Online Appendix – Not for Publication

A Proofs

A.1 Proof of Proposition 1

Part 1: From the sequence problem, one can show that c_1 and c_2 are non-decreasing in a, $c_1'(a), c_2'(a) \geq 0$ with $c_1(a), c_2(a) \to \infty$ as $a \to \infty$ and $c_1'(a), c_2'(a) < \infty$ for $a > \underline{a}$. Further $c_2(a) \geq c_1(a)$ and $s_2(a) \geq s_1(a)$ for all a. A sketch of the proof that $c_j'(a) \geq 0$ is as follows: the value function $v_j(a), j = 1, 2$ is the value of the sequence problem of maximizing (1) subject to (2), (3) when initial wealth $a_0 = a$ and initial income $y_0 = y_j$. This is a maximization problem with a concave objective function and convex constraint set and it therefore has a weakly concave value function (this follows from an appropriate version of the maximum theorem). Optimal consumption c_j solves the first-order condition $u'(c_j(a)) = v_j'(a)$ for all a. Since v_j is weakly concave, v_j' is weakly decreasing and therefore c_j is weakly increasing.

To prove that $s_1(\underline{a}) = 0$ but $s_1(a) < 0$ all $a > \underline{a}$, consider the "Euler equation" (18) for type $j = 1$. Rearranging

$$\frac{u''(c_1(a))}{u'(c_1(a))} c_1'(a) s_1(a) = \rho - r - \lambda_1 \left(\frac{u'(c_2(a))}{u'(c_1(a))} - 1 \right) \tag{50}$$

We have $c_2(a) \geq c_1(a)$ and hence $u'(c_2(a)) \leq u'(c_1(a))$ and hence the right-hand side of (50) is strictly positive. Since $u'' < 0, u' > 0$ and $c_1' \geq 0$, $s_1(a) \leq 0$ for all a. First consider $a > \underline{a}$: since $c_1'(a) < \infty$, we need $s_1(a) < 0$ for $a > \underline{a}$. Next consider $a = \underline{a}$. Since wealth a needs to obey the state constraint (3), $s_1(a) \leq 0$ for all a implies that saving must be zero at the constraint: $s_1(\underline{a}) = 0$.[47]

Part 2: Consider the "Euler equation" (18) for the low income type $j = 1$. Using $s_1'(a) = r - c_1'(a)$, and rearranging gives

$$(s_1'(a) - r) s_1(a) = \frac{(r - \rho) u'(c_1(a)) + \lambda_1 (u'(c_2(a)) - u'(c_1(a)))}{u''(c_1(a))} \tag{51}$$

As $a \to \underline{a}$, we have that $s_1(a) \to 0, c_1(a) \to \underline{c}_1 := y_1 + r\underline{a} > 0, c_2(a) \to \underline{c}_2 > 0$ and, by Assumption 1, $-u'(c_1(a))/u''(c_1(a)) \to 1/\underline{R} > 0$. Therefore

$$s_1(a) s_1'(a) \to \nu_1 \quad \text{with} \quad \nu_1 := \frac{(r - \rho) u'(\underline{c}_1) + \lambda_1 (u'(\underline{c}_2) - u'(\underline{c}_1))}{u''(\underline{c}_1)} > 0 \tag{52}$$

[47] The second part of the Proposition below shows that $c_1'(a) \to \infty$ as $a \to \underline{a}$ so that there is no contradiction with (50) holding.

as defined in (21). We have

$$\lim_{a\to\underline{a}} \frac{(s_1(a))^2}{a-\underline{a}} = \lim_{a\to\underline{a}} 2s_1(a)s_1'(a) = 2\nu_1$$

where the first equality follows from l'Hôpital's rule and the second equality uses (52). Hence

$$(s_1(a))^2 \sim 2\nu_1(a-\underline{a}).$$

Taking the square root yields (19). The approximation to ν_1 in the second line of (21) uses the Taylor series approximation $u'(\underline{c}_2) \approx u'(\underline{c}_1) + u''(\underline{c}_1)(\underline{c}_2 - \underline{c}_1)$ to substitute out $u'(\underline{c}_1) - u'(\underline{c}_2) \approx -u''(\underline{c}_1)(\underline{c}_2 - \underline{c}_1)$ in the first line.$\square$

Proposition 1' (MPCs and Saving at Borrowing Constraint) *Assume that $r < \rho, y_1 < y_2$ and that Assumption 1 is violated, i.e. $\underline{R} = \infty$. Then the solution to the HJB equation (7) and the corresponding saving policy function (9) have the following properties:*

1. *$s_1(\underline{a}) = 0$ but $s_1(a) < 0$ all $a > \underline{a}$. That is only individuals exactly at the borrowing constraint are constrained, whereas those with wealth $a > \underline{a}$ are unconstrained and decumulate assets.*

2. *as $a \to \underline{a}$, the saving and consumption policy function of the low income type and the corresponding instantaneous marginal propensity to consume satisfy*

$$s_1(a) \sim -\eta_1(a-\underline{a}), \tag{53}$$
$$c_1(a) \sim y_1 + ra + \eta_1(a-\underline{a}), \tag{54}$$
$$c_1'(a) \sim r + \eta_1, \tag{55}$$
$$\eta_1 := \frac{\rho - r + \lambda_1(1-\xi)}{\underline{\gamma}}, \quad \underline{\gamma} := -\lim_{a\to\underline{a}} \frac{u''(c_1(a))c_1(a)}{u'(c_1(a))}, \quad \xi := \lim_{a\to\underline{a}} \frac{u'(c_2(a))}{u'(c_1(a))} \tag{56}$$

   where $\underline{c}_j = c_j(\underline{a}), j = 1,2$ is consumption of the two types at the borrowing constraint and where ξ is zero if u satisfies the Inada condition $u'(c) \to \infty$ as $c \to 0$. This implies that the derivatives of c_1 and s_1 are bounded at the borrowing constraint, $c_1'(\underline{a}) < \infty$ and $|s_1'(\underline{a})| < \infty$.

3. *With CRRA utility (5) we have $\underline{\gamma} = \gamma$ and $\xi = 0$ so that (56) is $\eta_1 = (\rho - r + \lambda_1)/\gamma$.*

Proof: The proof of the first part is the same as that of Proposition 1. For the second part, recall from the discussion in the main text that Assumption 1 not being satisfied means that *both* (i) the borrowing constraint equals the natural borrowing constraint $\underline{a} = -y_1/r$ so that $c_1(\underline{a}) = 0$ *and* (ii) absolute risk aversion $R(c) := -u''(c)/u'(c) \to \infty$ as $c \to 0$. Next, note that

(51) in the proof of Proposition 1 still holds. However, we now have $-u'(c_1(a))/u''(c_1(a)) \to 0$ as $a \to \underline{a}$, and similarly $-u'(c_2(a))/u''(c_1(a)) = -\frac{u'(c_2(a))}{u'(c_1(a))}\frac{u'(c_1(a))}{u''(c_1(a))} \to 0$ as $a \to \underline{a}$. Therefore when Assumption 1 does not hold $s_1'(a)s_1(a) \to 0$ as $a \to \underline{a}$. We therefore pursue a slightly different strategy. Rearranging (18) for type $j = 1$:

$$(r - \rho - \lambda_1)c_1(a) = \gamma(c_1(a))c_1'(a)s_1(a) - \lambda_1 c_1(a)\frac{u'(c_2(a))}{u'(c_1(a))}, \quad \gamma(c) := -\frac{u''(c)c}{u'(c)}$$

Differentiate with respect to a

$$(r - \rho - \lambda_1)c_1' = \frac{d}{da}\left[\gamma(c_1)c_1'\right]s_1 + \gamma(c_1)c_1'(r - c_1') - \lambda_1 c_1'\frac{u'(c_2)}{u'(c_1)} - \lambda_1 c_1 \frac{d}{da}\left(\frac{u'(c_2)}{u'(c_1)}\right)$$

Evaluating at $\underline{a}$ so that $s_1(\underline{a}) = c_1(\underline{a}) = 0$ we have

$$r - \rho - \lambda_1 + \lambda_1\xi = \underline{\gamma}(r - c_1'(\underline{a}))$$

where $\underline{\gamma}$ and ξ are defined in (56). Finally, defining $\eta_1 := -s_1'(\underline{a}) = -(r - c_1'(\underline{a}))$ we have (56).□

A.2 Proof of Proposition 2

Part 1, existence of $a_{\max}$: We have already proven in Proposition 1 that the low income type always decumulates $s_1(a) < 0$ for $a > \underline{a}$. Hence to prove that there is an $a_{\max}$ such that $s_1(a), s_2(a) < 0$ for $a > a_{\max}$ we need to only consider the high income type. Rearranging (18) for $j = 2$

$$\frac{u''(c_2(a))}{u'(c_2(a))}c_2'(a)s_2(a) = \rho - r - \lambda_2\left(\frac{u'(c_1(a))}{u'(c_2(a))} - 1\right) \tag{57}$$

In contrast to the expression for type $j = 1$, (50), the sign of the right-hand side of (57) is ambiguous (in particular it may be negative). The economic intuition is that the term $\lambda_2\left(\frac{u'(c_1(a))}{u'(c_2(a))} - 1\right) \geq 0$ captures the precautionary saving motif. The proof strategy is to argue that this term becomes small for large a and that therefore the right-hand side of (57) is positive for large a.

To this end, recall the assumption that relative risk aversion is bounded above, $\gamma(c) = -cu''(c)/u'(c) \leq \bar{\gamma}$ for all c. Using this, we have

$$\frac{u'(c_1(a))}{u'(c_2(a))} \leq \left(\frac{c_2(a)}{c_1(a)}\right)^{\bar{\gamma}}. \tag{58}$$

Further

$$c_2(a) - c_1(a) = y_2 - y_1 - (s_2(a) - s_1(a)) = (y_2 - y_1)(1 - \theta(a)),$$

where $\theta(a) = (s_2(a) - s_1(a))/(y_2 - y_1) \geq 0$. Also note that $c_2(a) \geq c_1(a)$ implies that $\theta(a) \leq 1$. Hence

$$\frac{u'(c_1(a))}{u'(c_2(a))} \leq \left(1 + \frac{(y_2 - y_1)(1 - \theta(a))}{c_1(a)}\right)^{\bar{\gamma}}.$$

Since $c_1 \to \infty$ as $a \to \infty$, we have

$$\lim_{a \to \infty} \frac{u'(c_1(a))}{u'(c_2(a))} = 1.$$

Hence the right-hand side of (57) is strictly positive for a large enough. Since $u'' < 0, u' > 0, c_2' \geq 0$, we have $s_2(a) \leq 0$ for a large enough. Denoting the (largest) root of s_2 by $a_{\max}$, we obtain the first part of the Lemma.

Remark: note that the economically interesting case is the one in which the right-hand side of (57), $\rho - r - \lambda_2 \left(\frac{u'(c_1(a))}{u'(c_2(a))} - 1\right)$ is strictly positive for large a, strictly negative for small a (close to $\underline{a}$), and zero at $a_{\max}$. In such cases $a_{\max} > \underline{a}$, and $s_2(a) > 0$ for some $\underline{a} \leq a < a_{\max}$, i.e. some high-income types accumulate wealth. If instead, the right-hand side of (57) is strictly positive for all $a > \underline{a}$, then $a_{\max} = \underline{a}$ and $s_2(a) < 0$ for all $a > \underline{a}$, i.e. all high-income types decumulate wealth (just like the low income types).

Part 1, behavior of s_2 close to $a_{\max}$: Before laying out the proof we start with observation that some readers may find useful: to understand the behavior of s_2 at $a_{\max}$, one may be tempted to follow analogous steps to those in part 2 of Proposition 1. This is, however, not the right strategy given that the right-hand side of (57) equals zero by the definition of $a_{\max}$. We therefore pursue a strategy more akin to that in Proposition A.1.

Consider (18) for type $j = 2$:

$$(\rho - r + \lambda_2) u'(c_2(a)) = u''(c_2(a)) c_2'(a) s_2(a) + \lambda_2 u'(c_1(a)).$$

Differentiate with respect to a

$$(\rho - r + \lambda_2) u''(c_2) c_2' = \frac{d}{da}[u''(c_2) c_2'] s_2 + u''(c_2) c_2'(r - c_2') + \lambda_2 u''(c_1) c_1'.$$

Evaluating at $a_{\max}$ so that $s_2(a_{\max}) = 0$

$$(\rho - r + \lambda_2) c_2'(a_{\max}) = c_2'(a_{\max})(r - c_2'(a_{\max})) + \lambda_2 \frac{u''(c_1(a_{\max}))}{u''(c_2(a_{\max}))} c_1'(a_{\max}).$$

Define

$$\xi := c_2'(a_{\max}), \qquad \chi := \lambda_2 \frac{u''(c_1(a_{\max}))}{u''(c_2(a_{\max}))} c_1'(a_{\max}) > 0.$$

Using these definitions and rearranging

$$\xi^2 + (\rho - 2r + \lambda_2)\xi - \chi = 0.$$

Since $\chi > 0$, this quadratic has two real roots, one positive and one negative. Therefore ξ is the positive root and given by

$$c_2'(a_{\max}) = \xi = \frac{-(\rho - 2r + \lambda_2) + \sqrt{(\rho - 2r + \lambda_2)^2 + 4\chi}}{2}.$$

Also note that $c_2'(a_{\max}) = \xi < \infty$. Finally we have

$$\zeta_2 := -s_2'(a_{\max}) = c_2'(a_{\max}) - r = \frac{-(\rho + \lambda_2) + \sqrt{(\rho - 2r + \lambda_2)^2 + 4\chi}}{2}.$$

Hence $s_2(a) \sim \zeta_2(a_{\max} - a)$ as $a \to a_{\max}$.□

Remark: note that the behavior of s_2 at $a_{\max}$ is symmetric to that of s_1 near $\underline{a}$ *in the case in which Assumption 1 is violated* (see Proposition A.1). Suppose instead that there was a state constraint $a \leq \bar{a}$ with $\bar{a}$ tight (i.e. low) enough. Then, the behavior s_2 would instead satisfy $s_2'(a) \to -\infty$ as $a \to \bar{a}$, i.e. the behavior of s_2 would be symmetric to that of s_1 under Assumption 1 (see Proposition 1).

Part 2 of Proposition 2: Asymptotic Behavior with CRRA Utility Before proceeding to the proof of the result, we derive two auxiliary Lemmas. The first Lemma considers an auxiliary problem without labor income, $y_1 = y_2 = 0$, and shows that optimal policy functions are linear in wealth. The second Lemma shows that the problem with labor income and a borrowing constraint (7) satisfies a certain homogeneity property.

Lemma 3 *Consider the problem*

$$\rho v(a) = \max_c \; u(c) + v'(a)(ra - c) \tag{59}$$

where the utility function is given by (5). The optimal policy functions that solve (59) are linear in wealth and given by

$$c(a) = \frac{\rho - (1-\gamma)r}{\gamma} a, \qquad s(a) = \frac{r - \rho}{\gamma} a. \tag{60}$$

Proof of Lemma 3: Use a guess-and-verify strategy. Guess $v(a) = B\frac{a^{1-\gamma}}{1-\gamma}$ which implies

$$v'(a) = Ba^{-\gamma} \tag{61}$$

$$c(a) = v'(a)^{-1/\gamma} = B^{-1/\gamma} a \tag{62}$$

Substituting into (59) and dividing by $a^{1-\gamma}$

$$\rho B \frac{1}{1-\gamma} = \frac{1}{1-\gamma} B^{-(1-\gamma)/\gamma} + Br - BB^{-1/\gamma}$$

Dividing by B and collecting terms we have $B^{-1/\gamma} = \frac{\rho - r}{\gamma} + r$ and hence from (62) we have (60).□

Lemma 4 *Consider problem* (7)*. For any* $\xi > 0$*,*

$$v_j(\xi a) = \xi^{1-\gamma} v_{\xi,j}(a) \tag{63}$$

where $v_{\xi,j}$ *solves*

$$\rho v_{\xi,j}(a) = \max_c \ u(c) + v'_{\xi,j}(a)(y_j/\xi + ra - c) + \lambda_j(v_{\xi,-j}(a) - v_{\xi,j}(a)) \tag{64}$$

Proof of Lemma 4: Write (7) as

$$\rho v_j(a) = H(v'_j(a)) + v'_j(a)(y_j + ra) + \lambda_j(v_{-j}(a) - v_j(a)) \tag{65}$$
$$H(p) = \max_c \{u(c) - pc\} = \frac{\gamma}{1-\gamma} p^{\frac{\gamma-1}{\gamma}}$$

From (63), $v_j(a) = \xi^{1-\gamma} v_{\xi,j}(a/\xi)$, $v'_j(a) = \xi^{-\gamma} v'_{\xi,j}(a/\xi)$. Therefore $H(v'_j(a)) = H(v'_{\xi,j}(a/\xi))\xi^{1-\gamma}$. Substituting into (65) and dividing by $\xi^{1-\gamma}$ yields (64).□

Conclusion of Proof of Part 2 of Proposition 2: With these two Lemmas in hand we are ready to prove Part 2 of Proposition 2. Consider first the asymptotic behavior of the consumption policy function $c_j(a)$. From (63), $v_j(a) = \xi^{1-\gamma} v_{\xi,j}(a/\xi)$, $v'_j(a) = \xi^{-\gamma} v'_{\xi,j}(a/\xi)$ and therefore

$$c_j(a) = (v'_j(a))^{-1/\gamma} = \xi(v'_{\xi,j}(a/\xi))^{-1/\gamma} = \xi c_{\xi,j}(a/\xi)$$

In particular with $\xi = a$ we have

$$c_j(a) = a c_{a,j}(1)$$

Hence

$$\lim_{a\to\infty} \frac{c_j(a)}{a} = \lim_{\xi\to\infty} c_{\xi,j}(1) = c(1) = \frac{\rho - (1-\gamma)r}{\gamma},$$

where the second equality uses that problem (64) converges to that with no labor income (59) as $\xi \to \infty$ and therefore also $c_{\xi,j}(a) \to c(a)$ for all a as $\xi \to \infty$. The asymptotic behavior of $s_j(a)$ can be proved in an analogous fashion.□

A.3 Proof of Corollary 2

The marginal propensity to save (MPS) is simply the derivative of (29) with respect to starting wealth a:

$$\text{MPS}_{1,\tau}(a) \sim \left(1 - \tau\sqrt{\frac{\nu_1}{2(a-\underline{a})}}\right)^+. \tag{66}$$

To find the marginal propensity to consume (MPC) in (30) proceed as follows. Integrating the budget constraint $\dot{a}(t) + c(t) = y + ra(t)$ between $t = 0$ and $t = \tau$ and using $a_0 = a$ as well as the definitions of $S_\tau(a)$ and $C_\tau(a)$, we have[48]

$$S_\tau(a) + C_\tau(a) = a + \int_0^\tau (y + ra(t))dt \approx a + \tau(y + ra).$$

Differentiating with respect to starting wealth a, we have $\text{MPS}_\tau(a) + \text{MPC}_\tau(a) = 1 + \tau r$. Using (66) we obtain (30).□

A.4 Proof of Proposition 3

Integrating (31), we have

$$\log g_j(a) = \kappa_j - \log s_j(a) - \int_{\underline{a}}^a \left(\frac{\lambda_j}{s_j(x)} + \frac{\lambda_{-j}}{s_{-j}(x)}\right) dx, \quad j = 1, 2$$

or equivalently (33). Since $s_1(a)g_1(a) + s_2(a)g_2(a) = 0$ for all a as discussed in the main text, we need $\kappa_1 + \kappa_2 = 0$. The level of κ_1 and κ_2 is as explained in Appendix A.4.4 below.

A.4.1 Part 1: Close to the borrowing constraint

Now consider the behavior of g_1 near the borrowing constraint $a = \underline{a}$. The argument for Part 2, i.e. the behavior of g_2 near $a = a_{\max}$, is exactly symmetric and will be presented afterwards. The proof that g_1 features a Dirac point mass at $a = \underline{a}$ has already been stated in the text, right after the Proposition.

Consider our analytic expression for g_1 in (33), and its behavior near $a = \underline{a}$. The key is to understand

$$\lim_{a\to\underline{a}} \frac{-1}{s_1(a)} \exp\left(-\int_{a_0}^a \frac{\lambda_1}{s_1(x)}dx\right).$$

We will show that this limit equals either 0 or ∞ and since s_2 is bounded as $a \to \underline{a}$, the behavior of g_1 in (33) will be identical to the behavior of this limit. Assume that the *leading*

[48] We can also proceed without the approximation $\int_0^\tau ra(t)dt \approx ra_0$ and compute the term exactly using our closed-form solution in (22). But this adds only minor corrective terms.

term of s_1 around $\underline{a}$ is $-\vartheta(a-\underline{a})^\alpha$ for constants $\vartheta > 0, \alpha > 0$. Denote

$$L(\lambda_1, \vartheta, \alpha) := \lim_{a\to\underline{a}} \frac{1}{\vartheta(a-\underline{a})^\alpha} \exp\left(\int_{a_0}^{a} \frac{\lambda_1}{\vartheta(x-\underline{a})^\alpha} dx\right).$$

Then there are three different cases for the value of $L(\lambda_1, \vartheta, \alpha)$.

1. $0 < \alpha < 1$.

$$L(\lambda_1, \vartheta, \alpha) = \lim_{a\to\underline{a}} \frac{1}{\vartheta(a-\underline{a})^\alpha} \exp\left(\frac{\lambda_1}{\vartheta}\frac{1}{1-\alpha}\left((a-\underline{a})^{1-\alpha} - (a_0-\underline{a})^{1-\alpha}\right)\right) = +\infty$$

2. $\alpha > 1$

$$L(\lambda_1, \vartheta, \alpha) = \lim_{a\to\underline{a}} \frac{1}{\vartheta(a-\underline{a})^\alpha} \exp\left(\frac{\lambda_1}{\vartheta}\frac{1}{1-\alpha}\left((a-\underline{a})^{1-\alpha} - (a_0-\underline{a})^{1-\alpha}\right)\right) = 0$$

3. $\alpha = 1$.

$$\begin{aligned} L(\lambda_1, \vartheta, \alpha) &= \lim_{a\to\underline{a}} \frac{1}{\vartheta(a-\underline{a})} \exp\left(\frac{\lambda_1}{\vartheta}\left(\log(a-\underline{a}) - \log(a_0-\underline{a})\right)\right) \\ &= \lim_{a\to\underline{a}} \frac{(a-\underline{a})^{\lambda_1/\vartheta-1}}{\vartheta(a_0-\underline{a})^{\lambda_1/\vartheta}} \end{aligned}$$

   (a) If $\lambda_1 > \vartheta$, then $L(\lambda_1, \vartheta, 1) = 0$.

   (b) If $\lambda_1 = \vartheta$, then $L(\lambda_1, \vartheta, 1) \propto 1/\vartheta$.

   (c) If $\lambda_1 < \vartheta$, then $L(\lambda_1, \vartheta, 1) = +\infty$.

Now we come back to our problem of understanding the behavior of g_1 at $\underline{a}$. There are two cases.

(i) If Assumption 1 holds, we know from Proposition 1 that the leading term of s_1 at $\underline{a}$ is $-(2\nu_1(a-\underline{a}))^{1/2}$. Therefore, we are in the case $\alpha < 1$ and we have $g_1(a) \to +\infty$ as $a \to \underline{a}$.

(ii) If Assumption 1 does not hold, we know from Proposition 1' in Appendix A.1 that the leading term of s_1 at $\underline{a}$ is $-\eta_1(a-\underline{a})$. Therefore we are in the case $\alpha = 1$ and $g_1(\underline{a}) = 0$ if $\lambda_1 > \eta_1$ and $g_1(a) \to \infty$ as $a \to \underline{a}$ if $\lambda_1 < \eta_1$.

A.4.2 Part 2: In the right tail

Next, consider the behavior of g_2 at $a_{\max}$. The argument is exactly symmetric to Part 1 and we need to understand

$$\lim_{a\to a_{\max}} \frac{-1}{s_2(a)} \exp\left(-\int_{a_0}^{a} \frac{\lambda_2}{s_2(x)} dx\right)$$

Analogous to before denote the leading term of s_2 by $\vartheta(a_{\max} - a)^\alpha$ with $\vartheta > 0, \alpha > 0$ and

$$L(\lambda_2, \vartheta, \alpha) = \lim_{a \to a_{\max}} \frac{1}{\vartheta(a - a_{\max})^\alpha} \exp\left(\int_{a_0}^{a} \frac{\lambda_2}{\vartheta(x - a_{\max})^\alpha} dx\right)$$

There are again three cases depending on whether $\alpha \gtreqless 1$. From Proposition 2, we know that the leading term of s_2 is $\zeta_2(a_{\max} - a)$, i.e. we are in the case $\alpha = 1$. Therefore

$$L(\lambda_2, \vartheta, \alpha) = \lim_{a \to a_{\max}} \frac{(a - a_{\max})^{\lambda_2/\vartheta - 1}}{\vartheta(a_0 - a_{\max})^{\lambda_2/\vartheta}}$$

and further using $\vartheta = \zeta_2$, we have

$$g_2(a) \sim \xi(a_{\max} - a)^{\lambda_2/\zeta_2 - 1} \quad \text{as } a \to a_{\max}$$

for a constant ξ. Since $g_1(a_{\max}) = 0$ and $g(a) = g_1(a) + g_2(a)$ we obtain (72).

A.4.3 Part 3: Smoothness

That g_1 and g_2 are continuous and differentiable for all $a > \underline{a}$ follows directly from the analytic solution (33) and the fact that s_1, s_2 are continuous and differentiable.

A.4.4 Constants of Integration for Stationary Distribution (33)

In all cases of Proposition 3 we can express g_1, g_2 as functions of $(s_1, s_2, \lambda_1, \lambda_2)$ only by using the normalization condition (32), i.e. we can pin down the constants of integration (κ_1, κ_2) in (33). In the case with the Dirac mass (if Assumption 1 holds), this condition is

$$m_1 + \lim_{\varepsilon \to 0} \int_{\underline{a}+\varepsilon}^{a_{\max}} g_1(a) da = \frac{\lambda_2}{\lambda_1 + \lambda_2}, \qquad \int_{\underline{a}}^{a_{\max}} g_2(a) da = \frac{\lambda_1}{\lambda_1 + \lambda_2} \tag{67}$$

The following auxiliary Lemma is useful:

Lemma 5 *Under Assumption 1, we have the following relationship between the density at $a = \underline{a}$ and the Dirac mass m_1:*

$$0 = -\lim_{\varepsilon \to 0} s_1(\underline{a} + \varepsilon) g_1(\underline{a} + \varepsilon) - \lambda_1 m_1 \tag{68}$$

The Lemma states that the inflow of type 1 individuals into the borrowing constraint equals the outflow out of the constraint.

Proof: Integrating the stationary KF equation (8) between $\underline{a}+\varepsilon$ and $a_{\max}$ yields

$$0 = s_1(\underline{a}+\varepsilon)g_1(\underline{a}+\varepsilon) - \lambda_1 \int_{\underline{a}+\varepsilon}^{a_{\max}} g_1(a)da + \lambda_2 \int_{\underline{a}+\varepsilon}^{a_{\max}} g_2(a)da$$

Combining with (67), we have (68).□

Equation (68) can be used as a boundary condition for (33). From (33) for type $j=1$, we have $\lim_{\varepsilon\to 0} s_1(\underline{a}+\varepsilon)g_1(\underline{a}+\varepsilon) = \kappa_1$ and hence $\kappa_1 = -\lambda_1 m_1$. Since $\kappa_1+\kappa_2 = 0$, $\kappa_2 = \lambda_1 m_1$. Therefore

$$g_1(a) = -\frac{\lambda_1 m_1}{s_1(a)} \exp\left(-\int_{\underline{a}}^{a}\left(\frac{\lambda_1}{s_1(x)}+\frac{\lambda_2}{s_2(x)}\right)dx\right), \tag{69}$$

$$g_2(a) = +\frac{\lambda_1 m_1}{s_2(a)} \exp\left(-\int_{\underline{a}}^{a}\left(\frac{\lambda_1}{s_1(x)}+\frac{\lambda_2}{s_2(x)}\right)dx\right). \tag{70}$$

Substituting (70) into (67) we have

$$\lambda_1 m_1 \int_{\underline{a}}^{a_{\max}} \left\{\frac{1}{s_2(a)} \exp\left(-\int_{\underline{a}}^{a}\left(\frac{\lambda_1}{s_1(x)}+\frac{\lambda_2}{s_2(x)}\right)dx\right)\right\} da = \frac{\lambda_1}{\lambda_1+\lambda_2}$$

Rearranging, we have m_1 as a function of $(s_1, s_2, \lambda_1, \lambda_2)$ only:

$$m_1 = \frac{\lambda_2}{\lambda_1+\lambda_2}\tilde{m}_1, \quad \frac{1}{\tilde{m}_1} = \lambda_2 \int_{\underline{a}}^{a_{\max}} \left\{\frac{1}{s_2(a)} \exp\left(-\int_{\underline{a}}^{a}\left(\frac{\lambda_1}{s_1(x)}+\frac{\lambda_2}{s_2(x)}\right)dx\right)\right\} da. \tag{71}$$

Given m_1, we also know g_1 and g_2 as functions of $(s_1, s_2, \lambda_1, \lambda_2)$ only. In the case without the Dirac mass (if Assumption 1 does not hold) we have (67) with $m_1 = 0$ and these two equations pin down the constants of integration in (33).

A.5 Additional Characterizations of the Stationary Wealth Distribution that Follow from Proposition 3

The following corollary to Proposition 3 summarizes some additional properties of the stationary wealth distribution for the special case with two income types.

Corollary 3 *The stationary wealth distribution in the special case with two income types (33) has the following properties in addition to those already listed in Proposition 3:*

1. *(In the right tail) At its upper bound $a_{\max}$, the wealth distribution $g(a) := g_1(a) + g_2(a)$ satisfies*

$$g(a) \sim \xi(a_{\max} - a)^{\lambda_2/\zeta_2 - 1} \quad \text{as } a \to a_{\max} \tag{72}$$

where $\zeta_2 = |s'_2(a_{\max})|$ and ξ is a constant. Therefore $g(a_{\max}) = 0$ for large λ_2 (so that $\lambda_2 > \zeta_2$). In contrast, $g_2(a) \to \infty$ as $a \to a_{\max}$ for small λ_2. In neither case is there a Dirac mass.

2. *(Shape of the wealth distribution) The exact shape of g_1 and g_2 is ambiguous. However, both g_1 and g_2 are ratios of well-understood functions, in particular $g_j(a) = \kappa_j f(a)/s_j(a), j = 1, 2$ where $f(a) := \exp\left(-\int_{\underline{a}}^{a}\left(\frac{\lambda_1}{s_1(x)} + \frac{\lambda_2}{s_2(x)}\right)dx\right)$ and $\kappa_1 < 0, \kappa_2 > 0$. The function f is strictly log-concave and single-peaked with $f'(a)/f(a) \to \infty$ as $a \downarrow \underline{a}$ and $f'(a)/f(a) \to -\infty$ as $a \uparrow a_{\max}$.*

3. *(Joint distribution of labor income and wealth) For any given wealth level a, the fraction of individuals that have the high income y_2, $\Pr(y_2|a) := \frac{g_2(a)}{g_1(a)+g_2(a)}$ satisfies $\Pr(y_2|a) = \frac{1}{1-s_2(a)/s_1(a)}$ and similarly $\Pr(y_1|a) = 1 - \Pr(y_2|a) = \frac{1}{1-s_1(a)/s_2(a)}$.*

Proof of Corollary 3

1. (In the right tail) This was already proven when proving Part 2 in Proposition 3.

2. (Shape of the wealth distribution) Consider $f(a) := \exp\left(-\int_{\underline{a}}^{a}\left(\frac{\lambda_1}{s_1(x)} + \frac{\lambda_2}{s_2(x)}\right)dx\right)$. We have

$$\frac{d\log f(a)}{da} = -\left(\frac{\lambda_1}{s_1(a)} + \frac{\lambda_2}{s_2(a)}\right)$$

Further, since $s_1(a)$ and $s_2(a)$ are strictly decreasing, we have that

$$\frac{d^2\log f(a)}{da^2} = \frac{\lambda_1}{(s_1(a))^2}s'_1(a) + \frac{\lambda_2}{(s_2(a))^2}s'_2(a) < 0,$$

i.e. $d\log f(a)/da$ is strictly decreasing or, equivalently, $f(a)$ is strictly log-concave. Since $s_1(a) < 0$ for all $a \in (\underline{a}, a_{\max})$ and $s_1(\underline{a}) = 0$, we have $1/s_1(a) \to -\infty$ as $a \downarrow \underline{a}$. Similarly $1/s_2(a) \to +\infty$ as $a \uparrow a_{\max}$. Therefore

$$\lim_{a\downarrow\underline{a}} \frac{d\log f(a)}{da} = \infty, \qquad \lim_{a\uparrow a_{\max}} \frac{d\log f(a)}{da} = -\infty.$$

Since $d\log f(a)/da$ is strictly decreasing, there is a critical point a^* such that $f'(a) > 0$ for $a < a^*$ and $f'(a) < 0$ for $a > a^*$. Summarizing f is single-peaked and strictly log-concave.

3. (Joint distribution of labor income and wealth) As discussed in the main text, $s_1(a)g_1(a) + s_2(a)g_2(a) = 0$ for all a. Therefore $g_1(a) = -g_2(a)s_2(a)/s_1(a)$ and so

$$\Pr(y_2|a) := \frac{g_2(a)}{g_1(a) + g_2(a)} = \frac{1}{1 - s_2(a)/s_1(a)}.\square$$

Discussion of Corollary 3. Part 1 of Corollary 3 provides a more complete characterization of the wealth distribution's tail in the vicinity of its upper bound. From (72) top wealth inequality is high (g declines towards zero at $a_{\max}$ only slowly) if individuals face a high likelihood of dropping out of the high income state (λ_2 is high) and if high-income types accumulate wealth only slowly (ζ_2 is low). Intuitively, wealth accumulation requires both time and luck (consecutive high income draws). And under the circumstances just mentioned, only a few individuals obtain sufficiently long enough high income spells to accumulate large riches. Hence, wealth inequality is high.

Part 2 characterizes the wealth distribution for intermediate wealth levels. It shows that the shapes of g_1 and g_2 in Figure 6 are not simply due to a particular numerical example. Instead both density functions are simple ratios of well-understood functions $g_j(a) = \kappa_j f(a)/s_j(a)$ where f is defined in the Proposition and hump-shaped. For instance consider g_1 as a increases: as in Figure 6, g_1 tends to be first decreasing, then increasing again and finally decreasing. Similarly, consider g_2 as a increases: it tends to be first increasing and then decreasing (hump-shaped), again as in the Figure.

Part 3 characterizes the joint distribution of income and wealth. The fraction of high income types conditional on wealth $\Pr(y_2|a)$ depends only on the saving rates s_1 and s_2 but, perhaps surprisingly and in contrast to the fraction of high income types in the population $\Pr(y_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$, it does not depend directly on the intensities λ_1 and λ_2.

A.6 Supplement to Proposition 3: Dirac or No Dirac?

We here briefly return to part 1 of Proposition 3 and illustrate in more detail when and, if so, why the wealth distribution features a Dirac mass at the borrowing constraint. To this end, consider again the two special cases without income risk from Section 3.2. In the first special case with exponential utility $s(a) = -\sqrt{2\nu a}$. in the second special case with CRRA utility $s(a) = -\eta a$. To obtain a stationary wealth distribution in the absence of income risk, assume that individuals die at rate λ. When an individual dies, she is replaced by a newborn with starting wealth $a_{\max}$.[49] Because $r < \rho$ so that everyone decumulates wealth, $a_{\max}$ is also the upper bound of the wealth distribution (hence the notation).

It turns out to be convenient to work with the cumulative distribution function $G(a)$ which satisfies[50]

$$0 = -s(a)G'(a) - \lambda G(a), \quad 0 < a < a_{\max} \tag{73}$$

with boundary condition $G(a_{\max}) = 1$. This equation can be solved easily: integrating

[49]Alternatively, we could assume that newborns draw their starting wealth from some distribution Ψ with support $[0, a_{\max}]$. In this case, (73) below is identical but with an additional term $+\lambda\Psi(a)$.

[50]The KF equation is $0 = -(s(a)g(a))' - \lambda g(a)$ for $0 < a < a_{\max}$. Integrating and using $G(a) = \int_0^a g(x)dx$ yields (73). Working with the CDF is only more convenient in this special case. In the case with two income types above, it is instead more convenient to work with the densities.

$G'(a)/G(a) = -\lambda/s(a)$ with $G(a_{\max}) = 1$ we have

$$G(a) = \exp\left(\int_a^{a_{\max}} \frac{\lambda}{s(x)}dx\right). \tag{74}$$

In the first special case with $s(a) = -\sqrt{2\nu a}$, the CDF in (74) becomes

$$G(a) = \exp\left(\lambda\sqrt{2a/\nu} - \lambda\sqrt{2a_{\max}/\nu}\right). \tag{75}$$

Note in particular that $m := G(0) = \exp\left(-\lambda\sqrt{2a_{\max}/\nu}\right) > 0$, i.e. there is a Dirac mass at the borrowing constraint $a = 0$. In contrast, in the second special case $s(a) = -\eta a$, (74) becomes

$$G(a) = \left(\frac{a}{a_{\max}}\right)^{\lambda/\eta}. \tag{76}$$

Therefore $G(0) = 0$ i.e. there are no individuals at the borrowing constraint. As can be seen clearly in their derivations, the difference between (75) and (76) is solely due to the saving behavior (linearity versus unbounded derivative at $a = 0$) which determines whether individuals hit the borrowing constraint in finite time (see Section 3.2).

The special case with exponential utility and no income risk also yields some instructive comparative statics that carry over to numerical solutions of the more general case. Death risk at rate λ results in a higher effective discount rate $\rho + \lambda$ and hence the natural formula for the parameter governing the speed at which individuals hit the constraint is $\nu = (\rho - r + \lambda)/\theta$. Using this, the number of individuals at the borrowing constraint is $m = G(0) = \exp\left(-\lambda\sqrt{\frac{2\theta a_{\max}}{\rho - r + \lambda}}\right)$. This quantity is decreasing in the coefficient of absolute risk aversion θ, increasing in the gap $\rho - r$ and decreasing in the Poisson rate λ.[51] Numerical experiments in the model with a two-state Poisson process for income show that the same comparative static holds with respect to λ_1, the Poisson rate of leaving the low income state.

A.7 Proof of Proposition 4

As mentioned in Section 3.6, our uniqueness result not only applies to the two-state income process analyzed in Sections 1 and 3 but also to much more general stationary Markovian income processes, e.g. the diffusion process of Appendix G.1. To treat the general case, we

[51] An increase in λ has two offsetting effects: on one hand, individuals approach the borrowing constraint faster (ν), thereby increasing m; on the other hand, individuals are more likely to die before they reach the constraint, thereby decreasing m. Differentiation of the expression for m shows that the latter effect always dominates.

write the HJB and KF equations as

$$\rho v = \max_c \; u(c) + (y + ra - c)\partial_a v + \mathcal{A}v, \tag{77}$$

$$0 = -\partial_a(s(a, y)g) + \mathcal{A}^* g, \tag{78}$$

with a state constraint $a \geq \underline{a}$. Here $\mathcal{A}$ is the infinitesimal generator ("infinite-dimensional transition matrix") of the stochastic process for income y_t and $\mathcal{A}^*$ is its adjoint. For instance, if y_t follows a two-state Poisson process as in Section 1, then $(\mathcal{A}v)(a, y_j) = \lambda_j(v(a, y_{-j}) - v(a, y_j))$. Or if y_t is a continuous diffusion as in Appendix G.1, then $\mathcal{A}v = \mu(y)\partial_y v + \frac{\sigma^2(y)}{2}\partial_{yy} v$. In all cases we assume that the income process is such that y_t is bounded above and below $\underline{y} \leq y_t \leq \bar{y}$ for some positive and finite $\underline{y}$ and $\bar{y} > \underline{y}$ (e.g. there is a finite number of income states or there are reflecting barriers at $\underline{y}$ and $\bar{y}$). For readers who are not familiar with infinitesimal generators and so on, Appendix A.7.1 contains the proof for the special case with two income states.

The proof of the Proposition makes use of the following Lemma that is the natural generalization of Lemma 1 and which derives the Euler equation corresponding to the HJB equation (77).

Lemma 6 *The consumption and saving policy functions $c(a, y)$ and $s(a, y)$ corresponding to the HJB equation (77) satisfy*

$$(\rho - r)u'(c) = u''(c)(\partial_a c)s + \mathcal{A}u'(c), \quad s = y + ra - c.$$

Proof: Differentiate the HJB equation (77) with respect to a (envelope condition) and use that $\partial_a v(a, y) = u'(c(a, y))$ and hence $\partial_{aa} v = u''(c)\partial_a c$.□

In the proofs below it will be useful to rearrange this equation as

$$u''(c)(\partial_a c)s = (\rho - r)u'(c) - \mathcal{A}u'(c). \tag{79}$$

Saving Behavior as $r \downarrow -\infty$. Consider the Euler equation (79). We have that $u'(c) > 0$ and $\mathcal{A}u'(c) < \infty$ for all (a, y) and therefore, as $r \downarrow -\infty$, the right-hand side of (79) is strictly positive for all income types $\underline{y} \leq y \leq \bar{y}$ (in fact it converges to $+\infty$). Since $-\infty < u''(c) < 0$ and $0 \leq \partial_a c(a, y) < \infty$ for all $a > \underline{a}$ and all $\underline{y} \leq y \leq \bar{y}$ it follows from (79) that $s(a, y) < 0$ for all $a > \underline{a}$ and all $\underline{y} \leq y \leq \bar{y}$. Therefore, all individuals decumulate wealth and hence aggregate saving in any stationary distribution (its first moment) must satisfy

$$\lim_{r \downarrow -\infty} S(r) = \underline{a}. \tag{80}$$

Saving Behavior as $r \uparrow \rho$. Consider the Euler equation (79). We claim that if $r = \rho$ the right-hand side of (79) is strictly negative for the highest income type $\bar{y}$. To see this, note that by the definition of the infinitesimal generator:

$$\mathcal{A}u'(c(a,y)) = \lim_{\Delta t \to 0} \frac{\mathbb{E}[u'(c(a,y_{\Delta t}))] - u'(c(a,y_0))}{\Delta t} \quad \text{with } y_0 = y.$$

That is $\mathcal{A}u'(c)$ denotes the expected instantaneous change in marginal utility due to income fluctuations over an infinitesimal time interval. At the upper bound $\bar{y}$, income cannot increase whereas it might decrease with strictly positive probability. Hence consumption decreases and marginal utility increases with strictly positive probability, i.e. the expected instantaneous change in marginal utility is strictly positive: $\mathcal{A}u'(c(a,\bar{y})) > 0$ for all a. On the other hand, $u''(c) < 0$ and $u'(c) > 0$ for all c while $\partial_a c \geq 0$ for all (a,y). Consequently, the highest income type always accumulates wealth regardless of her current wealth level, $s(a,\bar{y}) > 0$ for all a, and hence

$$\lim_{r \uparrow \rho} S(r) = \infty. \tag{81}$$

Continuity of S. First note that the optimal saving policy function $s(a,y;r)$ is continuous in r because it is the policy function of the HJB equation (77) which depends on r in a continuous fashion. We next show that this implies that the stationary wealth distribution is continuous in r. For a given trajectory of income realizations $Y = \{y_t\}_{t \geq 0}$, denote by $a_t^Y(r)$ the solution to $\dot{a}_t = s(a_t, y_t; r)$ *for a fixed initial condition* a_0. Given that the saving policy function s is continuous in r, then so is the wealth trajectory $a_t^Y(r)$ for any given income trajectory Y and at all times $t \geq 0$. Denote by $G(a,y;r)$ the CDF of the stationary joint distribution of income and wealth. Further denote by $G(a;r)$ the unconditional stationary CDF of wealth (in particular $G(a;r) = \int G(a,y;r)dy$ if y is continuous and $G(a;r) := \sum_y G(a,y;r)$ if y is discrete). By its definition, the stationary wealth distribution is the stationary distribution of the process $a_t^Y(r)$. Because a_t^Y is continuous in r for all Y and t so is G. Finally, given that the stationary wealth distribution G is continuous in r so is aggregate saving $S(r)$ in the stationary distribution (the first moment of that distribution).

Conclusion of proof. Given that S is continuous in r and satisfies (80) and (81), it must intersect zero at least once, proving the existence of a stationary equilibrium.□

A.7.1 Proof of Proposition 4 for Special Case with Two-State Poisson Process

For readers who are not familiar with the apparatus employed above (especially infinitesimal generators) we here sketch the proof of Proposition 4 for the special case with a two-state Poisson process as in the baseline Huggett model covered in the main text.

First note that $S(r)$ defined in (11) is continuous in r. This is because individual saving policy functions $s_j(a;r), j = 1,2$, i.e. the optimal controls in (7), are continuous as functions of r. From (33) therefore also the stationary densities $g_j(a;r), j = 1,2$ are continuous in r and hence so is $S(r)$ in (11).

Now consider individual saving behavior $s_1(a;r)$ and $s_2(a;r)$ which is characterized by (50) and (57) in the proofs of Propositions 1 and 2. First, consider the case $r \downarrow -\infty$. As argued in the proof of Proposition 1, $s_1(a;r) < 0$ for all $a > \underline{a}$ for all $r < \rho$. Next, consider $s_2(a;r)$ in (57). As $r \downarrow -\infty$, the right-hand side of(57) becomes strictly positive for all $a > \underline{a}$ and hence $\lim_{r\downarrow-\infty} s_2(a;r) < 0, a > \underline{a}$. Therefore, all individuals decumulate wealth and hence we have (80).

Next, consider the case $r \uparrow \rho$. Since $c_1(a) < c_2(a)$ and hence $u'(c_1(a)) > u'(c_2(a))$ for all $a < \infty$, the right-hand side of (57) becomes strictly negative as $r \uparrow \rho$. Therefore $\lim_{r\uparrow\rho} s_2(a;r) > 0, a < \infty$. Hence, high income types always accumulate assets and one can show using (33) that (81) holds.

Given that S is continuous in r and satisfies (80) and (81), it must intersect zero at least once, proving the existence of a stationary equilibrium.□

A.8 Proof of Proposition 5

As mentioned in Section 3.6, our uniqueness result not only applies to the two-state income process analyzed in Sections 1 and 3 but also to much more general stationary Markovian income processes, e.g. the diffusion process of Appendix G.1. To treat the general case, we consider the environment with a general income process we already analyzed in the proof of Proposition 4 (existence of a stationary equilibrium) in Appendix A.7 and in particular consider the HJB and KF equations in (77) and (78). Readers who are not familiar with the apparatus employed there (in particular infinitesimal generators), can easily follow the proof strategy of Parts 1 and 2 (consumption decreasing in r and saving increasing in r) by setting all terms involving the generator $\mathcal{A}$ equal zero. This corresponds to the case without income uncertainty.

The solutions to the HJB and KF equations v and g as well as the corresponding policy functions c and s depend on r – for example, consumption is $c(a, y; r)$. Even though it is precisely this dependence we are interested in, we suppress it throughout the proof for notational convenience. Hence $\partial c(a, y)/\partial r$ should be understood to mean $\partial c(a, y; r)/\partial r$ and so on.

A.8.1 Proof of Proposition 5, Part 1: Consumption is decreasing in r

We first prove that $c(a, y)$ is strictly decreasing in r for all (a, y) if (35) holds. The proof combines two Lemmas. The first Lemma is due to Olivi (2017).

Lemma 7 (Olivi, 2017) *Consider the HJB equation (77). The corresponding consumption policy function satisfies*

$$\frac{\partial c(a,y)}{\partial r} = \frac{1}{u''(c_0)}\mathbb{E}_0 \int_0^T e^{-\int_0^t \xi_s ds}\{u'(c_t) + u''(c_t)(\partial_a c_t)a_t\}dt \tag{82}$$

with $\xi_t := \rho - r + \partial_a c_t > 0$ and where $T := \inf\{t \geq 0 | a_t = 0\}$ is the stopping time at which wealth reaches the borrowing constraint $\underline{a} = 0$. Here the expectations are over sample paths of (a_t, y_t) starting from $(a_0, y_0) = (a, y)$ and $\partial_a c_t$ is short-hand notation for the instantaneous MPC, $\partial_a c_t = \partial_a c(a_t, y_t)$.

Proof of Lemma 7: Define $\eta(a, y) := \partial_a v(a, y)$. Differentiating (77) we have the envelope condition

$$(\rho - r)\eta = \partial_a \eta(y + ra - c(\eta)) + \mathcal{A}\eta$$

on the interior of the state space and where $c(\eta) = (u')^{-1}(\eta)$. Differentiating with respect to r we have

$$-\eta + (\rho - r)\partial_r \eta = \partial_a[\partial_r \eta]s + \partial_a \eta a - \partial_a \eta c'(\eta)\partial_r \eta + \mathcal{A}\partial_r \eta,$$

where we have used that $s = y + ra - c$. Since $\partial_a c = c'(\eta)\partial_a \eta$, we have $\partial_a \eta c'(\eta)\partial_r \eta = \partial_r \eta \partial_a c$ and hence

$$(\rho - r + \partial_a c)\partial_r \eta = \eta + a\partial_a \eta + \partial_a[\partial_r \eta]s + \mathcal{A}\partial_r \eta. \tag{83}$$

We next evaluate $\partial_r \eta$ in (83) along a particular sample path $(a_t, y_t)_{t\geq 0}$ ("along the characteristic $(a_t, y_t)_{t\geq 0}$") and integrate with respect to time. To this end note that by the appropriate variant of Ito's Formula[52]

$$\mathbb{E}_t[d(\partial_r \eta_t)] = [\partial_a(\partial_r \eta(a_t, y_t))s(a_t, y_t) + \mathcal{A}\partial_r \eta(a_t, y_t)]\, dt. \tag{84}$$

and hence (83) is

$$\xi_t \partial_r \eta_t = \eta_t + a_t \partial_a \eta_t + \frac{1}{dt}\mathbb{E}_t[d(\partial_r \eta_t)], \qquad \xi_t := \rho - r + \partial_a c_t > 0. \tag{85}$$

where we use the short-hand notation $\eta_t = \eta(a_t, y_t)$, $c_t = c(a_t, y_t)$ and so on. For any sample path (a_t, y_t) starting from $(a_0, y_0) = (a, y)$, denote by $T := \inf\{t \geq 0 | a_t = 0\}$ the first time the process a_t hits the borrowing constraint $\underline{a} = 0$. Note that T is a stopping time and itself

[52]First consider the case when y_t follows a diffusion process (114). The sample path $(a_t, y_t)_{t\geq 0}$ is determined by $da_t = s(a_t, y_t)dt, dy_t = \mu(y_t)dt + \sigma(y_t)dW_t$. By Ito's Formula $\partial_r \eta_t = \partial_r \eta(a_t, y_t)$ then follows

$$d(\partial_r \eta_t) = [\partial_a(\partial_r \eta(a_t, y_t))s(a_t, y_t) + \mathcal{A}\partial_r \eta(a_t, y_t)]\, dt + \sigma(y_t)\partial_y(\partial_r \eta(a_t, y_t))dW_t$$

Because the expected increment of a Wiener process is zero, $\mathbb{E}_t[dW_t] = 0$, we have (84). If y_t does not follow a diffusion process, the second term in the equation in this footnote is more complicated (e.g. it will feature jumps). However it still has an expectation of zero and hence (84) still holds.

a random variable. Integrating (85), we have that for any $\tau < T$

$$\partial_r \eta_0 = \mathbb{E}_0 \left[\int_0^\tau e^{-\int_0^t \xi_s ds} \{\eta_t + a_t \partial_a \eta_t\} dt + e^{-\int_0^\tau \xi_s ds} \partial_r \eta_\tau \right]. \tag{86}$$

Now consider the limit as $\tau \to T$. First, recall that from the state constraint boundary condition (15) we have $\eta(a_T, y_T) = u'(y_T + r a_T)$ and therefore $\partial_r \eta(a_T, y_T) < \infty$. On the other hand, as $\tau \to T$, we have $\partial_a c(a_\tau, y_\tau) \to \infty$ and therefore $\xi_\tau \to \infty$. Hence for any sample path (a_t, y_t) and corresponding stopping time T, we have

$$\lim_{\tau \to T} e^{-\int_0^\tau \xi_s ds} \partial_r \eta_\tau = 0.$$

Therefore (86) implies

$$\partial_r \eta_0 = \mathbb{E}_0 \left[\int_0^T e^{-\int_0^t \xi_s ds} \{\eta_t + a_t \partial_a \eta_t\} dt \right]$$

and from the first-order condition $\eta_t = u'(c_t)$ we immediately obtain (82).□

Lemma 8 *Assume that the IES is weakly greater than one, i.e.* (35) *holds. Then* $u'(c(a,y)) + u''(c(a,y))\partial_a c(a,y) a > 0$ *for all* $a \geq 0$ *and all* y.

Proof of Lemma 8: We have

$$\begin{aligned} u''(c(a,y))\partial_a c(a,y)a + u'(c(a,y)) &= -u''(c(a,y))(\text{IES}(c(a,y))c(a,y) - \partial_a c(a,y)a) \\ &\geq -u''(c(a,y))(c(a,y) - \partial_a c(a,y)a) \\ &\geq -u''(c(a,y))c(0,y) \\ &> 0 \quad \text{for all } a > 0. \end{aligned}$$

The equality uses that $u'(c) = -\text{IES}(c)u''(c)c$ from the definition of the IES in (35). The first weak inequality uses that the IES is greater than one from (35). The second weak inequality uses the weak concavity of the consumption function: because c is weakly concave in a, we have $c(a,y) \geq c(0,y) + \partial_a c(a,y)a$ for all $a \geq 0$.[53] The strict inequality at the end uses that $c(0,y) > 0$ for all y. □

Conclusion of Proof of Proposition 5, Part 1: The proof of Part 1 concludes by combining Lemmas 7 and 8. From Lemma 7 we see that $c(a,y)$ is strictly decreasing in r if (i) $T > 0$ so that the integral in (82) is different from zero, and if (ii) $u'(c_t) + u''(c_t)(\partial_a c_t)a_t > 0$

[53]There are two easy ways of seeing this. First, graphically. Second, from the observation that any concave function is bounded above by its first-order Taylor-series approximation: for any fixed (a,y), $c(b,y) \leq c(a,y) + \partial_a c(a,y)(b-a)$ for all b. Taking $b = 0$ we have $c(0,y) \leq c(a,y) - \partial_a c(a,y)a$ as claimed.

point-by-point in the integral in (82) over sample paths $(a_t, y_t)_{0\le t\le T}$. Requirement (i) that $T > 0$ holds if $a_0 > 0$. Requirement (ii) holds if $u'(c(a,y)) + u''(c(a,y))\partial_a c(a,y)a > 0$ for all (a,y) on the interior of the state space. But we have shown in Lemma 8 that a sufficient condition for this is that the IES is weakly greater than one.

Finally, it is interesting to note that Part 1 of the Proposition does not require the assumption of a strict no-borrowing constraint $a \ge 0$: because $u''(c) < 0$, $u'(c(a,y)) + u''(c(a,y))\partial_a c(a,y)a > 0$ for all $a < 0$, independently of Lemma 8. Hence consumption is strictly decreasing in r even if we allow for borrowing, $a < 0$.□

A.8.2 Proof of Proposition 5, Part 2: Saving is increasing in r

That $s(a,y)$ is strictly increasing in r for all $a > 0$ follows immediately from the budget constraint $s(a,y) = y + ra - c(a,y)$ and that consumption is strictly decreasing in r as shown in Part 1:

$$\frac{\partial s(a,y)}{\partial r} = a - \frac{\partial c(a,y)}{\partial r} > 0, \quad a > 0.$$

Note that the assumption of a strict no-borrowing limit is only needed in this part of the proof: if $a < 0$ we cannot sign $\partial s(a,y)/\partial r$.

A.8.3 Proof of Proposition 5, Part 3: First-order Stochastic Dominance

We have shown thus far that wealth a_t evolves according to $\dot{a}_t = s(a_t, y_t; r)$ for some exogenous stochastic process y_t, and where the function s is continuous and differentiable in a and r and satisfies $\partial s/\partial r > 0$ for all (a,y). We next show that this implies that the stationary wealth distribution implied by a high r first-order stochastically dominates that implied by a low r. As in the proof of Proposition 4, denote by $a_t^Y(r)$ the solution to $\dot{a}_t = s(a_t, y_t; r)$ for a given trajectory of income realizations $Y = \{y_t\}_{t\ge 0}$ and *for a fixed initial condition* a_0. With this notation in hand, we first state a simple Lemma.

Lemma 9 *Consider two interest rates r_h and r_ℓ with $r_h > r_\ell$. Then, for each given income trajectory $Y = \{y_t\}_{t\ge 0}$, the higher interest rate implies a higher wealth trajectory, $a_t^Y(r_h) > a_t^Y(r_\ell)$ for all $t > 0$.*

Proof of Lemma: For given Y and r, $a_t^Y(r)$ is the solution to a simple and well-behaved ordinary differential equation and therefore a continuous function of time t. At $t = 0$ we have $a_0^Y(r_h) = a_0^Y(r_\ell) = a_0$ by assumption. Since also the income trajectories are assumed to be the same we have $\dot{a}_0^Y(r_h) = s(a_0, y_0; r_h) > s(a_0, y_0; r_\ell) = \dot{a}_0^Y(r_\ell)$, i.e. $a_t^Y(r_h)$ is initially above $a_t^Y(r_\ell)$. For the Lemma to be false, there would therefore need to be a time $T \in (0,\infty)$ such that the two trajectories meet again, $a_T^Y(r_h) = a_T^Y(r_\ell)$, and at which the trajectory corresponding to r_ℓ approaches that corresponding to r_h from below, $\dot{a}_T^Y(r_h) \leq \dot{a}_T^Y(r_\ell)$. But this is impossible given that $\dot{a}_T^Y(r_h) = s(a_T, y_T; r_h) > s(a_T, y_T; r_\ell) = \dot{a}_T^Y(r_\ell)$.□

Conclusion of Proof of Part 3: To conclude the proof, denote by $G(a, y; r)$ the CDF of the stationary joint distribution of income and wealth. Further denote by $G(a; r)$ the unconditional stationary CDF of wealth (in particular $G(a; r) = \int G(a, y; r)dy$ if y is continuous and $G(a; r) := \sum_y G(a, y; r)$ if y is discrete). Similarly, denote by $G_t(a; r)$ the wealth distribution at a given point in time $t \geq 0$, starting from a given initial wealth a_0, $G_t(a; r) = \Pr(a_t^Y(r) \leq a)$. If G is a stationary distribution, then $\lim_{t\to\infty} G_t(a; r) = G(a; r)$. Finally, denote by $\Omega_t(r)$ the set of income trajectories Y such that $a_t^Y(r)$ is below a scalar a at time t, i.e. $\Omega_t(r) := \{Y \text{ such that } a_t^Y(r) \leq a\}$. Since from Lemma 9 $a_t^Y(r_h) > a_t^Y(r_\ell), t > 0$ we have $\Omega_t(r_h) \subseteq \Omega_t(r_\ell)$. Since $\Pr(a_t^Y(r) \leq a) = \Pr(\Omega_t(r))$, we have $\Pr(a_t^Y(r_h) \leq a) \leq \Pr(a_t^Y(r_\ell) \leq a)$ for all $t > 0$. Since the statement holds for all $t > 0$ it also holds as $t \to \infty$ and therefore $G(a; r_h) \leq G(a; r_\ell)$.

A.8.4 Proof of Proposition 5, Part 4: $S(r)$ is increasing in r

Lemma 9 also immediately implies uniqueness of the stationary equilibrium. In particular aggregate saving $S(r)$ is the first moment of the stationary wealth distribution $S(r) = \lim_{t\to\infty} \mathbb{E}[a_t^Y(r)]$ where the expectation is taken over all possible income realizations Y. For $r_h > r_\ell$, since from Lemma 9 $a_t^Y(r_h) > a_t^Y(r_\ell), t > 0$ *trajectory by trajectory*, then also $S(r_h) > S(r_\ell)$. Since $S(r)$ is strictly increasing, there can be at most one r solving $S(r) = B$.□

B Derivation of HJB and KF Equations

This Appendix shows how to derive the HJB equation with Poisson shocks (7) and that with a diffusion process (115) as well as the Kolmogorov Forward or Fokker-Planck equation with Poisson shocks (13) from a discrete-time environment with time periods of length Δ and then taking the limit as $\Delta \to 0$.

B.1 Hamilton-Jacobi-Bellman Equation with Poisson Process

Consider the following income fluctuation problem in discrete time. Periods are of length Δ, individuals discount the future with discount factor $\beta(\Delta) = e^{-\rho\Delta}$, and individuals with income y_j keep their income with probability $p_j(\Delta) = e^{-\lambda_j\Delta}$ and switch to state y_{-j} with probability $1-p_j(\Delta)$. The Bellman equation for this problem is:

$$v_j(a_t) = \max_c \, u(c)\Delta + \beta(\Delta)\left(p_j(\Delta)v_j(a_{t+\Delta}) + (1-p_j(\Delta))v_{-j}(a_{t+\Delta})\right) \quad \text{s.t.} \tag{87}$$

$$a_{t+\Delta} = \Delta(y_j + ra_t - c) + a_t \tag{88}$$

$$a_{t+\Delta} \geq \underline{a} \tag{89}$$

for $j = 1, 2$. We will momentarily take $\Delta \to 0$ so we can use that for Δ small

$$\beta(\Delta) = e^{-\rho\Delta} \approx 1 - \rho\Delta, \quad p_j(\Delta) = e^{-\lambda_j\Delta} \approx 1 - \lambda_j\Delta.$$

Substituting these into (87) we have

$$v_j(a_t) = \max_c \; u(c)\Delta + (1-\rho\Delta)\left((1-\Delta\lambda_j)v_j(a_{t+\Delta}) + \Delta\lambda_j v_{-j}(a_{t+\Delta})\right)$$

subject to (91) and (89). Subtracting $(1-\rho\Delta)v_j(a)$ from both sides and rearranging, we get

$$\Delta\rho v_j(a_t) = \max_c \; u(c)\Delta + (1-\rho\Delta)\left(v_j(a_{t+\Delta}) - v_j(a) + \Delta\lambda_j(v_{-j}(a_{t+\Delta}) - v_j(a_{t+\Delta}))\right)$$

subject to (91) and (89). Dividing by Δ, taking $\Delta \to 0$ and using that

$$\lim_{\Delta\to 0} \frac{v_j(a_{t+\Delta}) - v_j(a)}{\Delta} = \lim_{\Delta\to 0} \frac{v_j(\Delta(y_j + ra_t - c) + a_t) - v_j(a_t)}{\Delta} = v_j'(a_t)(y_j + ra_t - c)$$

yields (7) where we drop the t-subscripts on a_t for notational simplicity. Note also that the borrowing constraint (89) never binds in the interior of the state space because with Δ arbitrarily small $a_t > \underline{a}$ implies $a_{t+\Delta} > \underline{a}$. The time-dependent case (12) can be derived in an analogous fashion, and the derivation can be generalized to any number of income states $J > 2$.

B.2 Hamilton-Jacobi-Bellman Equation with Diffusion Process

Consider a more general version of the discrete-time Bellman equation in Section B.1 but where now y_t is assumed to follow any general Markov process

$$v(a_t, y_t) = \max_c \, u(c)\Delta + \beta(\Delta)\mathbb{E}[v(a_{t+\Delta}, y_{t+\Delta})] \quad \text{s.t.} \tag{90}$$

$$a_{t+\Delta} = \Delta(y_t + ra_t - c) + a_t \tag{91}$$

$$a_{t+\Delta} \geq \underline{a}$$

where $\mathbb{E}[\cdot]$ is the appropriate expectation over $y_{t+\Delta}$. Then, following similar steps as in Section B.1, we get

$$\rho v(a_t, y_t) = \max_c \; u(c) + \frac{\mathbb{E}[dv(a_t, y_t)]}{dt} \tag{92}$$

Now assume y_t follows a diffusion process (114) and wealth follows (2). Then by Ito's Lemma

$$\begin{aligned} dv(a_t, y_t) = & \left(\partial_a v(a_t, y_t)(y_t + ra_t - c_t) + \partial_y v(a_t, y_t)\mu(y_t) + \frac{1}{2}\partial_{yy} v(a_t, y_t)\sigma^2(y_t) \right) dt \\ & + \partial_y v(a_t, y_t)\sigma(y_t) dW_t. \end{aligned}$$

Therefore, using that the expectation of the increment of a standard Brownian motion is zero, $\mathbb{E}[dW_t] = 0$,

$$\mathbb{E}[dv(a_t, y_t)] = \left(\partial_a v(a_t, y_t)(y_t + ra_t - c_t) + \partial_y v(a_t, y_t)\mu(y_t) + \frac{1}{2}\partial_{yy} v(a_t, y_t)\sigma^2(y_t) \right) dt$$

Substituting into (92) yields (115). For completeness, note that the connection to the Poisson HJB equation (7) is that with a two-state Poisson process with states y_j and intensities λ_j

$$\mathbb{E}[dv_j(a_t)] = \left(v_j'(a_t)(y_j + ra_t - c_t) + \lambda_j (v_{-j}(a_t) - v_j(a_t)) \right) dt, \quad j = 1, 2$$

and hence substituting into (90) yields (7).

B.3 Kolmogorov Forward Equation with Poisson Process

First recall the continuous-time economy. There is a continuum of individuals who are heterogeneous in their wealth a and their income y. To avoid confusion, we here adopt the convention that variables with tilde superscripts, $\tilde{a}_t$ and $\tilde{y}_t$, denote stochastic variables and variables without superscripts denote the values these can take. Income takes two values $\tilde{y}_t \in \{y_1, y_2\}$ and follows a two-state Poisson process with intensities λ_1 and λ_2. Wealth evolves as

$$d\tilde{a}_t = s_j(\tilde{a}_t, t)dt \tag{93}$$

where the optimal saving policy function s_j is derived from individuals' utility maximization problem. The state of the economy is the density $g_j(a, t), j = 1, 2$.

Now consider the discrete-time analogue. The timing of events over a time period of length Δ is as follows: individuals of type $j = 1, 2$ first make their saving decisions according to the discrete-time analogue of (93)

$$\tilde{a}_{t+\Delta} = \tilde{a}_t + \Delta s_j(\tilde{a}_t) \tag{94}$$

where we suppress the dependence of s_j on t for notational simplicity. After saving decisions are made, next period's income $\tilde{y}_{t+\Delta}$ is realized: it switches from y_j to y_{-j} with probability $\Delta\lambda_j$.

It turns out to be easiest to work with the CDF (in the wealth dimension)

$$G_j(a,t) = \Pr(\tilde{a}_t \leq a, \tilde{y}_t = y_j). \tag{95}$$

This is the fraction of people with income y_j and wealth below a. It satisfies $G_1(\underline{a},t) + G_2(\underline{a},t) = 0$ and $\lim_{a\to\infty}(G_1(a,t)+G_2(a,t)) = 1$. The density g_j satisfies $g_j(a,t) = \partial_a G_j(a,t)$.

In order to derive a law of motion for G, consider first the wealth accumulation process. In particular, we will need an answer to the question: if a type j individual has wealth $\tilde{a}_{t+\Delta}$ at time $t+\Delta$, then what level of wealth $\tilde{a}_t$ did she have at time t? To this end, it turns out to be convenient to work not with (94) but with another (equally correct) discrete-time analogue of (93):[54]

$$\tilde{a}_t = \tilde{a}_{t+\Delta} - \Delta s_j(\tilde{a}_{t+\Delta}) \tag{96}$$

Intuitively, if the individual dissaves such that $s_j < 0$, her past wealth must have been larger than her current wealth. Now consider the fraction of individuals with wealth below a at date $t+\Delta$. Momentarily ignoring that some individuals' incomes switch and assuming that individuals decumulate wealth $s_j(a) \leq 0$ (the case with $s_j(a) > 0$ is symmetric), we have

$$\Pr(\tilde{a}_{t+\Delta} \leq a) = \underbrace{\Pr(\tilde{a}_t \leq a)}_{\text{already below threshold } a} + \underbrace{\Pr(a \leq \tilde{a}_t \leq a - \Delta s_j(a))}_{\text{cross threshold } a} = \Pr(\tilde{a}_t \leq a - \Delta s_j(a)).$$

Next also taking into account income switches, the fraction of individuals with wealth below a evolves as follows:

$$\begin{aligned}\Pr(\tilde{a}_{t+\Delta} \leq a, \tilde{y}_{t+\Delta} = y_j) = (1-\Delta\lambda_j)\Pr(\tilde{a}_t \leq a - \Delta s_j(a), \tilde{y}_t = y_j) \\ +\Delta\lambda_{-j}\Pr(\tilde{a}_t \leq a - \Delta s_{-j}(a), \tilde{y}_t = y_{-j})\end{aligned} \tag{97}$$

Using the definition of G_j in (95), we then have

$$G_j(a,t+\Delta) = (1-\Delta\lambda_j)G_j(a-\Delta s_j(a),t) + \Delta\lambda_{-j}G_{-j}(a-\Delta s_{-j}(a),t)$$

Subtracting $G_j(a,t)$ from both sides and dividing by Δ

$$\frac{G_j(a,t+\Delta) - G_j(a,t)}{\Delta} = \frac{G_j(a-\Delta s_j(a),t) - G_j(a,t)}{\Delta} - \lambda_j G_j(a-\Delta s_j(a),t) + \lambda_{-j}G_{-j}(a-\Delta s_{-j}(a),t)$$

[54] Note that from (93) $\tilde{a}_{t+\Delta} = \int_t^{t+\Delta} s_j(\tilde{a}_\tau)d\tau + \tilde{a}_t$. The integral is approximately equal to both $\Delta s_j(\tilde{a}_t)$ and $\Delta s_j(\tilde{a}_{t+\Delta})$ and therefore both (94) and (96) are meaningful discrete-time analogues. The difference is that the former looks forward in time and the latter looks backward in time.

Taking the limit as $\Delta \to 0$ gives

$$\partial_t G_j(a,t) = -s_j(a)\partial_a G_j(a,t) - \lambda_j G_j(a,t) + \lambda_{-j} G_{-j}(a,t), \tag{98}$$

where we have used that

$$\lim_{\Delta\to 0} \frac{G_j(a - \Delta s_j(a), t) - G_j(a,t)}{\Delta} = \lim_{x\to 0} \frac{G_j(a - x, t) - G_j(a,t)}{x} s_j(a) = -s_j(a)\partial_a G_j(a,t),$$

Differentiating with respect to a and using the definition of the density as $g_j(a,t) = \partial_a G_j(a,t)$, we obtain (13).

Equation (98) is the Kolmogorov Forward equation written in terms of the CDF $G_j(a,t)$ and it is entirely intuitive. The first term captures inflows and outflows due to continuous movements in wealth a, and the second and third terms capture inflows and outflows due to jumps in income y_j. To understand the first term, $-s_j(a)\partial_a G_j(a,t)$, consider the case where at a given point a and income y_j, savings are negative $s_j(a) < 0$. In that case, the fraction of individuals with wealth below a and income equal to y_j, $G_j(a,t) = \Pr(\tilde{a}_t \leq a, \tilde{y}_t = y_j)$, increases at a rate proportional to the density of individuals exactly at that point a, $g_j(a,t) = \partial_a G_j(a,t) = \Pr(\tilde{a}_t = a, \tilde{y}_t = y_j)$, i.e. there is an inflow of individuals into wealth levels below a. The reverse logic applies if $s_j(a) > 0$.

B.4 Kolmogorov Forward Equation with Diffusion Process

We are not aware of any intuitive derivations of the Kolmogorov Forward (Fokker-Planck) equation with a diffusion process (116). One relatively accessible derivation is provided by Kredler (2014).

C General Heterogeneous Agent Models: Mean Field Games in n Dimensions

We here spell out the backward-forward MFG system in n dimensions which is a natural generalization of the equations for the Aiyagari-Bewley-Huggett model. The mathematics MFG literature typically writes this system using the language of the modern theory of PDEs, in particular vector calculus notation. See for example Lasry and Lions (2007), Cardaliaguet (2013), Bertucci, Lasry, and Lions (2018) and Ryzhik (2018). In order to make this literature accessible to economists, we here make the connection between our formulation and the standard MFG notation. For background readings on modern PDE theory, see the short introduction by Evans (2008) as well as the book by the same author, Evans (2010).

C.1 Preliminaries: Vector Calculus Notation

The mathematics literature usually writes Mean Field Games using vector calculus notation and we will make this connection toward the end of this section. To this end, we define three useful operators: the gradient ∇, the Laplacian Δ and the divergence div. First, for a function $f : \mathbb{R}^n \to \mathbb{R}$, the *gradient vector* is the vector of first derivatives

$$\nabla f := \left[\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n}\right]^{\mathrm{T}}.$$

Second, for a function $f : \mathbb{R}^n \to \mathbb{R}$, the *Laplacian* is the sum of all the unmixed second derivatives

$$\Delta f := \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}.$$

Third, for a vector-valued function $\mathbf{v} : \mathbb{R}^n \to \mathbb{R}^n$, i.e. $\mathbf{v}(x_1, ..., x_n) = [v_1(x_1, ..., x_n), ..., v_n(x_1, ..., x_n)]^{\mathrm{T}}$, the *divergence* of $\mathbf{v}$ is

$$\mathrm{div}(\mathbf{v}) := \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i}.$$

A useful check on one's understanding of these definitions is to verify that $\Delta f = \mathrm{div}(\nabla f)$.

C.2 Backward-Forward MFG System in n Dimensions

The mathematics literature typically only considers the case where state variables follow diffusion processes rather than processes featuring jumps. Under this assumption, a general backward-forward MFG system in n dimensions is:

$$\begin{aligned}
&\rho v = \max_{\alpha}\{r(x,\alpha,g) + \sum_{i=1}^{n} \alpha_i \partial_i v\} + \frac{1}{2}\sum_{i=1}^{n} \sigma_i^2(x)\partial_{ii} v + \partial_t v \quad \text{in } \mathbb{R}^n \times (0,T) \\
&\partial_t g = -\sum_{i=1}^{n} \partial_i(\alpha_i^*(x,g)g) + \frac{1}{2}\sum_{i=1}^{n} \partial_{ii}(\sigma_i^2(x)g) \quad \text{in } \mathbb{R}^n \times (0,T) \\
&g(0) = g_0 \quad \text{and} \quad v(x,T) = V(x, g(T)) \quad \text{in } \mathbb{R}^n.
\end{aligned}$$

Here $v(x,t)$ is the value function, $g(x,t)$ the density, $x \in \mathbb{R}^n$ an n-dimensional state vector, $r(x,\alpha,g)$ a period return function, $\sigma_i^2(x)$ a diffusion coefficient, $\alpha \in \mathbb{R}^n$ a control vector and α^* its optimally chosen counterpart (the policy function). The first equation is the HJB equation, the second equation is the KF equation and the equations in the third line are the initial condition on the density and the terminal condition on the value function. We here purposely ignore the discussion of boundary conditions on v and g which are typically application-specific.[55] Note that the Huggett model with a diffusion process in Appendix G.1 is a two-dimensional special case, i.e. $x \in \mathbb{R}^2$ with $x_1 = a$ and $x_2 = y$.

C.3 Connection to Standard Notation used in MFG Literature

The MFG literature typically assumes that $\sigma_i^2(x) = 2\nu$ for all x and all i which implies that the second-order terms simplify. Although this is not necessary, we will also make this assumption from now. Furthermore, and as already discussed in footnote 18 in the main text, it is useful to work with the *Hamiltonian* $H(x, p, g) := \max_\alpha \{r(x, \alpha, g) + \sum_{i=1}^n \alpha_i p_i\}$. Hence:

$$\begin{aligned}
\rho v &= H(x, \nabla v, g) + \nu \sum_{i=1}^n \partial_{ii} v + \partial_t v \quad \text{in } \mathbb{R}^n \times (0, T) \\
\partial_t g &= -\sum_{i=1}^n \partial_i (\partial_{p_i} H(x, \nabla v, g)\ g) + \nu \sum_{i=1}^n \partial_{ii} g \quad \text{in } \mathbb{R}^n \times (0, T) \\
g(0) &= g_0, \quad v(x, T) = V(x, g(T)) \quad \text{in } \mathbb{R}^n.
\end{aligned}$$

Note that the KF equation uses that, from the definition of the Hamiltonian and the envelope theorem, the optimal drift of each state variable equals $\alpha_i^*(x, g) = \partial_{p_i} H(x, \nabla v, g)$. Finally, using the Laplacian and the divergence defined in Appendix C.1, we have the standard formulation of the backward-forward MFG system:

$$\begin{aligned}
\rho v &= H(x, \nabla v, g) + \nu \Delta v + \partial_t v \quad \text{in } \mathbb{R}^n \times (0, T) \\
\partial_t g &= -\text{div}(\nabla_p H(x, \nabla v, g)\ g) + \nu \Delta g \quad \text{in } \mathbb{R}^n \times (0, T) \\
g(0) &= g_0, \quad v(x, T) = V(x, g(T)) \quad \text{in } \mathbb{R}^n.
\end{aligned} \tag{99}$$

There are two remaining differences to the standard MFG notation: first, we have written v and g for the value function and density whereas the MFG literature typically uses u and m; and second the MFG literature typically sets $\rho = 0$ for simplicity, i.e. it ignores discounting. See in particular equation (1) in Bertucci, Lasry, and Lions (2018) and equation (1.2) in Ryzhik (2018) for the system (99) but using u and m in place of v and g and with $\rho = 0$.

C.4 MFGs with Aggregate Uncertainty – the "Master Equation"

The backward-forward MFG system (99) describes general heterogeneous agent models without aggregate uncertainty. However, in many economically interesting situations it is important to allow for aggregate risk in addition to idiosyncratic risk (as in Den Haan, 1997;

[55] In the MFG literature the state space is often specified as the n-dimensional torus $\mathbb{T}^n$ (i.e. an n-dimensional analogue of the circle for $n = 1$) rather than $\mathbb{R}^n$. The only reason for this is to sidestep the discussion of boundary conditions in the space dimension.

Krusell and Smith, 1998). Fortunately, the theory of MFGs has also studied that case, with mathematicians referring to aggregate uncertainty as "common noise." In the most general case, such MFGs can be written in terms of a so-called "Master equation" (Cardaliaguet, Delarue, Lasry, and Lions, 2019). This Master equation is an equation on the space of measures, i.e. it is an equation that is set in infinite-dimensional space. The logic why the problem with aggregate uncertainty becomes infinite-dimensional is the same as in discrete-time heterogeneous agent models in the economics literature (Den Haan, 1997; Krusell and Smith, 1998): the cross-sectional distribution across agents becomes a state variable in agents' dynamic programming problems and that distribution is an infinite-dimensional object. In the case without aggregate uncertainty, the Master equation reduces to the "backward-forward MFG system."

Finally, there is also a literature that treats MFGs from a probability-theoretic viewpoint instead of the partial differential equation viewpoint taken here. In particular see the books by Carmona and Delarue (2018a,b) which also feature a number of interesting potential applications of MFGs.

C.5 Existence and Uniqueness Results in the Mathematics Literature

A natural question is whether the mathematics literature contains any "off-the-shelf" results on backward-forward MFG systems that we can use to characterize the economic models we are interested in, in particular with regard to existence and uniqueness of solutions. The answer is "no" unfortunately. While there are several existence and uniqueness results in the literature, none of these apply to the heterogeneous agent models we are interested in studying.

This is because the theoretical results in the literature typically make several additional strong assumptions on the system (99) that rule out Aiyagari-Bewley-Huggett models and many other economically interesting models. For example, there are a number of existence and uniqueness results for the case in which the Hamiltonian H is additively separable in the gradient of the value function ∇v and the distribution g:

$$H(x, \nabla v, g) = \tilde{H}(x, \nabla v) - F(x, g). \tag{100}$$

See Section 4 in Ryzhik (2018) who discusses several of these results.[56] However, for the heterogeneous agent models we are interested in here, the Hamiltonian is typically not separable.

[56]These results typically make additional strong assumptions on the function F. For example, a common assumption is that F is continuous and "monotone" with respect to g in the following sense: $\int (F(x, g_1) - F(x, g_2))(g_1(x) - g_2(x))dx > 0$ for all $g_1 \neq g_2$.

For example, in a Huggett model, the Hamiltonian is defined as

$$H(x, \nabla v, g) = \max_c \{u(c) + \partial_{x_1} v \times (x_2 + r(g)x_1 - c)\}$$

which is clearly not additively separable between ∇v and g (instead there is a multiplicative dependence). Similarly, the MFG systems considered in the mathematics literature typically have not featured state constraints and the resulting Dirac point masses in the distribution g, and this is yet another reason why the standard results in the literature are not applicable to the heterogeneous agent models economists are interested in studying. That being said, recent work by Cannarsa, Capuani, and Cardaliaguet (2018) does treat MFGs with state constraints so this situation may well change in the not-too-distant future.

D Viscosity Solutions for Dummies (including Economists)

See the separate online appendix on Benjamin Moll's website at https://benjaminmoll.com/viscosity_for_dummies/.

E Measure-Valued Solutions to KF Equations, Extended to Allow for Mass on Boundary

We here explain the notion of a measure-valued solution and how to deal with a potential Dirac mass at the borrowing constraint. As for the discussion of viscosity solutions in Appendix D, the discussion aims to intuitively explain key ideas rather than providing a technically rigorous treatment. For a technical discussion of weak solutions to KF equations, see Bogachev, Krylov, Röckner, and Shaposhnikov (2015). Also see Cannarsa, Capuani, and Cardaliaguet (2018) for a rigorous study of MFGs with state constraints in a different setting than ours (namely one with a separable Hamiltonian as in equation (100) in Appendix C.5).

We consider a simplified version of the KF equation (8) (or the counterpart with a continuum of income types (116)). In particular assume that there is no income risk and wealth simply follows $\dot{a}_t = s(a_t)$ where s is a saving policy function (for the purpose of this Appendix it is immaterial whether it comes from an optimization problem). The wealth distribution is then simply a one-dimensional object and we denote the stationary density by $g(a)$ and its time-varying counterpart by $g(a, t)$. These follow the analogues of (8) and (13):

$$0 = -(s(a)g(a))' \quad \text{on } (\underline{a}, \infty), \tag{101}$$

$$\partial_t g(a, t) = -\partial_a (s(a)g(a, t)), \quad \text{on } (\underline{a}, \infty) \times (0, \infty) \quad \text{with } g(a, 0) = g_0(a). \tag{102}$$

Of course this model is "economically boring": the stationary wealth distribution is either degenerate with all mass at $a = \underline{a}$; or it does not exist. Nevertheless, the model is sufficiently rich to explain the appropriate solution concept for (101) and (102).

E.1 Measure-Valued Solution with No Mass on Boundary: Intuition

First, assume away the possibility that g features a Dirac point mass at $\underline{a}$. In this case we can use the standard notion of a measure-valued solution on $(\underline{a}, \infty)$. We extend this notion to allow for mass on the boundary below.

To motivate the notion of a measure-valued solution, consider for the moment the case where the KF equation has a classical solution g. The goal is to obtain a more general equation that also has meaning if this is not the case so that (102) is meaningless in the classical sense. Define the measure $\mu_t(a)$ by $d\mu_t(a) = g(a,t)da$. Also consider a "test function" φ, i.e. a "nice" function that is infinitely differentiable and assume, for now, that φ vanishes at the boundaries as $a \to \underline{a}$ or ∞. Next differentiate $\int_{\underline{a}}^{\infty} \varphi(a) d\mu_t(a)$ with respect to time, use the KF equation (102), and then integrate by parts, to get[57]

$$\frac{d}{dt} \int_{\underline{a}}^{\infty} \varphi(a) d\mu_t(a) = \int_{\underline{a}}^{\infty} \varphi'(a) s(a) d\mu_t(a). \tag{103}$$

Now comes the key observation: we have shown that the KF equation (102) implies that (103) holds for any test function φ. But the converse is not true, i.e. (103) is a weaker requirement than the KF equation (102). In particular (103) still has an interpretation if the KF equation (102) is meaningless because g does not exist or is not differentiable. All that is required is that the distribution admits a measure μ_t.

Summarizing, we say that $g(\cdot, t)$ is a measure-valued solution to (102) if the corresponding measure μ_t satisfies (103) for all test functions φ (that vanish at the boundaries). The stationary counterpart is obvious: we say that g is a measure-valued solution to (101) if the corresponding measure μ satisfies

$$0 = \int_{\underline{a}}^{\infty} \varphi'(a) s(a) d\mu(a) \tag{104}$$

[57] That is, we follow these steps:

$$\begin{aligned} \frac{d}{dt} \int_{\underline{a}}^{\infty} \varphi(a) d\mu_t(a) &= \int_{\underline{a}}^{\infty} \varphi(a) \partial_t g(a,t) da = \int_{\underline{a}}^{\infty} \varphi(a) \left[-\partial_a (s(a) g(a,t)) \right] da \\ &= \int_{\underline{a}}^{\infty} \varphi'(a) s(a) g(a,t) da = \int_{\underline{a}}^{\infty} \varphi'(a) s(a) d\mu_t(a) \end{aligned}$$

.

for all test functions φ. This solution concept is entirely intuitive: after all, the economic/physical object we are modeling with the KF equation is exactly a measure. As Bogachev, Krylov, Röckner, and Shaposhnikov (2015) state on the first page of their treatise on KF equations: "it is crucial to understand that a priori Fokker-Planck-Kolmogorov equations are equations for measures, not for functions."

E.2 Measure-Valued Solution with No Mass on Boundary: Derivation

In fact, (103) is exactly the equation that shows up in the derivation of the KF equation from first principles (see e.g. Kredler, 2014). That is, the argument above is basically the derivation of the KF equation in reverse and (103) is an intermediate step. We briefly go through this derivation. Define the measure μ_t such that for all test functions φ (that vanish at $\underline{a}$ and ∞), the expected value of $\varphi(a_t)$ can be computed as

$$\mathbb{E}[\varphi(a_t)] = \int_{\underline{a}}^{\infty} \varphi(a) d\mu_t(a).$$

In particular for $\varphi \equiv 1$ we have the familiar normalization condition $\int_{\underline{a}}^{\infty} d\mu_t(a) = 1$ for all t, i.e. the total mass equals one. Differentiating with respect to t, we then have

$$\frac{d}{dt}\int_{\underline{a}}^{\infty} \varphi(a) d\mu_t(a) = \frac{d}{dt}\mathbb{E}[\varphi(a_t)] = \mathbb{E}[\varphi'(a_t)s(a_t)] = \int_{\underline{a}}^{\infty} \varphi'(a)s(a)d\mu_t(a)$$

which is exactly condition (103). To derive the KF equation (101) we then usually *assume* that μ_t admits a density so that $d\mu_t(a) = g(a,t)da$. Therefore

$$\int_{\underline{a}}^{\infty} \varphi(a)\partial_t g(a,t)da = \int_{\underline{a}}^{\infty} \varphi'(a)s(a)g(a,t)da.$$

We then integrate by parts, basically following the steps in footnote 57 in reverse and using that φ vanishes at the boundaries to zero out the boundary terms, to get

$$\int_{\underline{a}}^{\infty} \varphi(a)[\partial_t g(a,t) + \partial_a(s(a)g(a,t))]da = 0.$$

This then implies (101). The derivation of a measure-valued solution of (101) basically stops half-way through this derivation – at (103) – and therefore does not require the assumption that μ admits a density.

E.3 Measure-Valued Solution with Mass on Boundary

Next consider our solution concept for the KF equation when there can be a Dirac mass at the boundary. The extension is straightforward. In particular the definition is exactly identical to the one for the case without mass at the boundary except for one crucial change: we now assume that (103) and (104) *must also hold* for test functions φ that *do not vanish at* $a = \underline{a}$, i.e. even if $\varphi(\underline{a}), \varphi'(\underline{a}) \neq 0$.

E.4 Two Applications: Dirac Mass at Boundary or in Interior

We briefly demonstrate the usefulness of this apparatus by considering two special cases that arise in the economic problems considered in this paper.

Application 1: Dirac Mass on Boundary but Density in Interior In the Huggett model of Sections 1 and 3, the wealth distribution typically has a Dirac point mass at the borrowing constraint $\underline{a}$ but admits a smooth density for all $a > \underline{a}$. Motivated by this observation we look for a stationary measure μ as

$$d\mu(a) = g(a)d\mathcal{L}(a) + m\delta_{\underline{a}}, \tag{105}$$

Here $\mathcal{L}$ is the Lebesgue measure on $(\underline{a}, \infty)$ and g is a Lebesgue-integrable non-negative real-valued function on $(\underline{a}, \infty)$ which we call the density of wealth a. Similarly, $\delta_{\underline{a}}$ is the Dirac delta function at $a = \underline{a}$ and m is a non-negative real-valued scalar which we call the Dirac point mass at $a = \underline{a}$. Hence

$$\int_{\underline{a}}^{\infty} \varphi(a)d\mu(a) = \varphi(\underline{a})m + \int_{\underline{a}}^{\infty} \varphi(a)g(a)d\mathcal{L}(a) \tag{106}$$

for all φ and, in particular, $1 = m + \int_{\underline{a}}^{\infty} g(a)d\mathcal{L}(a)$. Further (104) becomes

$$0 = \int_{\underline{a}}^{\infty} \varphi'(a)s(a)g(a)d\mathcal{L}(a) + \varphi'(\underline{a})s(\underline{a})m. \tag{107}$$

Integrating the first term by parts, we can again see that, in $(\underline{a}, \infty)$, g is a measure-valued solution to (101). The difference is the second term.[58]

Summarizing, whenever we make statements of the sort, "g is a solution to (101)", the precise meaning is that there is a non-negative real-valued function g and a non-negative

[58]Note that it is not possible to derive an explicit boundary condition for g at $a = \underline{a}$ because the term $\varphi'(\underline{a})s(\underline{a})m$ cannot be expressed as a term multiplying $\varphi(\underline{a})$ (e.g. it is not possible to integrate it by parts – in mathematics language: $\varphi'(\underline{a})s(\underline{a})m$ cannot be expressed "as a distribution acting on $\varphi(\underline{a})$").

real-valued scalar m that satisfy (107) for any test function φ, including ones that do not vanish at $a = \underline{a}$.

Application 2: Dirac Mass in Interior The model with soft borrowing constraints in Appendix G.3 gives rise to an alternative scenario: there is a Dirac mass at $a = 0$ and a smooth density both to the left and the right of zero. See Figure 17(b). In this case we look for a stationary measure μ as $d\mu(a) = g(a)d\mathcal{L}(a) + m_0\delta_0$ where $\mathcal{L}$ is the Lebesgue measure on $(\underline{a}, \infty)$ and g is a function that is integrable with respect to this Lebesgue measure; δ_0 is the Dirac delta function at $a = 0$ and the scalar m_0 is the Dirac mass at $a = 0$.[59] The analogue of (107) is then

$$0 = \int_{\underline{a}}^{\infty} \varphi'(a)s(a)g(a)d\mathcal{L}(a) + \varphi'(0)s(0)m_0.$$

E.5 Generalization to KF equations in Paper and Beyond

This solution concept generalizes in a straightforward fashion to the KF equations used in the paper. For simplicity consider only the case with a Dirac mass at the boundary, the other case being analogous. First consider the stationary KF equation in the Huggett model with two income types in Section 1. The statement "$g_j, j = 1, 2$ is a solution to (8)" means: for any test functions (φ_1, φ_2) defined on $[\underline{a}, \infty)$ and potentially not vanishing at $\underline{a}$, the functions (g_1, g_2) defined on $(0, \infty)$ and the scalars (m_1, m_2) satisfy

$$\begin{aligned} 0 = &\int_{\underline{a}}^{\infty} \left[\varphi_1'(a)s_1(a) + \lambda_1(\varphi_2(a) - \varphi_1(a))\right] g_1(a)d\mathcal{L}(a) + [\varphi_1'(\underline{a})s_1(\underline{a}) + \lambda_1(\varphi_2(\underline{a})) - \varphi_1(\underline{a})]m_1 \\ &+ \int_{\underline{a}}^{\infty} \left[\varphi_2'(a)s_2(a) + \lambda_2(\varphi_1(a) - \varphi_2(a))\right] g_2(a)d\mathcal{L}(a) + [\varphi_2'(\underline{a})s_2(\underline{a}) + \lambda_2(\varphi_1(\underline{a})) - \varphi_2(\underline{a})]m_2. \end{aligned}$$

Next consider the stationary KF equation in the Huggett model with a continuum of income types in Appendix G.1. The statement "g is a solution to (116)" means: for any test function φ defined on $[\underline{a}, \infty) \times (\underline{y}, \bar{y})$ and potentially not vanishing at $\underline{a}$, the non-negative real-valued function g defined on $(\underline{a}, \infty) \times (\underline{y}, \bar{y})$ and the function m defined on $(\underline{y}, \bar{y})$ satisfy

$$\begin{aligned} 0 = &\int_{\Omega} \left[\partial_a\varphi(a, y)s(a, y) + \partial_y\varphi(a, y)\mu(y) + \frac{1}{2}\partial_{yy}v(a, y)\sigma^2(y)\right] g(a, y)d\mathcal{L}(a, y) \\ &+ \int_{\underline{y}}^{\bar{y}} \left[\partial_a\varphi(\underline{a}, y)s(\underline{a}, y) + \partial_y\varphi(\underline{a}, y)\mu(y) + \frac{1}{2}\partial_{yy}\varphi(\underline{a}, y)\sigma^2(y)\right] m(y)d\mathcal{L}(y), \end{aligned}$$

where $\Omega := (\underline{a}, \infty) \times (\underline{y}, \bar{y})$ is the state space. The solutions to the time-dependent KF equations are, of course, defined in the analogous fashion.

[59] Note that the Lebesgue measure does not see the single point $a = 0$.

Note that both of these definitions are special cases of a more general definition. Suppose we have N state variables, $x \in \mathbb{R}^N$. Consider an open subset $\Omega \subset \mathbb{R}^N$, denote its closed counterpart by $\bar{\Omega}$ and its boundary by $\partial\Omega$. Assume that there is a state constraint $X_t \in \bar{\Omega}$. Denote the infinitesimal generator that governs the evolution of X_t by $\mathcal{A}$, its adjoint by $\mathcal{A}^*$. Then the time-dependent and stationary KF equations are

$$\partial_t g = \mathcal{A}^* g \text{ with } g(\cdot, t) = g_0 \qquad \text{and} \qquad 0 = \mathcal{A}^* g.$$

Assume further that the measure μ admits a density for all $x \in \Omega$ but there may be a Dirac mass on the boundary $\partial\Omega$. Then "g satisfies the stationary KF equation $0 = \mathcal{A}^* g$" means that there are functions g defined on Ω and m defined on $\partial\Omega$ such that for any test function φ defined on $\bar{\Omega}$

$$0 = \int_\Omega [\mathcal{A}\varphi(x)]\, g(x) d\mathcal{L}(x) + \int_{\partial\Omega} [\mathcal{A}\varphi(x)]\, m(x) d\mathcal{L}_{\partial\Omega}(x),$$

where $\mathcal{L}_\Omega$ is the N-dimensional Lebesgue measure on Ω and $\mathcal{L}_{\partial\Omega}(x)$ is the Lebesgue measure on the boundary of Ω. The definition is again analogous for the time-dependent KF equation.

F Accuracy of Finite Difference Schemes

This appendix considers various ways of assessing the accuracy of numerical solutions to our continuous-time heterogeneous agent models computed using FD methods discussed in Section 4. Appendix F.1 considers the numerical solution of HJB equations and makes a comparison with discrete-time methods. Appendix F.2 considers the KF equation.

F.1 Accuracy of FD Scheme for Hamilton-Jacobi-Bellman Equation

We here conduct three exercises. Appendix F.1.1 provides the details for the exercise in Section 4.6 in which we compared the computational speed of continuous- and discrete-time methods for given accuracy. Appendix F.1.2 explains why standard discrete-time accuracy metrics like Euler equation errors are not applicable for HJB equations and discusses other candidate accuracy metrics from the mathematics literature on HJB equations. Finally, in Appendix F.1.3, we take advantage of the closed-form solution for a special case without income risk from Section 3.2 and use it as a benchmark to which to compare numerical solutions.

F.1.1 Details for Section 4.6 – Computational Speed and Accuracy: Continuous vs Discrete Time

We compared the computational performance of a continuous- versus a discrete-time formulation of the income fluctuation problem with a borrowing constraint in partial equilibrium, i.e. with an exogenously given interest rate. Section 4.6 summarizes the results. This Appendix describes in more detail the two variants of the income fluctuation problem, their parameterization, our computational methods for solving them as well as the construction of Figure 9 which compares the speed-accuracy tradeoffs between the two methods.

Income fluctuation problem in continuous time. We consider an income fluctuation with a continuum of income types as in Appendix G.1. The corresponding HJB equation is given by (115) and we choose an Ornstein-Uhlenbeck (OU) process for the logarithm of income:

$$d \log y_t = -\theta \log y_t dt + \sigma dW_t, \tag{108}$$

that is the special case of (114) with $\mu(y) = (-\theta \log(y) + \sigma^2/2)y$ and $\sigma(y) = \sigma y$ (from Ito's formula). The stationary distribution of log income is then a normal distribution with mean zero and variance $\nu^2 = \sigma^2/(2\theta)$. The autocorrelation of log income over a time period of one year is given by $\varrho := \text{Corr}(\log y_{t+1}, \log y_t) = e^{-\theta}$.

To compute our second accuracy metric, we also need to compute stationary aggregate consumption which requires finding the stationary joint distribution of income and wealth. We solve both the HJB equation and the KF equation for the stationary distribution using the finite difference method described in the main text.

Income fluctuation problem in discrete time. The individual solves

$$\begin{aligned}
&\max_{\{a_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t) \quad \text{s.t.} \\
&a_{t+1} = y_t + (1+r)a_t - c_t, \\
&\log y_{t+1} = \varrho \log y_t + \sigma \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1), \\
&a_{t+1} \geq \underline{a}.
\end{aligned}$$

where $\beta \in (0,1)$ is a discount factor and all other variables have the same interpretation as in the continuous-time problem. Note that the income process is an AR(1) process for the logarithm of income. Our continuous-time process, the OU process (108), is the continuous-time analogue of this process. Analogous to the continuous-time variant, the stationary distribution of log income is a normal distribution with mean zero and variance $\nu^2 = \sigma^2/(1-\rho^2)$.

To find the optimal consumption and saving policy functions of this discrete-time income fluctuation problem we solve the corresponding Euler equation using the endogenous grid method (Carroll, 2006). As already noted, to compute our second accuracy metric, we also need to compute stationary aggregate consumption which requires finding the stationary joint distribution of income and wealth. We use a Monte-Carlo method to find this stationary distribution: using the optimal saving policy function and the income process we simulate 50,000 individuals over 500 time periods.

Parameterization. For both variants, we assume CRRA utility (5) with $\gamma = 2$ and impose a strict no-borrowing limit $\underline{a} = 0$. We use a discount factor of $\beta = 0.95$ for the discrete-time variant and a discount rate of $\rho = 1/\beta - 1 = 0.0526$ for the continuous-time one, and set the interest rate to $r = 0.03$ in both cases.

Next consider the income process. We set the yearly autocorrelation of income to $\varrho = 0.95$. For the OU process (108) we therefore set $\theta = -\ln(\varrho) = 0.0513$. We set the standard deviation of the stationary income distribution to $\nu = 0.2$ which implies $\sigma = \nu\sqrt{2\theta} = 0.064$ for the OU process (108) (from $\nu^2 = \sigma^2/(2\theta)$) and $\sigma = \nu\sqrt{1-\varrho^2} = 0.062$ for the AR(1) process (from $\nu^2 = \sigma^2/(1-\varrho^2)$). In both cases, we normalize the process such that the stationary mean equals one. In both cases, we discretize the income process on a grid with 9 points. In discrete time we use the Rouwenhorst (1995) method; in continuous time we use our finite difference method to convert the OU process (108) into a discrete-state Poisson process (see the main text and computational appendix).

In both cases, we use a non-uniform grid for wealth: $a_i = \underline{a} + (a_{\max} - \underline{a})x_i^{\chi}$ with $\chi = 2$ where $x_i, i = 1, ..., I$ is a uniformly spaced grid on the interval $[0, 1]$. The lower bound is $\underline{a} = 0$ and the upper bound is $a_{\max} = 50$.

In both cases, we use the following convergence criterion for the value/consumption policy function: that the difference between the function in the current iteration and the previous one as measured by the sup-norm is below a tolerance of 10^{-7}, e.g. $||v^n - v^{n-1}|| \leq 10^{-7}$.

Finally, a brief comment is in order: while we have specified and parameterized the continuous-time and discrete-time versions of our test problem to be as comparable as possible, these are still different models. For example, while the OU process (108) is the continuous-time analogue of the discrete-time AR(1) process and we parameterize the two processes in an analogous fashion, these are still different stochastic processes: one process moves continuously over time whereas the other only moves at discrete time intervals. As a result, there are small differences in model outputs even putting aside any numerical error, and similarly for the "true" solutions computed with an extremely fine grid with 10,000 wealth grid points. For example, "true" aggregate consumption equals 1.0259 in the continuous-time version and 1.0475 in the discrete-time version (recall that we normalized stationary mean income to one in both cases).

Table 1: Computational Speed and Accuracy: Continuous vs Discrete Time

(a) Policy Function

	Continuous time (finite difference method)		Discrete time (endogenous grid method)	
Grid points (1)	Speed in seconds (2)	Policy function error (%) (3)	Speed in seconds (4)	Policy function error (%) (5)
10	0.04	5.29	2.88	4.47
25	0.08	2.14	2.90	1.29
50	0.10	1.07	2.93	0.29
100	0.12	0.53	3.24	0.08
1,000	0.65	0.05	14.75	0.00
10,000	7.92	0	428.96	-

(b) Aggregate Consumption

	Continuous time (finite difference method)		Discrete time (endogenous grid method)	
Grid points (6)	Speed in seconds (7)	Error in aggregate consumption (%) (8)	Speed in seconds (9)	Error in aggregate consumption (%) (10)
10	0.05	0.12	28.64	58.76
25	0.08	0.07	29.41	7.41
50	0.10	0.04	30.36	2.03
100	0.13	0.02	32.13	0.57
1,000	0.68	0.01	49.16	0.01
10,000	8.18	0	462.21	-

Notes: The table reports speed and accuracy measures for the numerical solution of an income fluctuation problem in both continuous and discrete time. The code is available at https://benjaminmoll.com/comparison/.

Accuracy metrics and construction of Figure 9. As discussed in the main text, we assess accuracy using one of two metrics: (i) the mean percentage error in the policy function relative to its counterpart computed using an extremely fine grid, and (ii) the deviation of stationary aggregate consumption from its counterpart computed using an extremely fine grid. More precisely, the two metrics we use in Figure 9 are given by

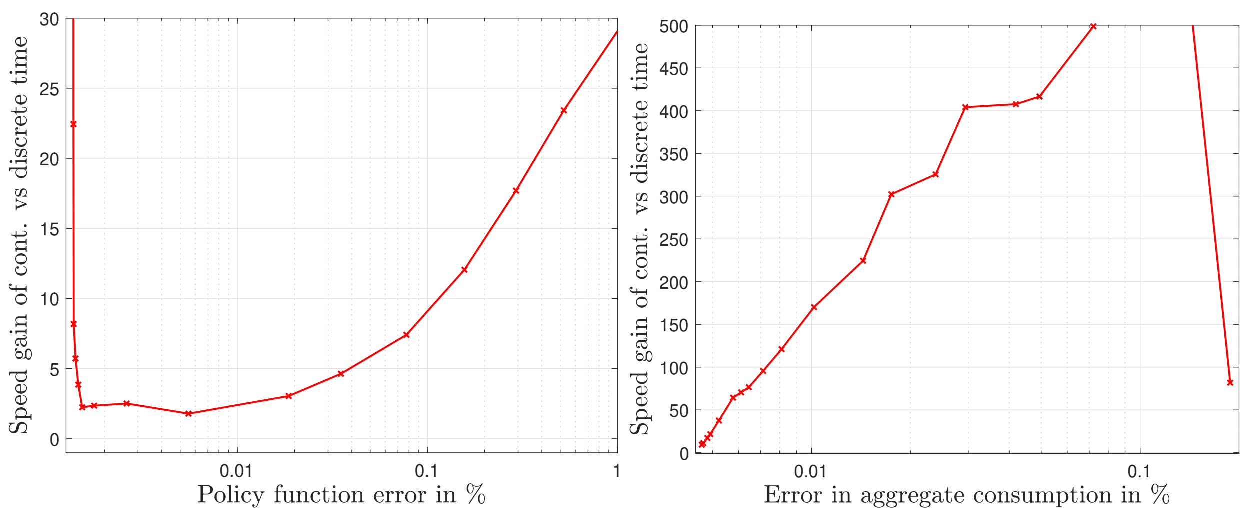
$$\text{error}_1 = 100 \times \frac{1}{I^*}\frac{1}{J^*}\sum_{i=1}^{I^*}\sum_{j=1}^{J^*}\frac{|\tilde{c}(a_i,y_j) - c^*(a_i,y_j)|}{c^*(a_i,y_j)}, \qquad \text{error}_2 = 100 \times \frac{C - C^*}{C^*}, \tag{109}$$

where c^* is the consumption policy function computed using the extremely fine grid with I^* and J^* grid points, $\tilde{c}$ is the consumption policy function computed using a coarser grid and then interpolated onto the finer grid, and C and C^* are stationary aggregate consumption computed using the coarse and fine grid.

To construct the speed-accuracy tradeoffs in Figure 9 we first compute the solutions to the two test problems but with different numbers of grid points. Table 1 contains output from a subset of the computations. First consider panel (a) of Table 9 which is the input used to construct panel (a) of Figure 9. Column (1) shows the number of grid points used in each experiment ranging from a very coarse discretization with 25 grid points to an extremely fine one with 10,000 grid points (the “true” solution). Columns (2)-(3) report results for the continuous-time FD method whereas columns (4)-(5) report those for the discrete-time endogenous-grid method. Columns (2) and (4) report the time until the algorithm converged measured in seconds. Columns (3) and (5) report the first accuracy metric in (109). Unsurprisingly, in both continuous and discrete time, the computations become slower but more accurate as the number of grid points grows. Next consider panel (b) of Table 9 which is the input used to construct panel (b) of Figure 9. All columns are exactly analogous to those in panel (a) with two differences: first, the computational speed in columns (7) and (9) now includes the time to compute the stationary distribution; second, columns (8) and (10) now report the second accuracy metric in (109).

With the input from Table 1 in hand, the construction of Figure 9 is straightforward: in panel (a) of the figure, the blue line with circles is column (2) in the table plotted against column (3), the red line with crosses in the figure is column (4) plotted against column (5); similarly, in panel (b), the blue line with circles is column (7) plotted against column (8), the red line with crosses is column (9) plotted against column (10).

Finally, Figure 11 reports the ratios of the two computational times for given accuracy metrics in Figure 9.



(a) Policy function

(b) Aggregate consumption

Figure 11: Performance of continuous-time method relative to discrete-time method

Notes: in each panel, the speed gain is computed from Figure 9 as the ratio of the discrete-time speed to the continuous-time speed, i.e. the ratio of the red line with crosses to the blue line with circles.

Code and Computational Infrastructure. The code is available at https://benjaminmoll.com/comparison/. There are two files to generate the "true" values: they run the code for the superfine grid disc_true.m and cont_true.m and write the benchmark duration and aggregate consumption to a txt/csv files disc_true.txt and cont_true.txt. Then there is a main file (main.m), which runs the code first for the continuous-time variant and then for the discrete-time one. For each variant it first checks to see whether the .txt file containing the "true" value exists in the path/current folder or not. If it does, it reads the "true" aggregate consumption and speed from it and loops over different grid sizes. Otherwise, it first calls the benchmarking file and then loops over coarser grids. Finally fig_tradeoff.m generates Figures 9 and 11.

The computations which the results in Figure 9 are based on were performed on a 2018 MacBook Pro with a 2.9GHz Intel Core i9 CPU.

Additional Exercise. We also compared the numerical solution of the continuous-time to a simpler discrete-time problem with an iid income process rather than a persistent AR(1) process as in the exercise above. The discrete-time income fluctuation problem can then be solved more efficiently because iid income shocks imply that one can work with one state variable only, namely cash-on-hand. That is, the discrete-time algorithm with an iid income process has a head start relative to the continuous time one. Nevertheless, results analogous to those in Figure 9 (not reported here) show that, for any given number of wealth grid points, our continuous-time code for solving the problem with a persistent income process runs about 10 times faster than the discrete-time code for solving the problem with the iid process while at the same time being more accurate.

F.1.2 Applicability of Euler-equation errors and continuous-time accuracy metrics

In the exercise in Section 4.6, we chose a pragmatic approach for assessing the accuracy of our finite-difference method but one without a deep theoretical foundation. Why did we not develop the continuous-time analogue of a standard accuracy metric from the discrete-time literature like the Euler equation error? Or alternatively take an existing accuracy metric from the continuous-time literature?

For discrete-time problems, the standard strategy for assessing the accuracy of a numerical solution to a Bellman equation is to examine the associated Euler equation errors. It turns out that the rationale for examining these does not apply to HJB equations. To see this consider the analogous discrete-time problem in Section 4.1 with Euler equation (37). As explained by Santos (2000), the rationale for examining the residuals in this Euler equation is that it is the first-order condition of the maximization problem in the Bellman equation. And by bounding the error in this first-order condition, one can bound the error in the policy function and more importantly the value function, i.e. the welfare loss from suboptimal behavior due to numerical error.[60] But for HJB equations like (41) and the associated finite-difference approximation (42), there is no error in the first-order condition (36) because it can be solved by hand. Instead, any error in the numerical solution of this PDE stems only from the finite-difference approximation of its derivatives.

To assess this error, the mathematics literature contains a number of results on the accuracy of numerical solutions to HJB equations. Closest in spirit to Euler-equation errors are so-called "a posteriori error estimates" (see e.g. Albert, Cockburn, French, and Peterson, 2002; Cockburn, Merev, and Qian, 2013). In particular, they share three features with the latter: (i) they need to be calculated after the numerical solution has been computed (hence the "a posteriori" name), (ii) they depend solely on the approximate solution, and (iii) and they do not depend on the particular numerical method used to compute the approximate solution.[61] Unfortunately, however, the mathematics literature has not derived any such a posteriori estimates for problems with state constraints like ours.

The literature also contains another set of results on the accuracy of FD approximations to HJB equations, so-called "a priori error estimates." Most of these are variants of a classic result by Crandall and Lions (1984) which states that the distance between the finite-difference approximation to an HJB equation and its viscosity solution scales with the square root of the grid step in the approximation. For example, in the context of the value function v solving the HJB equation (41) and its numerical approximation v_i solving (42), we have $|v_i - v(a_i)| \leq \kappa\sqrt{\Delta a}$ for all $i = 1, ..., I$ where κ is a constant that depends on parameters.[62] While this may seem like a useful result for judging accuracy, typically no characterization is available for the constant κ. This means that these a priori estimates typically cannot be used to judge the accuracy of a numerical solution (because κ may be very large); instead, their main use is a characterization of the rate at which the error converges to zero as the grid is refined more and more, i.e. as $\Delta a \to 0$. The introduction of Albert, Cockburn, French, and Peterson (2002) contains an accessible discussion of the differences between a

[60] As Santos (2000, Lemma 2.1) explains, the result is a generalization of a simple result about optimization problems (here stated in one dimension). Consider a twice-differentiable and sufficiently concave function $f(x)$ and denote by x^* its interior optimum. If a candidate optimum x satisfies the first-order condition up to an error $\varepsilon > 0$, $|f'(x)| \leq \varepsilon$, then $|x - x^*| \leq \varepsilon/\eta$ and $|f(x) - f(x^*)| \leq \varepsilon^2/\eta$ for a constant $\eta > 0$.

[61] The literature further distinguishes between "global" and "local" a posteriori estimates. Global error estimates bound the maximum error over the entire state space (typically the L^∞-norm between the numerical approximation and true solution computed over the entire state space) whereas local error estimates bound the error in particular regions of the state space. An advantage of local estimates is therefore to give information on where the grid should be refined in order to improve accuracy.

[62] For readers consulting their original paper, the following remark may be helpful. Crandall and Lions are mostly concerned with time-dependent problems so that the discretization additionally features a time step Δt. Their main result (their Theorem 1 and in particular equation (5)) is obtained under the assumption that the grid steps are proportional to the time step Δt (see their assumption that λ^x and λ^y are fixed). On p.15 they state that the result can be extended to stationary problems and this has been confirmed by subsequent work.

priori and a posteriori error estimates.

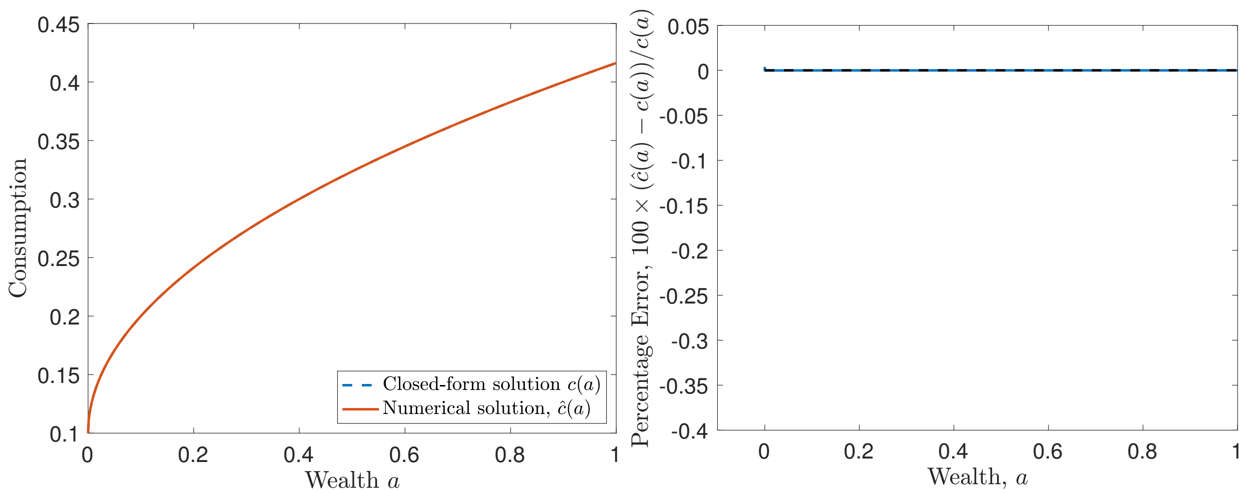
Given the inapplicability of Euler equation errors and the absence of off-the-shelf continuous-time accuracy metrics, we have therefore opted to assess accuracy pragmatically using the two metrics discussed above: (i) the mean percentage error in the policy function relative to its counterpart computed using an extremely fine grid, and (ii) the deviation of stationary aggregate consumption from its counterpart computed using an extremely fine grid.

F.1.3 Accuracy of FD Scheme: Comparison with Closed-Form Solution

In Section 3.2 ("Intuition for Proposition 1 and Corollary 1: Two Useful Special Cases") we derived a closed-form solution for individuals' consumption policy function under the assumption that utility is exponential (6), that there is no income risk and that the interest rate $r = 0$, namely $c(a) = y + \sqrt{2\nu a}$ where $\nu := \rho/\theta$. The HJB equation corresponding to the same problem is

$$\rho v(a) = \max_c \; -e^{-\theta c}/\theta + v'(a)(y - c),$$

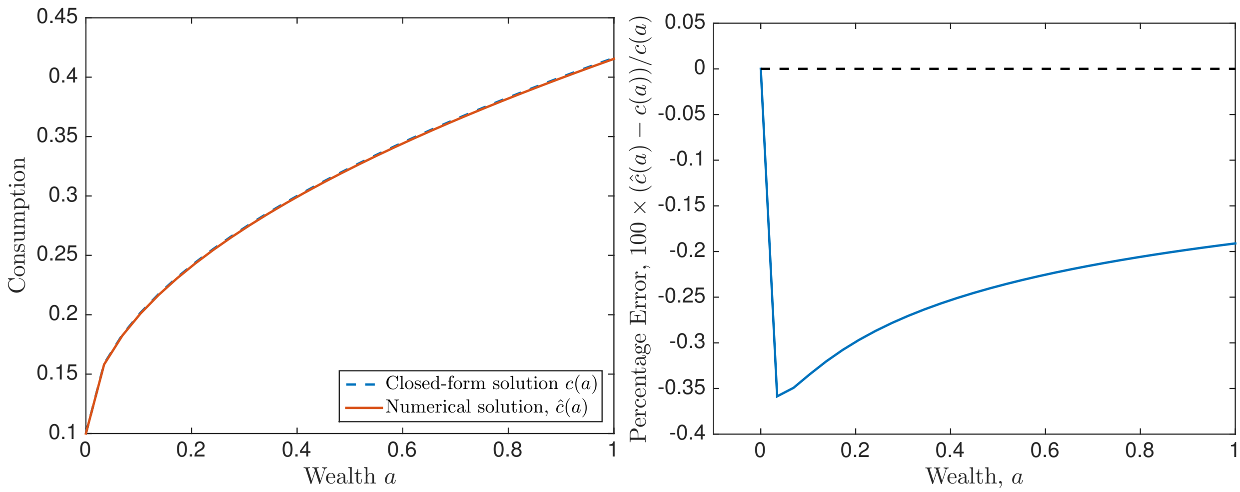
with a state constraint $a \geq 0$. We numerically solve this HJB equation using the FD method laid out in Section 4. The code is available at http://benjaminmoll.com/HJB_accuracy1/. We denote the consumption policy function computed in this fashion by $\widehat{c}(a)$. Figures 12 and 13 compare the numerically computed consumption policy function $\widehat{c}(a)$ to the analytic solution $c(a)$. Figure 12 does this for a fine grid with $I = 1000$ wealth grid points. Panel (a) shows that the two consumption policy functions are virtually indistinguishable. Panel (b) computes the percentage difference between the true and numerical solutions and shows that the error is extremely small, on the order of 0.01%. Figure 13 repeats the exercise for a



(a) Consumption Policy Function (b) Percentage Error in Policy Function

Figure 12: Approximation error in consumption policy function from FD method, fine grid $I = 1000$ wealth grid points

much coarser grid with $I = 30$ grid points. The numerical solution is still relatively accurate with a somewhat larger approximation error of up to -0.4%.



(a) Consumption Policy Function

(b) Percentage Error in Policy Function

Figure 13: Approximation error in consumption policy function from FD method, coarse grid $I = 30$ wealth grid points

F.2 Accuracy of FD Scheme for Kolmogorov Forward Equation

Readers may worry that the existence of the Dirac mass at the borrowing constraint (see Proposition 3) may cause problems because our FD scheme explained in Section 4.4 does not explicitly take into account its existence. This Appendix shows that this is not a valid concern. We proceed first theoretically and then numerically.

F.2.1 Finite Differences and Dirac Masses: Theoretical Considerations

Appendix E explained how to think about solutions to the KF equation when there is a Dirac point mass at the boundary by introducing an appropriate notion of weak solution. We now show that the numerical scheme in Section 4.4 is consistent with this notion as long as some care is taken when *interpreting* the output of the numerical algorithm. For simplicity we make the argument in the context of the simplified model without income risk already studied in Section E. Recall from there that we denoted the wealth density for $a > \underline{a}$ by $g(a)$ and the Dirac point mass at $a = \underline{a}$ by m.

Discretization: As in Section 4.4, we discretize the distribution g as $g_i, i = 0, ..., I$ on an equi-spaced grid $a_i, i = 0, ..., I$ with step size Δa. Our claim is that this discretization is consistent with the continuous problem if: (a) we view $g_i, i > 0$ as the discrete counterpart to the density g, and (b) we view $g_0 \Delta a$ as the discrete counterpart to the Dirac mass m.

Intuition: The simplest way of getting intuition for this interpretation is to consider the normalization conditions in the continuous and discrete approximation side by side:

$$1 = m + \int_{\underline{a}}^{\infty} g(a) d\mathcal{L}(a),$$
$$1 = g_0 \Delta a + \sum_{i>0} g_i \Delta a.$$

Clearly, the two equations are consistent if we take $m \approx g_0 \Delta a$ and $g(a_i) d\mathcal{L}(a_i) \approx g_i \Delta a,\ i > 0$. More generally, for any test function φ, we approximate $\mathbb{E}[\varphi(a_t)] = \int_{\underline{a}}^{\infty} \varphi(a) d\mu(a) \approx \sum_{i=0}^{I} \varphi_i g_i \Delta a$, where $\varphi_i := \varphi(a_i)$. From (106) this is equivalent to

$$\varphi(\underline{a}) m + \int_{\underline{a}}^{\infty} \varphi(a) g(a) d\mathcal{L}(a) \approx \varphi_0 g_0 \Delta a + \sum_{i>0} \varphi_i g_i \Delta a,$$

and this again yields the same conclusion.

It is again important to emphasize that the only thing that is at stake in this discussion is the *interpretation* of the discretized distribution $g_i, i \geq 0$. In particular none of it affects how we calculate macroeconomic aggregates and other moments of the distribution. Such moments are always approximated as $\mathbb{E}[\varphi(a_t)] \approx \sum_{i=0}^{I} \varphi_i g_i \Delta a$ which is the right thing to do independent of whether there is a Dirac point mass at the boundary.

More Systematic Approach via Discrete KF Equation: A more systematic approach is to make the connection between the numerical scheme laid out in Section 4.4 and the discrete counterpart of the weak formulation of the KF equation (104). When discretizing the term $\varphi'(a)s(a)$ we upwind it as explained in Section 4.3

$$\varphi'(a_i) s(a_i) \approx \frac{\varphi_{i+1} - \varphi_i}{\Delta a} s_i^+ + \frac{\varphi_i - \varphi_{i-1}}{\Delta a} s_i^-,$$

where $s_i := s(a_i)$. The discrete counterpart to (104) is then

$$0 = \sum_{i \geq 0} \left(\frac{\varphi_{i+1} - \varphi_i}{\Delta a} s_i^+ + \frac{\varphi_i - \varphi_{i-1}}{\Delta a} s_i^- \right) g_i \Delta a. \tag{110}$$

Performing a discrete integration by parts, we have

$$0 = \sum_{i \geq 0} \left(- \frac{s_i^+ g_i - s_{i-1}^+ g_{i-1}}{\Delta a} - \frac{s_{i+1}^- g_{i+1} - s_i^- g_i}{\Delta a} \right) \varphi_i \Delta a,$$

which, in turn, yields

$$0 = -\frac{s_i^+ g_i - s_{i-1}^+ g_{i-1}}{\Delta a} - \frac{s_{i+1}^- g_{i+1} - s_i^- g_i}{\Delta a}, \qquad i = 1, ..., I. \tag{111}$$

Note that this is exactly the discretization of the KF equation advocated in Section 4.4. In particular, note that (111) can be written in matrix form as $0 = \mathbf{A}^{\mathrm{T}}\mathbf{g}$ with $\mathbf{g} = (g_0, ..., g_I)^{\mathrm{T}}$ and where the matrix $\mathbf{A}$ has the same form as in Section 4.3.

To see if the numerical scheme is consistent with the continuous problem, we only have to check if the weak formulation of the discretized KF equation (110) is consistent with the weak formulation of the continuous KF equation (107). The two problems are again consistent if we interpret $g_i, i > 0$ as the discrete version of g, and $g_0\Delta a$ as the discrete version of m.

FD Scheme with Dirac Mass in Interior: Finally, consider the approximation of the density when there is a Dirac mass at $a = 0$ in the interior of the state space $[\underline{a}, \infty)$ as with a soft borrowing constraint (Section G.3) and as briefly discussed in Section E.4 (Application 2). Consider a grid that places a point at $a = 0$, e.g. $a_i, i = 0, ..., I$ with $a_k = 0$ for some $k > 0$. Denoting by $\mathcal{L}$ the Lebesgue measure on $(\underline{a}, \infty)$, the continuous and discrete normalization conditions are

$$1 = m_0 + \int_{\underline{a}}^{\infty} g(a) d\mathcal{L}(a),$$
$$1 = g_k \Delta a + \sum_{i \neq k} g_i \Delta a.$$

which suggests that the correct interpretation is to view $g_k\Delta a$ as the discrete counterpart of the Dirac mass m_0 at $a = 0$ and $g_i, i \neq k$ as the counterpart of the density everywhere else. This can again be made more rigorous by following the same steps as above.

Points at which there is both positive density and a Dirac mass: In some of our applications, at some points a there is a Dirac mass for some income types and positive but finite density for other types. For example in Figure 17 (b), there is a Dirac mass for income type 1 but positive, finite density for types 0 and 2. The discussion thus far implies that, for a fixed grid, our numerical scheme cannot distinguish between the two: at that point a_k, $g_k\Delta a$ will be positive in both scenarios. However, we can distinguish between the two by varying the grid spacing. In particular, a Dirac mass implies that $g_k \approx m/(\Delta a)$. Hence we can conclude that there is a Dirac mass at a point when, as $\Delta a \to 0$, g_k scales like $1/(\Delta a)$. If instead g_k converges to a positive constant as $\Delta a \to 0$ then there is no Dirac mass.

F.2.2 Numerical Experiments

In addition to these theoretical considerations, we take advantage of our closed-form solution for the stationary wealth distribution from Proposition 3 to assess the accuracy of the FD scheme for the KF equation in practice. We show that, in practice, the numerical solution closely approximates the analytic solution.

Our closed form for the distribution in (33) involves the term $-\int_{\underline{a}}^{a}\left(\frac{\lambda_1}{s_1(x)}+\frac{\lambda_2}{s_2(x)}\right)dx$ which is difficult to evaluate numerically for the optimal saving policy functions because these satisfy $s_1(\underline{a}) = s_2(a_{\max}) = 0$. Rather than evaluating (33) at the optimal saving policy functions, we therefore make use of our expansions of these policy functions around the points $\underline{a}$ and $a_{\max}$. That is, we here compute the KF equation for the case when this characterization is exact for all a and assume that

$$s_1(a) = -\sqrt{2\nu_1}\sqrt{a-\underline{a}}, \qquad s_2(a) = -\zeta_2(a - a_{\max}) \tag{112}$$

for all a (and not just at $a = \underline{a}$ and $a = a_{\max}$). Under this assumption, we have

$$\begin{aligned}
-\int_{\underline{a}}^{a}\left(\frac{\lambda_1}{s_1(x)}+\frac{\lambda_2}{s_2(x)}\right)dx &= -\int_{\underline{a}}^{a}\left(\frac{\lambda_1}{-\sqrt{2\nu_1}\sqrt{x-\underline{a}}}+\frac{\lambda_2}{-\zeta_2(x-a_{\max})}\right)dx \\
&= \frac{\lambda_1}{\sqrt{\nu_1/2}}\sqrt{a-\underline{a}} + \frac{\lambda_2}{\zeta_2}\Big(\log(a_{\max}-a) - \log(a_{\max}-\underline{a})\Big)
\end{aligned}$$

Therefore (33) and the Dirac point mass m_1 defined in Proposition 3 become

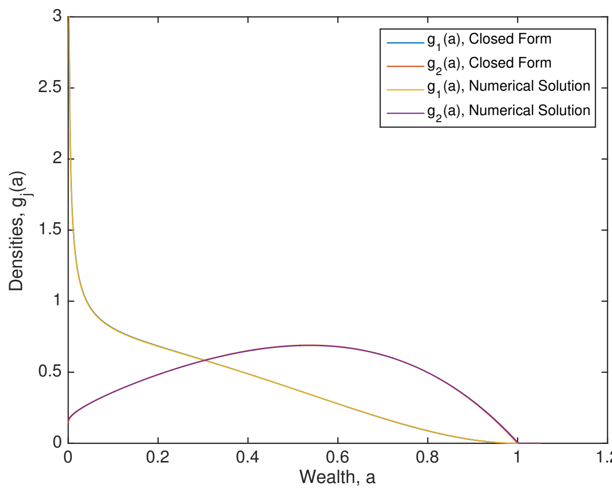
$$\begin{aligned}
g_1(a) &= \frac{\kappa_1}{-\sqrt{2\nu_1}\sqrt{a-\underline{a}}}\exp\left(\frac{\lambda_1}{\sqrt{\nu_1/2}}\sqrt{a-\underline{a}}\right)\left(\frac{a_{\max}-a}{a_{\max}-\underline{a}}\right)^{\lambda_2/\zeta_2} \\
g_2(a) &= \frac{\kappa_2}{-\zeta_2(a-a_{\max})}\exp\left(\frac{\lambda_1}{\sqrt{\nu_1/2}}\sqrt{a-\underline{a}}\right)\left(\frac{a_{\max}-a}{a_{\max}-\underline{a}}\right)^{\lambda_2/\zeta_2}
\end{aligned} \tag{113}$$

$$m_1 = \frac{\lambda_2}{\lambda_1+\lambda_2}\tilde{m}_1, \quad \frac{1}{\tilde{m}_1} = \frac{\lambda_2}{\zeta_2}(a_{\max}-\underline{a})^{-\lambda_2/\zeta_2}\int_{\underline{a}}^{a_{\max}}\exp\left(\frac{\lambda_1}{\sqrt{\nu_1/2}}\sqrt{a-\underline{a}}\right)(a_{\max}-a)^{\lambda_2/\zeta_2-1}da.$$

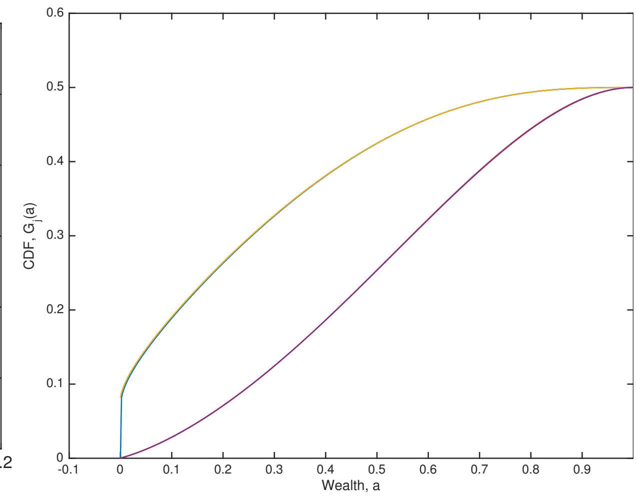
Finally, as explained in Appendix A.4.4, we have $\kappa_1 = -\lambda_1 m_1$ and $\kappa_2 = \lambda_2 m_2$.

Figure 14 plots this solution and compares it to the numerical solution computed using the algorithm laid out in Section 4.4 with $I = 500$ wealth grid points (of course, also assuming that s_1 and s_2 are given by (112)). We here assume that $\underline{a} = 0, a_{\max} = 1, \nu_1 = 0.05, \zeta_2 = 0.25, \lambda_1 = 0.5, \lambda_2 = 0.5$. With $I = 500$ grid points, the results are extremely similar for all other parameter combinations we have tried. Panel (a) plots the densities and panel (b) plots the corresponding cumulative distribution functions (CDFs).[63]

[63] To obtain the CDFs corresponding to the closed-form solution (113), we numerically integrate g_1 and g_2 in (113). In this regard, a difficulty is that $g_1(a) \to \infty$ as $a \to \underline{a}$. We therefore compute the cumulative

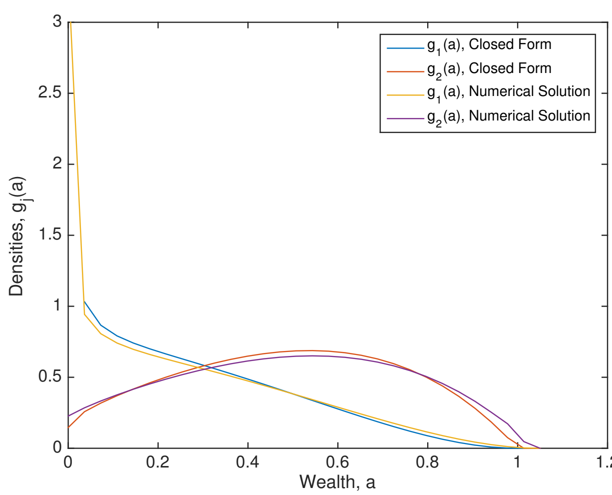


(a) Densities

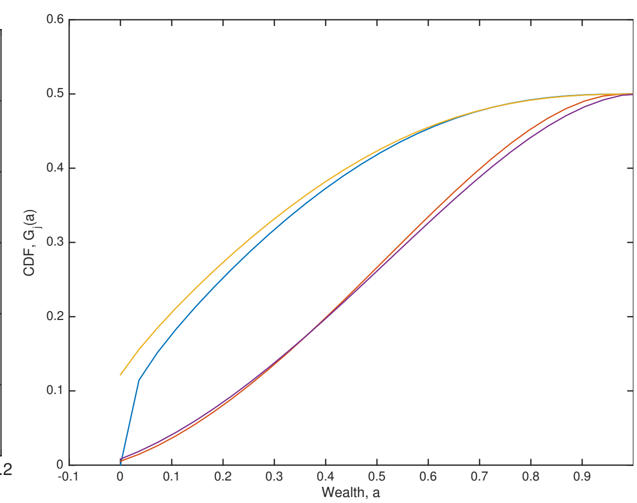


(b) Cumulative Distribution

Figure 14: A fine grid with $I = 500$ points results in the FD scheme being highly accurate



(a) Densities



(b) Cumulative Distribution

Figure 15: A coarse grid with only $I = 30$ points results in the FD scheme being relatively inaccurate

Figure 15 repeats the exercise but with only $I = 30$ wealth grid points. With this much lower number of grid points the approximation is naturally of lower quality. That being said, the approximation can easily be improved by employing a non-equispaced grid. See the online Appendix at https://benjaminmoll.com/HACT_Numerical_Appendix/ for a discussion on how to do this.

The codes used to generate Figures 14 and 15 are available online at https://benjaminmoll.com/KFE_accuracy_check/. The interested reader can try out herself how varying the number of grid points and parameter values affects the accuracy of the numerical solution.

distribution function for type 1 as $G_1(a) = p_1 - \int_a^{a_{\max}} g_1(a)da$ where $p_1 = \frac{\lambda_2}{\lambda_1+\lambda_2}$ is the total mass of type 1 individuals.

G Details for Section 6: Generalizations and Extensions

G.1 More General Income Processes

Our baseline model assumed that income y_t takes one of two values, high and low. We now extend many of our theoretical results to an environment with a continuum of productivity types. Furthermore, the computational algorithm laid out in Section 4 carries over without change. This is true even though the system of equations describing an equilibrium will be a system of PDEs rather than a system of ODEs.

As in Section 1.1, there is a continuum of individuals that are heterogeneous in their wealth a and income y. The state of the economy is the joint distribution of income and wealth $g(a, y, t)$. The simplest way of introducing a continuum of income types is to work with a continuous diffusion process. Individual income evolves stochastically over time on a bounded interval $[\underline{y}, \bar{y}]$ with $\bar{y} > \underline{y} \geq 0$, according to the stationary diffusion process[64]

$$dy_t = \mu(y_t)dt + \sigma(y_t)dW_t. \tag{114}$$

This is simply the continuous-time analogue of a Markov process (without jumps). W_t is a Wiener process or standard Brownian motion and the functions μ and σ are called the drift and the diffusion of the process. We normalize the process such that its stationary mean equals one. An individual's problem is now to maximize (1) subject to (2), (3) and (114), taking as given the evolution of the interest rate r_t for $t \geq 0$.[65]

Similarly to Section 1, a *stationary* equilibrium can be written as a system of partial differential equations. The problem of individuals and the joint distribution of income and wealth satisfy stationary HJB and KF equations:

$$\rho v(a, y) = \max_c \; u(c) + \partial_a v(a, y)(y + ra - c) + \partial_y v(a, y)\mu(y) + \frac{1}{2}\partial_{yy} v(a, y)\sigma^2(y), \tag{115}$$

$$0 = -\partial_a(s(a, y)g(a, y)) - \partial_y(\mu(y)g(a, y)) + \frac{1}{2}\partial_{yy}(\sigma^2(y)g(a, y)). \tag{116}$$

on $(\underline{a}, \infty) \times (\underline{y}, \bar{y})$. The function s is the saving policy function

$$s(a, y) = y + ra - c(a, y), \quad \text{where} \quad c(a, y) = (u')^{-1}(\partial_a v(a, y)). \tag{117}$$

[64] The process (114) either stays in the interval $[\underline{y}, \bar{y}]$ by itself or is reflected at $\underline{y}$ and $\bar{y}$. From a theoretical perspective there is no need for restricting the process to a bounded interval, and unbounded processes can be easily analyzed. Instead the motivation for this assumption is purely practical: we ultimately solve the problem numerically and any computations necessarily require income to lie in a bounded interval.

[65] The corresponding "natural borrowing constraint" is now $a_t \geq -\underline{y} \int_t^\infty \exp\left(-\int_t^s r_\tau d\tau\right) ds$. As before, the borrowing constraint $\underline{a}$ only binds if it is tighter than this "natural" borrowing limit.

The function v again satisfies a state constraint boundary condition at $a = \underline{a}$ which is now

$$\partial_a v(\underline{a}, y) \geq u'(y + r\underline{a}), \quad \text{all } y. \tag{118}$$

Because the diffusion is reflected at $\underline{y}$ and $\bar{y}$, the value function also satisfies the boundary conditions

$$\partial_y v(a, \underline{y}) = 0, \qquad \partial_y v(a, \bar{y}) = 0, \quad \text{all } a. \tag{119}$$

A stationary equilibrium is a scalar r and functions v and g satisfying the PDEs (115) and (116) with s given by (117), boundary conditions (118), (119), with an equilibrium condition analogous to (11), namely $\int_{\underline{y}}^{\bar{y}} \int_{\underline{a}}^{\infty} a g(a, y) da dy = B$. Transition dynamics again satisfy a system of time-dependent PDEs analogous to that in Section 1.

Importantly, the computational algorithm laid out in Section 4 carries over without change: from a computational perspective it is immaterial whether we solve a system of ODEs like (7) and (8) or a system of PDEs like (115) and (116). This would not be true if we had relied on an pre-built ODE solver (say one that is part of Matlab) to solve the ODEs (7) and (8).

Other income processes are possible as well. For instance, Kaplan, Moll, and Violante (2018) consider a "jump-drift process" with transitory and permanent components. As in (114) there is a continuum of types for each component; but rather than moving continuously over time as in (114), each component is subject to Poisson jumps. Income could also follow a jump-diffusion process. At the most general level, we can accommodate any income process that can be represented with an "infinitesimal generator." To treat the general case, we write the HJB and KF equations as

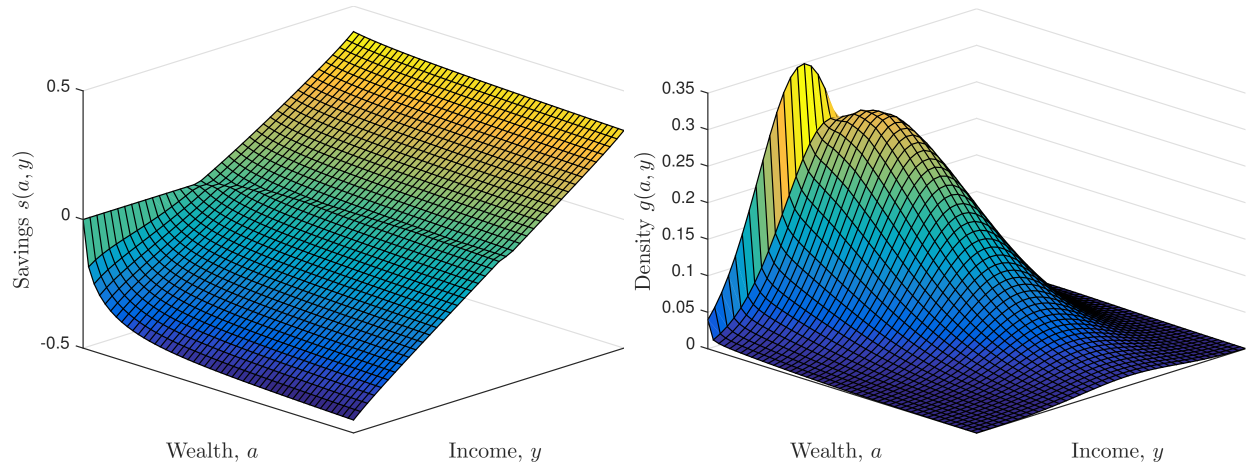
$$\rho v = \max_c \; u(c) + (y + ra - c)\partial_a v + \mathcal{A}v, \tag{120}$$

$$0 = -\partial_a (s(a, y)g) + \mathcal{A}^* g, \tag{121}$$

with a state constraint $a \geq 0$. Here $\mathcal{A}$ is the infinitesimal generator ("infinite-dimensional transition matrix") of the stochastic process for income y_t and $\mathcal{A}^*$ is its adjoint. For instance, if y_t follows a two-state Poisson process as in Section 1, then $(\mathcal{A}v)(a, y_j) = \lambda_j(v(a, y_{-j}) - v(a, y_j))$. Or if y_t is a continuous diffusion like aboves, then $\mathcal{A}v = \mu(y)\partial_y v + \frac{\sigma^2(y)}{2}\partial_{yy} v$.

Figure 16 plots the stationary saving policy function and wealth distribution when income follows a diffusion. Both inherit all important properties of the saving policy function and wealth distribution from the baseline model with two income types from Sections 1 and 3. This is not just a numerical result. Instead Propositions 6 and 7 below generalize Propositions 1 and 2 from the case with a two-state Poisson process to other processes including the diffusion process (114).

Proposition 6 (Generalization of Proposition 1 to Other Income Processes) *Assume*



(a) Saving as Function of Income and Wealth

(b) Distribution of Income and Wealth

Figure 16: Stationary Equilibrium of Huggett Model with Diffusion Process

that $r < \rho$ and that Assumption 1 holds with y_1 replaced by $\underline{y}$. Then the solution to the HJB equation (77) and the corresponding saving policy function have the following properties:

1. *There is a cutoff y^* such that $s(\underline{a}, y) = 0$ for all $y \leq y^*$ but $s(a, y) < 0$ for all $a > \underline{a}, y \leq y^*$. That is, individuals with $y \leq y^*$ and wealth exactly at the borrowing constraint are constrained, whereas those with income $y \leq y^*$ and wealth $a > \underline{a}$ are unconstrained and decumulate assets. Those with income $y > y^*$ are always unconstrained and they accumulate assets even at the constraint $s(\underline{a}, y) > 0$ for all $y > y^*$.*

2. *as $a \to \underline{a}$, the saving and consumption policy function of individuals with income below the threshold, $y \leq y^*$, and the corresponding instantaneous marginal propensity to consume satisfy*

$$\begin{aligned} s(a, y) &\sim -\sqrt{2\nu(y)}\sqrt{a - \underline{a}}, \\ c(a, y) &\sim y + ra + \sqrt{2\nu(y)}\sqrt{a - \underline{a}}, \\ \partial_a c(a, y) &\sim r + \frac{1}{2}\frac{\sqrt{2\nu(y)}}{\sqrt{a - \underline{a}}}, \\ \nu(y) &= \frac{(\rho - r)u'(\underline{c}(y)) - (\mathcal{A}u'(\underline{c}))(y)}{-u''(\underline{c}(y))}, \end{aligned}$$

where $\underline{c}(y) = c(\underline{a}, y)$ is consumption at the borrowing constraint. For the special case *with a diffusion process* (114) *so that* $\mathcal{A}v = \mu(y)\partial_y v + \frac{\sigma^2(y)}{2}\partial_{yy}v$*:*

$$\nu(y) = \frac{(\rho - r)u'(\underline{c}(y)) - \mu(y)\partial_y u'(\underline{c}(y)) - \frac{\sigma^2(y)}{2}\partial_{yy}u'(\underline{c}(y))}{-u''(\underline{c}(y))}$$
$$= (\rho - r)IES(\underline{c}(y))\underline{c}(y) + \left(\mu(y) - \frac{\sigma^2(y)}{2}\mathcal{P}(\underline{c}(y))\right)\underline{c}'(y) + \frac{\sigma^2(y)}{2}\underline{c}''(y)$$

where $\mathcal{P}(c) := -u'''(c)/u''(c)$ *is absolute prudence.*

This implies that for $y \leq y^*$ *the derivatives of* $c(a, y)$ *and* $s(a, y)$ *are unbounded at the borrowing constraint,* $\partial_a c(a, y) \to \infty$ *and* $\partial_a s(a, y) \to -\infty$ *as* $a \to \underline{a}$*. Therefore individuals with wealth* $a > \underline{a}$ *and successive low income draws* $y \leq y^*$ *decumulate wealth and hit the borrowing constraint in finite time at speed governed by* $\nu(y)$ *analogous to Corollary 1.*

This Proposition shows that – as can be seen in panel (a) of the Figure – the saving policy function has an unbounded derivative at $a = \underline{a}$ for income y below some threshold, and that therefore individuals with persistent low income realizations hit the borrowing constraint in finite time. This results in the spike in the wealth distribution at $a = \underline{a}$ in panel (b).[66]

Proposition 7 (Generalization of Proposition 2 to Other Income Processes) *Consider the HJB equation* (77) *with a general income process on* $[\underline{y}, \bar{y}]$ *and the corresponding policy functions. Assume that* $r < \rho$ *and that relative risk aversion* $\gamma(c) := -cu''(c)/u'(c)$ *is bounded above for all* c*.*

1. *Then there exists* $a_{\max} < \infty$ *such that* $s(a, y) < 0$ *for all* $a \geq a_{\max}$ *and all* y*.*

2. *In the special case of CRRA utility* (5) *individual policy functions are asymptotically linear in* a*. As* $a \to \infty$*, they satisfy*

$$s(a, y) \sim \frac{r - \rho}{\gamma}a, \qquad c(a, y) \sim \frac{\rho - (1 - \gamma)r}{\gamma}a, \quad \text{all } y.$$

We do not state the extended Propositions 4 and 5 here because the wording is unchanged. The proofs of Proposition 4 and 5 already covered the case of a general income process.

[66] Given that the figure plots the density $g(a, y)$, some readers may wonder why the spike representing the Dirac mass at $a = \underline{a}$ is finite. The answer is that the figure plots the output of our numerical scheme, $g_{i,j}$ over grids $a_i, i = 0, ..., I$ and $y_j, j = 1, ..., J$. As explained in Appendix F.2 the correct interpretation is that $g_{i,j} \approx g(a_i, y_j)$ for all grid points in the interior $i > 0$. But at the boundary $g_{0,j}\Delta a \approx m(y_j)$ where $m(y)$ is the Dirac mass. In the figure for example, $g_{0,j}$ equals about 0.35 at its highest point. The correct interpretation is: since the computation uses $\Delta a = 0.3$, the corresponding Dirac mass is $g_{0,j}\Delta a = 0.35 \times 0.3 = 0.105$.

Proof of Proposition 6: Analogous to Proposition 1, we start by differentiating (115) with respect to a (envelope condition) and using the FOC $u'(c(a,y)) = \partial_a v(a,y)$ to obtain the "Euler equation"

$$(\rho - r)u'(c) = u''(c)(\partial_a c)s + \mathcal{A}u'(c) \tag{122}$$

The proof of Part 1 follows the same steps as in Proposition 1. For Part 2, use $\partial_a c = r - \partial_a s$ in (122) and rearrange to get

$$(\partial_a s - r)s = \frac{(r-\rho)u'(c) + \mathcal{A}u'(c)}{u''(c)}$$

Now consider low income types $y \le y^*$ for whom $s(\underline{a}, y) = 0$. As $a \to \underline{a}$, we additionally have, $c(a,y) \to \underline{c}(y) := y + r\underline{a} > 0$ and $-u'(c(a,y))/u''(c(a,y)) \to 1/\underline{R} > 0$. Therefore

$$s(a,y)\partial_a s(a,y) \to \nu(y) \quad \text{with} \quad \nu(y) := \frac{(r-\rho)u'(\underline{c}(y)) + (\mathcal{A}u'(\underline{c}))(y)}{u''(\underline{c}(y))}$$

as defined in the Proposition. Using l'Hôpital's rule we have

$$\lim_{a \to \underline{a}} \frac{(s(a,y))^2}{a - \underline{a}} = \lim_{a \to \underline{a}} 2s(a,y)\partial_a s(a,y) = 2\nu(y)$$

and hence $(s(a,y))^2 \sim 2\nu(y)(a - \underline{a})$. Taking the square root yields $s(a,y) \sim -\sqrt{2\nu(y)}\sqrt{a - \underline{a}}$.□

Proof of Proposition 7, Part 1: The proof of Part 1 follows identical steps as in the proof of Proposition 2.

Proof of Proposition 7, Part 2: Also the proof of Part 2 follows similar steps as the proof of Proposition 2. We still have Lemma 3 which shows that the case without labor income has a closed-form solution. We next prove a Lemma about a scaling property of the value function. We then combine the two Lemmas to prove the statement in the Proposition.

Lemma 10 *Consider problem (120). For any $\xi > 0$,*

$$v(\xi a, y) = \xi^{1-\gamma} v_\xi(a,y) \tag{123}$$

where v_ξ solves

$$\rho v_\xi(a,y) = \max_c \; u(c) + \partial_a v_\xi(a,y)(y/\xi + ra - c) + (\mathcal{A}v_\xi)(a,y) \tag{124}$$

Proof of Lemma 10: Write (120) as

$$\rho v(a,y) = H(\partial_a v(a,y)) + \partial_a v(a,y)(y + ra) + (\mathcal{A}v)(a,y) \tag{125}$$
$$H(p) = \max_c \{u(c) - pc\} = \frac{\gamma}{1-\gamma} p^{\frac{\gamma-1}{\gamma}}$$

From (123), $v(a,y) = \xi^{1-\gamma} v_\xi(a/\xi, y)$, $\partial_a v(a,y) = \xi^{-\gamma}\partial_a v_\xi(a/\xi,y)$. Therefore $H(\partial_a v(a,y)) = H(\partial_a v_\xi(a/\xi,y))\xi^{1-\gamma}$. Substituting into (125) and dividing by $\xi^{1-\gamma}$ yields (124).□

Conclusion of Proof of Part 2 of Proposition 7: With Lemmas 3 and 10 in hand we are ready to prove Part 2 of Proposition 7. Consider first the asymptotic behavior of the consumption policy function $c(a,y)$. From (123), $v(a,y) = \xi^{1-\gamma} v_\xi(a/\xi, y)$, $\partial_a v(a,y) = \xi^{-\gamma}\partial_a v_\xi(a/\xi,y)$ and therefore

$$c(a,y) = (\partial_a v(a,y))^{-1/\gamma} = \xi(\partial_a v_\xi(a/\xi,y))^{-1/\gamma} = \xi c_\xi(a/\xi,y)$$

In particular with $\xi = a$ we have

$$c(a,y) = a c_a(1,y)$$

Hence

$$\lim_{a\to\infty} \frac{c(a,y)}{a} = \lim_{\xi\to\infty} c_\xi(1,y) = c(1) = \frac{\rho-(1-\gamma)r}{\gamma},$$

where the second equality uses that problem (124) converges to that with no labor income (59) as $\xi \to \infty$ and therefore also $c_\xi(a,y) \to c(a)$ for all a as $\xi \to \infty$. The asymptotic behavior of $s(a,y)$ can be proved in an analogous fashion.□

G.2 An Alternative Way of Closing the Model: Aiyagari (1994)

Section 1 assumed that wealth takes the form of bonds that are in fixed supply. It is, of course, possible to make other assumptions. In particular, we can assume as in Aiyagari (1994) that wealth takes the form of productive capital that is used by a representative firm which also hires labor. Each individual's income is the product of an economy-wide wage w_t and her idiosyncratic labor productivity z_t and her wealth follows (2) with $y_t = w_t z_t$. The total amount of capital supplied in the economy equals the total amount of wealth. In a stationary equilibrium it is given by

$$K = \int_{\underline{z}}^{\bar{z}} \int_{\underline{a}}^{\infty} a g(a,z) da dz := S(r,w). \tag{126}$$

Capital depreciates at rate δ. There is a representative firm with a constant returns to scale production function $Y = F(K, L)$. Since factor markets are competitive, the wage and the interest rate are given by

$$r = \partial_K F(K, 1) - \delta, \qquad w = \partial_L F(K, 1), \tag{127}$$

where we use that the mean of the stationary distribution of productivities z equals one.

Because the income fluctuation problem at the heart of the Aiyagari model is the same as that in the Huggett model all of Propositions 1 to 3 apply without change. So does Proposition 4. Proposition 5 applies by exploiting a homogeneity property noted by Auclert and Rognlie (2016), namely that individual policy functions and therefore aggregate saving is homothetic in the wage rate, $S(r, w) = wS(r, 1)$ for all $w > 0$.[67] The computational algorithm is again unchanged except that, in Step 3, it imposes (126) and (127) rather than (11).

G.3 Soft Borrowing Constraints and Non-Participation

Empirical wealth distributions typically have the following properties: there are individuals with both positive and negative net worth but there is a spike at close to zero net worth. This empirical observation does not square well with the Aiyagari-Bewley-Huggett model we have considered thus far. If we set the borrowing constraint to $\underline{a} = 0$, we get the spike at zero but there are no individuals with negative net worth; if we set $\underline{a} < 0$, we get a spike at a strictly negative wealth level. Both are counterfactual. A simple way of generating the empirical observation is to model a wedge between the interest rates at which people can borrow and save. Such a wedge functions as a "soft" borrowing constraint. In particular, it is a soft version of a hard no-borrowing constraint (3) with $\underline{a} = 0$ in the sense that the soft constraint becomes closer and closer to the hard constraint as the interest-rate wedge increases (and is equivalent if the wedge is infinite). Such a soft constraint is used in a number of recent papers including Alonso (2018) and Kaplan, Moll, and Violante (2018). In this section, we provide the first theoretical characterization of such soft borrowing constraints.

Consider the Huggett model from Section 1 with one modification: there is a wedge between borrowing and lending rates. That is, we replace the budget constraint (2) by

$$\dot{a}_t = y_t + r(a_t)a_t - c_t, \quad r(a) = \begin{cases} r_+, & a \geq 0 \\ r_-, & a < 0 \end{cases}, \qquad r_- > r_+.$$

[67] Uniqueness requires one additional technical assumption about the production function F. To see this note that the homotheticity property implies that (126) becomes $S(r, 1) = k(r)$ where $k(r) := K(r)/w(r)$ is normalized capital demand. Since $S(r, 1)$ slopes upward by Proposition 5, the equilibrium is unique if $k(r)$ slopes downward. Auclert and Rognlie show that this is indeed the case if $\alpha < \varepsilon$ where α is the capital share and ε is the elasticity of substitution corresponding to F (both of which may depend on K/L).

We show below that, in order to obtain a stationary wealth distribution with a spike at zero and positive mass on both sides of zero, it is necessary to introduce more than two income types. In particular, the simplest extension of the model that yields the desired result is to have three income types i.e. $y_t \in \{y_0, y_1, y_2\}$ with $y_0 < y_1 < y_2$.

The next Proposition characterizes the saving behavior with a soft borrowing constraint. To avoid the somewhat cluttered notation resulting from considering three income types, it only considers the deterministic case $y_t = y$ for all t. This case has all the intuition and the extension to stochastic income is straightforward.

Proposition 8 (Saving Behavior with Soft Borrowing Constraint) *Assume that $r_+ < \rho < r_-$, that $y_t = y$ for all t and that $y > 0$ (so that $-u''(y)/u'(y) < \infty$, the analogue of Assumption 1). Then the solution to the HJB equation (7) and the corresponding saving policy function (9) have the following properties:*

1. *$s(0) = 0$ but $s(a) < 0$ all $a > 0$ and $s(a) > 0$ all $a < 0$.*
2. *close to $a = 0$, the saving and consumption policy functions satisfy*

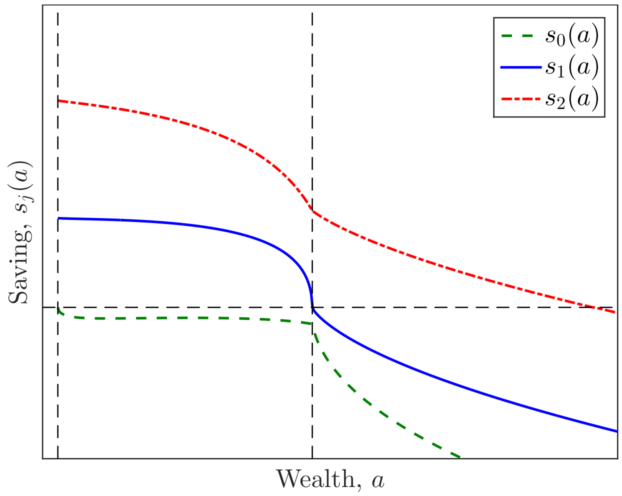
$$s(a) \sim -\sqrt{2\nu_+ a}, \quad c'(a) \sim r_+ + \frac{1}{2}\sqrt{\frac{2\nu_+}{a}}, \quad \nu_+ := \frac{(\rho - r_+)u'(y)}{-u''(y)} > 0 \quad \text{as } a \downarrow 0,$$
$$s(a) \sim \sqrt{2\nu_- a}, \quad c'(a) \sim r_- + \frac{1}{2}\sqrt{\frac{2\nu_-}{a}}, \quad \nu_- := \frac{(\rho - r_-)u'(y)}{-u''(y)} < 0 \quad \text{as } a \uparrow 0.$$

   This implies that the derivatives of s and c are unbounded at zero, with $s'(a) \to -\infty$ and $c'(a) \to \infty$ both as $a \uparrow 0$ and $a \downarrow 0$.

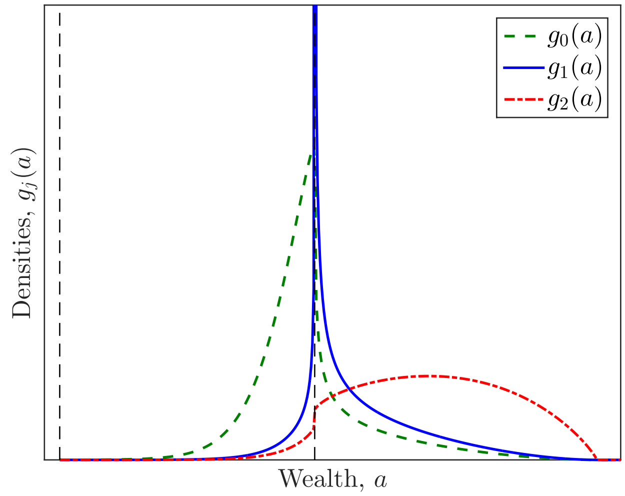
3. *Individuals with $a > 0$ decumulate wealth and hit $a = 0$ in finite time. Individuals with $a < 0$ instead accumulate wealth and also hit $a = 0$ in finite time.*

The main takeaway from the Proposition is that a soft borrowing constraint results in an interesting symmetry in the saving policy function around zero net worth. To understand this property consider the blue solid line labelled $s_1(a)$ in Figure 17(a) (we will return to the other two lines below). The behavior for $a > 0$ with a soft borrowing constraint is identical to that with a hard borrowing constraint but at $\underline{a} = 0$: as $a \downarrow 0$ it behaves like $-\sqrt{a}$. See for example Figure 1(b). The main takeaway from the Proposition is that the behavior of the saving policy function for $a < 0$ is simply a mirror image around the forty-five degree line of the behavior for $a > 0$: as $a \uparrow 0$ it behaves like $\sqrt{-a}$.[68] A simple extension of Corollary 1 then implies that individuals with $a > 0$ *decumulate* wealth and hit $a = 0$ in finite time; individuals with $a < 0$ instead *accumulate* wealth and also hit $a = 0$ in finite time.

[68] Interestingly, for $a < 0$ the consumption function is *convex*, that is, the instantaneous MPC $c'(a)$ is *increasing* in wealth a. See the Proposition which, ignoring constants, shows that $c'(a) \sim \sqrt{1/(-a)}$ as $a \uparrow 0$.



(a) Stationary Saving Policy Function



(b) Stationary Densities

Figure 17: Saving Behavior and Wealth Distribution with Soft Borrowing Constraint

Of course with only one income type, the stationary wealth distribution will only be a Dirac point mass at $a = 0$. With two income types, it will be a Dirac mass at $a = 0$ combined with some mass either to the left ($a < 0$) or to the right ($a > 0$) *but not both.* Therefore to speak to the empirical observation of a spike at zero combined with mass both to the left and the right of zero it is necessary to introduce (at least) another income type. Figure 17(a) plots the saving policy functions in such a version with three income types $y_0 < y_1 < y_2$. Figure 17(b) plots the resulting wealth densities g_0, g_1 and g_2. The unconditional wealth distribution is the sum of these three densities. As expected, it has a spike at zero and mass both to the left and the right.

Proof of Proposition 8 The proof follows the same steps as Proposition 1. In particular, from the Euler equation (envelope condition)

$$(s'(a) - r(a))s(a) = \frac{(r(a) - \rho)u'(c(a))}{u''(c(a))}$$

Therefore, taking the left and right limits as $a \downarrow 0$ and as $a \uparrow 0$, we have

$$s(a)s'(a) \to \nu_+, \quad \nu_+ := \frac{(\rho - r_+)u'(y)}{-u''(y)} \quad \text{as } a \downarrow 0$$
$$s(a)s'(a) \to \nu_-, \quad \nu_- := \frac{(\rho - r_-)u'(y)}{-u''(y)} \quad \text{as } a \uparrow 0$$

Note that $\nu_+ > 0$ because we have assumed $r_+ < \rho$ and $\nu_- < 0$ because $r_- > \rho$. By again following the same steps as in Proposition 1 we then obtain the expressions in Part 2. Part 3 follows from Corollary 1.

G.4 Fat Tails in a Huggett Model with Two Assets

In this section we show how to extend the Huggett model of Section 1 to feature a fat-tailed stationary wealth distribution. We do this by introducing a risky asset in addition to the riskless bond. The insight that the introduction of "investment risk" into a Bewley model generates a Pareto tail for the wealth distribution is due to Benhabib, Bisin, and Zhu (2015) and our argument mimics several of their steps.[69] Our result differs from theirs in three regards. First, Benhabib, Bisin and Zhu make their argument in an environment with a risky asset only and no market for bonds, i.e. no borrowing and lending. In contrast, we analyze a framework with two assets, a risky asset and a riskless bond that is in zero net supply. Our framework therefore nests both the standard Aiyagari-Bewley-Huggett model and the framework of Benhabib, Bisin and Zhu as special cases. Conveniently, in continuous time, analyzing a model with two assets poses no extra difficulty relative to the one-asset case.[70] Second, we obtain an easily interpretable analytic solution for the tail exponent of the wealth distribution and we show that, somewhat counterintuitively, top wealth inequality is *decreasing* in the riskiness of the risky asset. Finally, we explore the effects of both linear and progressive capital income taxation on top wealth inequality and macroeconomic performance.

Setup The setup is similar to that described in Section 1. We here keep the description as short as possible and focus on highlighting the differences between the two setups. The main difference to the previous setup is that individuals now have access to a real risky asset k_t in addition to the riskless bond which we now denote by b_t. With this additional asset, the budget constraint (2) now becomes

$$dk_t + db_t = (z_t + \tilde{R}_t k_t + r_t b_t - c_t)dt \tag{128}$$

where r_t is the return on the riskless bond, i.e. the real interest rate, as before and $\tilde{R}_t$ is the return on the risky asset. The return of the risky asset is stochastic and given by[71]

$$\tilde{R}_t dt = R dt + \sigma dW_t$$

[69] Also see Benhabib, Bisin, and Zhu (2011) and Benhabib, Bisin, and Zhu (2016). Quadrini (2009) and Cagetti and De Nardi (2006) argue for the importance of entrepreneurial risk in explaining the right tail of the wealth distribution, which is one particular form of investment risk. Also see Krusell and Smith (1998) and Castaneda, Diaz-Gimenez, and Rios-Rull (2003) for alternative mechanisms accounting for skewed wealth distributions in the data.

[70] In particular, the two assets can still be summarized by a single state variable, "net worth."

[71] This notation is somewhat unconventional. The more conventional notation is to denote the return over a time interval of length dt by $d\widehat{R}_t = Rdt + \sigma dW_t$ and to write the budget constraint (128) as $dk_t + db_t = (z_t + r_t b_t - c_t)dt + d\widehat{R}_t k_t$. This is equivalent and yields the same budget constraint in terms of net worth $a_t = k_t + b_t$, namely (128) below.

where R is a parameter and W_t is a standard Brownian motion, that is $dW_t \approx \lim_{\Delta t \to 0} \varepsilon_t \sqrt{\Delta t}$, with $\varepsilon_t \sim \mathcal{N}(0,1)$. The risky asset is a real asset in the sense that k_t units produce $\tilde{R}_t k_t$ units of physical output, and only positive asset positions are possible $k_t \geq 0$. One particularly appealing interpretation of the risky asset is that $\tilde{R}_t$ is the return from owning and running a private firm.[72] A negative $\tilde{R}_t$ captures strong enough depreciation. But other interpretations are possible as well. Finally, there is still a borrowing constraint which we now write as $b_t \geq -\phi$ with $\phi \geq 0$.

The problem of an individual can be simplified by writing the budget constraint in terms of wealth or net worth $a_t = b_t + k_t$:

$$da_t = (z_t + ra_t + (R - r)k_t - c_t)dt + \sigma k_t dW_t \tag{129}$$

Because capital satisfies $k_t \geq 0$, there is a state constraint $a_t \geq \underline{a} = -\phi$ as before. Similarly, the borrowing constraint $b_t \geq -\phi$ can be written as

$$k_t \leq a_t + \phi \tag{130}$$

Individuals maximize (1) subject to (129), (130) and the processes for z_t and $\tilde{R}_t$, taking as given the evolution of the equilibrium interest rate r_t for $t \geq 0$.

Stationary Equilibrium As before, individuals' saving decisions and the joint distribution of income and wealth can be summarized by means of a Hamilton-Jacobi-Bellman equation and a Kolmogorov Forward equation

$$\begin{aligned}\rho v_j(a) = \max_{c, 0 \leq k \leq a+\phi} \; & u(c) + v_j'(a)(z_j + ra + (R-r)k - c) \\ & + \frac{1}{2} v_j''(a)\sigma^2 k^2 + \lambda_j (v_{-j}(a) - v_j(a))\end{aligned} \tag{131}$$

$$0 = -\frac{d}{da}[s_j(a)g_j(a)] + \frac{1}{2}\frac{d^2}{da^2}[\sigma^2 k_j(a)^2 g_j(a)] - \lambda_j g_j(a) + \lambda_{-j} g_{-j}(a). \tag{132}$$

for $j = 1, 2$. As before, $s_j(a)$ is the optimal saving policy function and $k_j(a)$ is the optimal choice of the risky asset. It can be seen that (131) is an optimal portfolio allocation problem as in Merton (1969) and $k_j(a)/a$ is the share of the individual's portfolio invested in the risky asset. For example, $k_j(a) > a$ means that the individual borrows in riskless bonds so

[72] For example, assume that private firms produce using capital and labor using a constant returns to scale production functions $Z_t f(k_t, \ell_t)$ as in Angeletos (2007), and define

$$\tilde{R}_t k_t = \max_{\ell_t} \{ Z_t f(k_t, \ell_t) - w_t \ell_t - \delta_t k_t \}.$$

Then the process for $\tilde{R}_t$ inherits the properties of the process for Z_t. Also see Quadrini (2009) and Cagetti and De Nardi (2006) for related models of private firms.

as to invest into the risky asset. The interest rate r is determined in equilibrium by the fact that bonds are in zero net supply. The bond market clearing condition can be written as:

$$\int_{\underline{a}}^{\infty} k_1(a)g_1(a)da + \int_{\underline{a}}^{\infty} k_2(a)g_2(a)da = \int_{\underline{a}}^{\infty} ag_1(a)da + \int_{\underline{a}}^{\infty} ag_2(a)da$$

The Tail of the Wealth Distribution We now show that if individuals have CRRA utility (5), the stationary wealth distribution has a Pareto tail, and derive an analytic expression for the tail parameter. The key to this result is the following result.

Proposition 9 *With CRRA utility* (5)*, individual policy functions are asymptotically linear in a (as $a \to \infty$) and given by*

$$c_j(a) \sim \left(\frac{\rho - (1-\gamma)r}{\gamma} - \frac{1-\gamma}{2\gamma}\frac{(R-r)^2}{\gamma\sigma^2}\right)a \tag{133}$$

$$s_j(a) \sim \left(\frac{r-\rho}{\gamma} + \frac{1+\gamma}{2\gamma}\frac{(R-r)^2}{\gamma\sigma^2}\right)a \tag{134}$$

$$k_j(a) \sim \frac{R-r}{\gamma\sigma^2}a. \tag{135}$$

where for any two functions f and g, $f \sim g$ means $\lim_{a\to\infty} f(a)/g(a) = 1$*.*

The key idea of this result is that for large enough wealth a, labor income and the borrowing constraint become irrelevant, and so individual behavior will be like in a problem without labor income and without a borrowing constraint. And with CRRA utility, this problem is the portfolio allocation problem of Merton (1969) which can be solved analytically with the policy functions in Lemma 9.

Before proceeding to the main result of this section, we make one additional assumption on parameter values.

Assumption 2 $(R-r)^2 < 2\sigma^2(\rho - r)$.

This assumption states that the excess return on the risky asset cannot be too large relative to the riskiness of assets and the gap between the interest rate and the rate of time preference. With this assumption in hand, we obtain the following analytic solution for the fatness of the stationary wealth distribution.

Proposition 10 *With CRRA utility* (5) *and under Assumption 2, there is a unique stationary wealth distribution which follows an asymptotic power law, that is $1 - G(a) \sim ma^{-\zeta}$ with tail exponent*

$$\zeta = \gamma\left(\frac{2\sigma^2(\rho - r)}{(R-r)^2} - 1\right). \tag{136}$$

Therefore top wealth inequality $1/\zeta$ is decreasing in volatility σ, risk aversion γ, and the rate of time preference ρ, and increasing in the stationary interest rate r, and the excess return of risky assets $R - r$.

Somewhat counterintuitively, top wealth inequality is *decreasing* in the volatility of the risky asset. The reason for this is that there are two offsetting effects. On one hand, a higher σ has a direct effect in that more randomness in the risky asset leads to higher inequality. On the other hand, if σ increases, risk averse individuals optimally choose a smaller portfolio share of risky assets (see (135)) which is a force towards lower top wealth inequality. Formula (136) shows that the latter effect always dominates so that top wealth inequality $1/\zeta$ is unambiguously decreasing in volatility σ. Another way of stating this is that what matters for top wealth inequality is the volatility of wealth $\sigma k_j(a)$ and from (135) we have

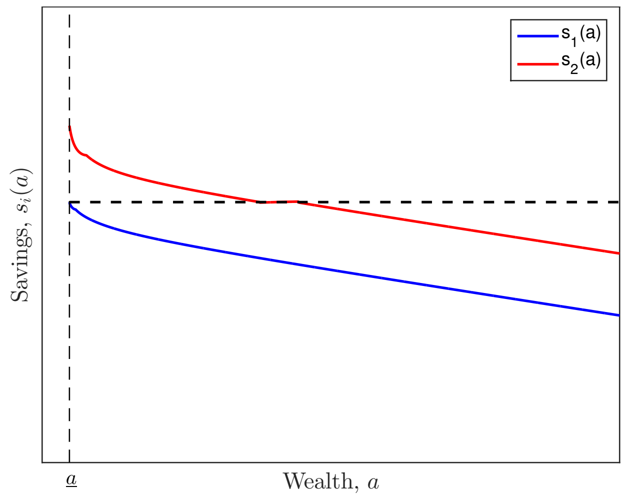
$$\sigma k_j(a) \sim \frac{R-r}{\gamma\sigma} a \tag{137}$$

which is decreasing in σ. The behavior of top wealth inequality with respect to the other parameter values is more intuitive: individuals invest a large share of their assets into risky assets when they are not too risk averse, or when the excess return of risky assets is high, and this also implies that wealth inequality is high. Also note that the fatness of the tail parameter does *not* depend in any way on the properties of the stochastic process for labor income (the income levels z_1, z_2 or the Poisson intensities λ_1, λ_2). This property of models with investment risk was first pointed out by Benhabib, Bisin, and Zhu (2011) using the theory of "Kesten processes" in a discrete-time model with investment risk.

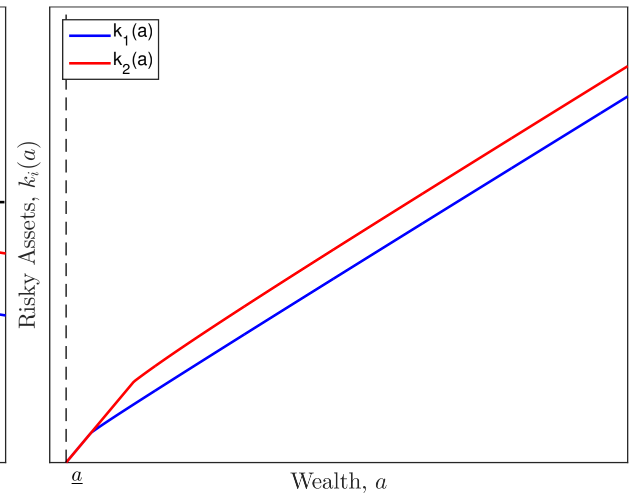
Proposition 10 provides a powerful formula for calibrating models with investment risk. Empirically, wealth distributions for developed countries like the United States feature a high degree of concentration with a tail exponent of $\zeta \approx 1.5$. From (136) it can be seen that the model can generate such high wealth concentration quite easily. For example with a standard risk aversion parameter of $\gamma = 2$, an excess return of four percent, $R - r = 0.04$, a gap between interest rate and rate of time preference of $\rho - r = 0.035$, and a standard deviation of returns of twenty percent, $\sigma = 0.2$, we get $\zeta = 1.5$ just like in the data.

Figure 18 plots individuals' optimal choices and the resulting wealth distribution for the model with both a risky and a riskless asset. Panel (d) in particular shows that the distribution behaves asymptotically like a Pareto distribution by showing that the logarithm of the density of log wealth $f_i(x)$ is asymptotically linear in the logarithm of wealth $x = \log(a)$.[73] Table 2 reports the results of a calibration exercise for the wealth distribution in a stationary equilibrium. It can be seen that the model matches the empirical wealth

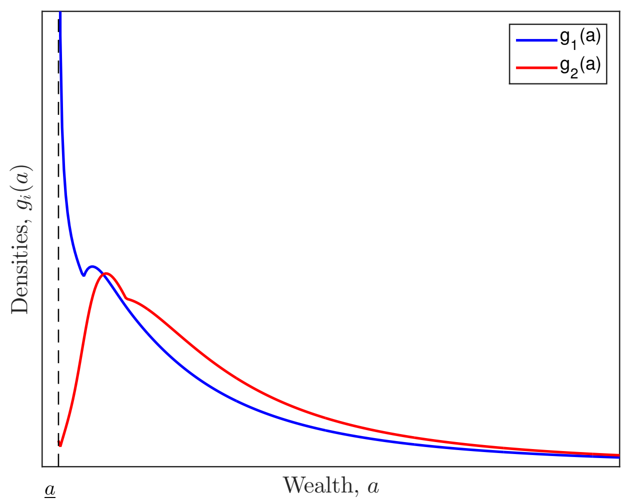
[73]We here use the fact that if a variable a follows a Pareto distribution $g(a) \propto a^{-\zeta-1}$, then $x = \log a$ follows an exponential distribution $f(x) \propto e^{-\zeta x}$ and hence $\log f(x)$ is a linear function of x where the slope equals the tail exponent ζ.



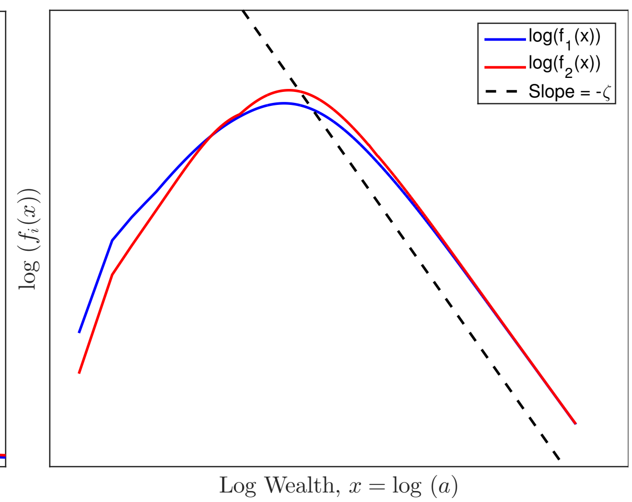
(a) Saving Policy Function



(b) Investment in Risky Assets



(c) Stationary Densities



(d) Pareto Tail of Stationary Distribution

Figure 18: Optimal Choices and Pareto Tail of Wealth Distribution in Two-Asset Model

distribution of the United States quite well, particularly at the top.[74]

Effect of Capital Income Taxation on Top Wealth Inequality We briefly examine the question how a tax on capital income affects top wealth inequality. To this end, we introduce a linear tax on capital income into our version of Huggett's model with both a risky and a riskless asset. We modify the budget constraint (128) to

$$dk_t + db_t = (z_t + (1-\tau)(\tilde{R}_t k_t + r_t b_t) + T_t - c_t)dt$$

[74]The parameter values are $\gamma = 2, \rho = 0.05, \sigma = 0.56, \lambda_1 = \lambda_2 = 0.5, z_1 = 0.4, z_2 = 0.6, \phi = 1.5$ and the equilibrium interest rate is $r = 0.0492$. It should be possible to further improve the fit at the bottom by allowing for a looser borrowing limit ϕ so that a larger fraction of individuals hold negative wealth.

	U.S. Data	Model
Tail Exponent ζ	1.5	1.55
Top 1% wealth share	34.6%	34.6%
Next 9% wealth share	38.0%	32.8%
Next 40 % wealth share	26.7%	26.3%
Bottom 50 % wealth share	0.7%	6.3%

Table 2: Wealth Distribution in Model vs. Data (Source: Survey of Consumer Finances)

where τ is the linear tax on capital income and T_t are lump-sum transfers. We assume that the government balances its budget each period and redistributes revenues from capital income taxation equally to all individuals. It is not hard to show that the formula for top wealth inequality (136) becomes

$$\zeta = \gamma \left(\frac{2\sigma^2(\rho - r(1-\tau))}{(R-r)^2} - 1 \right)$$

A higher capital income tax rate τ lowers top wealth inequality. Interestingly, capital taxation affects top wealth inequality *only* through its effect on the return of the riskless asset. This is because a linear capital income tax does not affect the volatility of wealth in (137). On one hand, a high tax rate directly lowers the effective variance of the risky asset $\sigma(1-\tau)$. On the other hand, this reduced riskiness implies that individuals invest a larger fraction of their wealth into risky assets. The two effects exactly offset each other as can be seen from (137).

Proof of Proposition 9 Before proceeding to the proof of the result, we derive two auxiliary Lemmas. The first Lemma considers an auxiliary problem without labor income, $y_1 = y_2 = 0$, and without a borrowing constraint, $\phi = \infty$ and shows that optimal policy functions are linear in wealth. The second Lemma shows that the problem with labor income and a borrowing constraint (131) satisfies a certain homogeneity property.

Lemma 11 *Consider the problem*

$$\rho v(a) = \max_{c,k} \; u(c) + v'(a)(ra + (R-r)k - c) + \frac{1}{2}v''(a)\sigma^2 k^2 \tag{138}$$

where $u(c) = c^{1-\gamma}/(1-\gamma), \gamma > 0$*. The optimal policy functions that solve* (138) *are linear* *in wealth and given by*

$$c(a) = \left(\frac{\rho - (1-\gamma) r}{\gamma} - \frac{1}{2} \frac{(R-r)^2}{\sigma^2} \frac{1-\gamma}{\gamma^2} \right) a \tag{139}$$

$$s(a) = \left(\frac{r-\rho}{\gamma} + \frac{1+\gamma}{2\gamma} \frac{(R-r)^2}{\gamma \sigma^2} \right) a \tag{140}$$

$$k(a) = \frac{R-r}{\gamma \sigma^2} a \tag{141}$$

Proof of Lemma 11 Grouping terms by the relevant maximization problems and solving these, we can write

$$\rho v(a) = H(v'(a)) + G(v'(a), v''(a)) + v'(a) r a \tag{142}$$

$$H(p) = \max_c \{u(c) - pc\} = \frac{\gamma}{1-\gamma} p^{\frac{\gamma-1}{\gamma}}$$

$$G(p,q) = \max_k \left\{ p(R-r)k + \frac{1}{2} q \sigma^2 k^2 \right\} = \frac{1}{2} \frac{p^2}{-q} \frac{(R-r)^2}{\sigma^2}$$

and from the first-order conditions

$$u'(c(a)) = v'(a), \quad k(a) = -\frac{v'(a)}{v''(a)} \frac{R-r}{\sigma^2} \tag{143}$$

Guess and verify $v(a) = Ba^{1-\gamma}$ and hence $v'(a) = (1-\gamma)Ba^{-\gamma}, v''(a) = -\gamma(1-\gamma)Ba^{-\gamma-1}$

$$H(v'(a)) = \frac{\gamma}{1-\gamma} (v'(a))^{\frac{\gamma-1}{\gamma}} = \frac{\gamma}{1-\gamma} ((1-\gamma)B)^{\frac{\gamma-1}{\gamma}} a^{1-\gamma}$$

$$\frac{(v'(a))^2}{-v''(a)} = \frac{(1-\gamma)B}{\gamma} a^{1-\gamma}$$

$$G(v'(a), v''(a)) = \frac{1}{2} \frac{(v'(a))^2}{-v''(a)} \frac{(R-r)^2}{\sigma^2} = \frac{1}{2} \frac{(R-r)^2}{\sigma^2} \frac{(1-\gamma)B}{\gamma} a^{1-\gamma}$$

Substituting into (142) and dividing by $Ba^{1-\gamma}$, we have

$$\rho = \gamma((1-\gamma)B)^{-\frac{1}{\gamma}} + \frac{1}{2} \frac{(R-r)^2}{\sigma^2} \frac{1-\gamma}{\gamma} + (1-\gamma) r. \tag{144}$$

From (143) $c(a) = ((1-\gamma)B)^{-\frac{1}{\gamma}} a$ and hence using (144) we obtain (139), (140) and (141).□

Lemma 12 *Consider the problem* (131)*. For any* $\xi > 0$*,*

$$v_j(\xi a) = \xi^{1-\gamma} v_{\xi,j}(a) \tag{145}$$

where $v_{\xi,j}$ *solves*

$$\rho v_{\xi,j}(a) = \max_{c,0\le k\le a+\phi/\xi} \quad u(c) + v_j'(a)(y_j/\xi + ra + (R-r)k - c) \\ + \frac{1}{2}v_{\xi,j}''(a)\sigma^2 k^2 + \lambda_j(v_{\xi,-j}(a) - v_{\xi,j}(a)) \tag{146}$$

Proof of Lemma 12 The proof follows exactly the same steps as the proof of the second part of Proposition 2 and is therefore ommitted.□

Also the conclusion of the proof combines the preceding two Lemmas in exactly the same manner as in the proof of the second part of Proposition 2. We therefore again omit it.□

Proof of Proposition 10 The following argument shows that if there exists a stationary distribution, it must have a Pareto tail with a tail parameter (136). Adding the two Kolmogorov Forward (Fokker-Planck) equations (132)

$$0 = -\frac{d}{da}[s_1(a)g_1(a) + s_2(a)g_2(a)] + \frac{\sigma^2}{2}\frac{d^2}{da^2}[k_1(a)^2 g_1(a) + k_2(a)^2 g_2(a)]. \tag{147}$$

From Proposition 9, for large a we have $s_j(a) = \tilde{s}_j + \bar{s}a$ and $k_j(a) = \tilde{k}_j + \bar{k}a$ where

$$\bar{s} = \frac{r-\rho}{\gamma} + \frac{1+\gamma}{2\gamma}\frac{(R-r)^2}{\gamma\sigma^2}, \quad \bar{k} = \frac{R-r}{\gamma\sigma^2} \tag{148}$$

A heuristic argument is to use a "guess-and-verify" strategy, i.e. guess that $g(a) = g_1(a) + g_2(a) = \xi a^{-\zeta-1}$, and verify that the guess solves (147) for large enough a (all other terms go to zero as $a \to \infty$). We here present a more rigorous and constructive proof. Integrating (147)

$$\frac{\sigma^2}{2}\frac{d}{da}[k_1(a)^2 g_1(a) + k_2(a)^2 g_2(a)] = [s_1(a)g_1(a) + s_2(a)g_2(a)] + C. \tag{149}$$

As in the proof of Proposition 3, we choose $C = 0$ as an implicit boundary condition. Later we will check that the solution does satisfy this condition. Now we define $y_j(a) = \sigma^2 k_j(a)^2 g_j(a)/2$, and rewrite (149) as

$$y_1'(a) + y_2'(a) = \frac{2s_1(a)}{\sigma^2 k_1(a)^2}y_1(a) + \frac{2s_2(a)}{\sigma^2 k_2(a)^2}y_2(a). \tag{150}$$

Define $y(a) = y_1(a) + y_2(a)$. After collecting the leading term, (150) is written as

$$y'(a) = \frac{\theta}{a}y(a) + h_1(a)y_1(a) + h_2(a)y_2(a), \tag{151}$$

$$\theta = \frac{2\bar{s}}{\sigma^2\bar{k}^2}, \quad h_j(a) = \frac{2}{\sigma^2}\left(\frac{\tilde{s}_j + \bar{s}a}{(\tilde{k}_j + \bar{k}a)^2} - \frac{\bar{s}}{\bar{k}^2 a}\right), \ j = 1, 2.$$

Dividing (151) by $y(a)$ and integrating both sides from a_1 to a_2 where $a_1 < a_2$ are large enough, we have

$$\ln\left(\frac{y(a_2)}{a_2^\theta}\right) - \ln\left(\frac{y(a_1)}{a_1^\theta}\right) = \int_{a_1}^{a_2} \frac{h_1(x)y_1(x)}{y(x)}dx + \int_{a_1}^{a_2} \frac{h_2(x)y_2(x)}{y(x)}dx. \tag{152}$$

Note that there exists a positive constant $\bar{C}$ such that $|h_j(a)| \leq \bar{C}/a^2,\ j = 1, 2$ and $y_j > 0$. Therefore we have

$$\left|\ln\left(\frac{y(a_2)}{a_2^\theta}\right) - \ln\left(\frac{y(a_1)}{a_1^\theta}\right)\right| \leq \int_{a_1}^{a_2} \frac{\bar{C}}{x^2}\left(\frac{y_1(x)}{y(x)} + \frac{y_2(x)}{y(x)}\right)dx \leq \bar{C}\left(\frac{1}{a_1} - \frac{1}{a_2}\right).$$

Hence there exists $\bar{\xi}$ such that

$$\lim_{a\to\infty} \ln\left(\frac{y(a)}{a^\theta}\right) = \bar{\xi}.$$

Recalling the definition of $y(a) = \sigma^2 g(a)(k_1(a)^2 + k_2(a)^2)/2$, we have

$$\lim_{a\to\infty} \frac{g(a)}{a^{\theta-2}} = \frac{2\exp(\bar{\xi})}{\sigma^2\bar{k}^2}. \tag{153}$$

Equivalently

$$g(a) \sim \xi a^{-\zeta-1}, \quad \zeta = 1 - \theta = 1 - \frac{2\bar{s}}{\sigma^2\bar{k}^2}, \quad \xi = \frac{2\exp(\bar{\xi})}{\sigma^2\bar{k}^2}.$$

Finally, substituting the expressions for $\bar{s}$ and $\bar{k}$ in (148) into the expression for ζ yields (136).□

Appendix References

Albert, S., B. Cockburn, D. A. French, and T. E. Peterson (2002): "A Posteriori Error Estimates for General Numerical Methods for Hamilton-Jacobi Equations. Part I: The Steady State Case," *Mathematics of Computation*, 71(237), 49–76.

Alonso, C. (2018): "Hard vs. soft financial constraints: Implications for the effects of a credit crunch," .

Angeletos, G.-M. (2007): "Uninsured Idiosyncratic Investment Risk and Aggregate Saving," *Review of Economic Dynamics*, 10, 1–30.

Bertucci, C., J. M. Lasry, and P. L. Lions (2018): "Some remarks on Mean Field Games," *ArXiv e-prints*.

Bogachev, V., N. Krylov, M. Röckner, and S. Shaposhnikov (2015): *Fokker-Planck-Kolmogorov Equations*. American Mathematical Society.

Cagetti, M., and M. De Nardi (2006): "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, 114(5), 835–870.

Cannarsa, P., R. Capuani, and P. Cardaliaguet (2018): "Mean Field Games with state constraints: from mild to pointwise solutions of the PDE system," .

Cardaliaguet, P., F. Delarue, J.-M. Lasry, and P.-L. Lions (2019): *The master equation* *and the convergence problem in mean field games*, vol. 201 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ.

Carmona, R., and F. Delarue (2018a): *Probabilistic theory of mean field games with applications. I*, vol. 83 of *Probability Theory and Stochastic Modelling*. Springer, Cham, Mean field FBSDEs, control, and games.

——— (2018b): *Probabilistic theory of mean field games with applications. II*, vol. 84 of *Probability Theory and Stochastic Modelling*. Springer, Cham, Mean field games with common noise and master equations.

Castaneda, A., J. Diaz-Gimenez, and J.-V. Rios-Rull (2003): "Accounting for the U.S. Earnings and Wealth Inequality," *Journal of Political Economy*, 111(4), 818–857.

Cockburn, B., I. Merev, and J. Qian (2013): "Local a Posteriori Error Estimates for Time-dependent Hamilton-Jacobi Equations," *Mathematics of Computation*, 82(281), 187–212.

Crandall, M. G., and P. L. Lions (1984): "Two Approximations of Solutions of Hamilton-Jacobi Equations," *Mathematics of Computation*, 43(167), 1–19.

Evans, L. (2008): "Partial Differential Equations," in *The Princeton Companion to Mathematics*, ed. by T. Gowers, J. Barrow-Green, and I. Leader. Princeton University Press, https://math.berkeley.edu/~evans/evans_pcam.pdf.

Kredler, M. (2014): "The Kolmogorov Forward Equation, http://www.eco.uc3m.es/~mkredler/ContTime/KolmForwEqu.pdf," Lecture notes.

Merton, R. C. (1969): "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," *The Review of Economics and Statistics*, 51(3), 247–57.

Quadrini, V. (2009): "Entrepreneurship in macroeconomics," *Annals of Finance*, 5(3), 295–311.

Rouwenhorst, K. G. (1995): *Asset Pricing Implications of Equilibrium Business Cycle Models*pp. 294–330. Princeton University Press, Princeton, NJ.