = 1$.[5]

1.3 Transition experiments

The perfect-foresight transition experiment is computationally straightforward. Given the calibrated initial steady state $(K^\star, \underline{r}^\star, T^\star)$, we solve for the new steady state $(K^{\star\star}, \underline{r}^{\star\star}, T^{\star\star})$ given the new exogenous environment. Next, we search for a fixed point in $(K_t, \underline{r}_t, T_t)_{t=t_0+1}^{t_1}$-space where $t_1 - t_0$ is chosen to be large enough such that $(K_{t_1}, \underline{r}_{t_1}, T_{t_1}) \approx (K^{\star\star}, \underline{r}^{\star\star}, T^{\star\star})$. For each iteration, we first solve for the value functions and corresponding (inverse) savings decisions backwards and subsequently roll the distribution forward, as described in the previous sections for the steady state. Note that now the grids and integration weights for the earnings process components are time-varying.[6]

The myopic transition experiment is conceptually very different. Given a period t distribution $G_{j,m}^t(k_i)$ and savings decisions $s_{j,m,q,h}^t(k)$ (reflecting factor prices $\underline{r}_t$, w_t, transfers T_t and exogenous environment θ_t, all naively assumed to persist forever), $G_{j,m}^{t+1}(k_i)$ is obtained as in (6).[7] In turn, $G_{j,m}^{t+1}(k_i)$ and θ_{t+1} determine K_{t+1} (thus w_{t+1}), $\underline{r}_{t+1}$, and T_{t+1}. The surprised agents expect this new endogenous and exogenous environment to prevail forever; hence, we solve the dynamic programming problem given this environment and accordingly obtain $s_{j,m,q,h}^{t+1}(k)$. Note that no fixed point problem has to be solved, and the capital stock converges to the same new steady state as under perfect foresight. Theoretically, this strategy could give rise to oscillatory paths of capital. However, this turns out not to be the case in our application.

References

Aiyagari, S. R. (1994). Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics*, *109*(3), pp. 659–684.

Carroll, C. D. (2006). The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems. *Economics Letters*, *91*(3), 312–320.

[5] Of course one could also use Gauss-Hermite quadrature here, as the corresponding weights and results coincide for all practical purposes.

[6] In particular, as the variance of the innovation term of the persistent earnings component σ_t^P is time-varying, $p_{t|t+1}$ is no longer equal to $p_{t+1|t}$ in distribution (but still normal); hence the integration weights for the decision problem (forward-looking) and the cross-sectional distribution (backward-looking) differ.

[7] Again, the grids and integration weights for the earnings process components are time-varying.