Online Appendix – Not for Publication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. These results are available upon request. Now, consider the "Euler equation" (18) for type $j = 1$. Rearranging

$$\frac{u''(c_1(a))}{u'(c_1(a))} c_1'(a) s_1(a) = \rho - r - \lambda_1 \left(\frac{u'(c_2(a))}{u'(c_1(a))} - 1 \right) \tag{63}$$

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 (63) 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: $s_1(\underline{a}) = 0$.[58]

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

$$(s_1'(a) - r) s_1(a) = \frac{(r - \rho) u'(c_1(a)) + \lambda_1 (u'(c_2(a)) - u'(c_1(a)))}{u''(c_1(a))} \tag{64}$$

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 $-u'(c_1(a))/u''(c_1(a)) \to 1/\underline{R} > 0$. Therefore

$$s_1(a) s_1'(a) \to \nu_1 \quad \text{with} \quad \nu_1 := \frac{(r - \rho) u'(\underline{c}_1) + \lambda_1 (u'(\underline{c}_2) - u'(\underline{c}_1))}{u''(\underline{c}_1)} > 0 \tag{65}$$

as defined in (21). We have

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

where the first equality follows from l'Hôpital's rule and the second equality uses (65). Hence

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

[58] 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 (63) holding.

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.□

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*

$$s_1(a) \sim -\eta_1(a - \underline{a}), \tag{66}$$

$$c_1(a) \sim y_1 + ra + \eta_1(a - \underline{a}), \tag{67}$$

$$c_1'(a) \sim r + \eta_1, \tag{68}$$

$$\eta_1 := \frac{\rho - r + \lambda_1(1 - \xi)}{\underline{\gamma}}, \quad \underline{\gamma} := -\lim_{a \to \underline{a}} \frac{u''(c_1(a))c_1(a)}{u'(c_1(a))}, \quad \xi := \lim_{a \to \underline{a}} \frac{u'(c_2(a))}{u'(c_1(a))} \tag{69}$$

   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 (69) 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 (64) in the proof of Proposition 1 still holds. However, we now have $-u'(c_1(a))/u''(c_1(a)) \to 0$ as $a \to \underline{a}$. 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 (69). Finally, defining $\eta_1 := -s_1'(\underline{a}) = -(r-c_1'(\underline{a}))$ we have (69).□

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 $s_1(a) \leq 0$. 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 $j = 2$

$$\frac{u''(c_2(a))}{u'(c_2(a))}c_2'(a)s_2(a) = \rho - r - \lambda_2\left(\frac{u'(c_1(a))}{u'(c_2(a))} - 1\right) \tag{70}$$

In contrast to the expression for type $j = 1$, (63), the sign of the right-hand side of (70) is ambiguous. However, 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

$$\frac{u'(c_1(a))}{u'(c_2(a))} \leq \left(\frac{c_2(a)}{c_1(a)}\right)^{\bar{\gamma}}. \tag{71}$$

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 (70) 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.

Part 1, behavior of s_2 close to $a_{\max}$: 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

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

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 $a \to a_{\max}$.$\square$

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*

$$\rho v(a) = \max_c \; u(c) + v'(a)(ra - c) \tag{72}$$

where the utility function is given by (5)*. The optimal policy functions that solve* (72) *are linear in wealth and given by*

$$c(a) = \frac{\rho - (1-\gamma)r}{\gamma} a, \qquad s(a) = \frac{r-\rho}{\gamma} a. \tag{73}$$

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

$$v'(a) = Ba^{-\gamma} \tag{74}$$

$$c(a) = v'(a)^{-1/\gamma} = B^{-1/\gamma} a \tag{75}$$

Substituting into (72) 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 (75) we have (73).□

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

$$v_j(\xi a) = \xi^{1-\gamma} v_{\xi,j}(a) \tag{76}$$

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

$$\rho v_{\xi,j}(a) = \max_c \; u(c) + v'_{\xi,j}(a)(y_j/\xi + ra - c) + \lambda_j(v_{\xi,-j}(a) - v_{\xi,j}(a)) \tag{77}$$

Proof of Lemma 4: Write (7) as

$$\rho v_j(a) = H(v'_j(a)) + v'_j(a)(y_j + ra) + \lambda_j(v_{-j}(a) - v_j(a)) \tag{78}$$

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

From (76), $v_j(a) = \xi^{1-\gamma} v_{\xi,j}(a/\xi)$, $v'_j(a) = \xi^{-\gamma} v'_{\xi,j}(a/\xi)$. Therefore $H(v'_j(a)) = H(v'_{\xi,j}(a/\xi))\xi^{1-\gamma}$. Substituting into (78) and dividing by $\xi^{1-\gamma}$ yields (77).□

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

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

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 (77) converges to that with no labor income (72) 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 (34) with respect to starting wealth a:

$$\text{MPS}_{1,\tau}(a) \sim \left(1 - \tau\sqrt{\frac{\nu_1}{2(a-\underline{a})}}\right)^+. \tag{79}$$

To find the marginal propensity to consume (MPC) in (35) 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[59]

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

Differentiating with respect to starting wealth a, we have $\text{MPS}_\tau(a) + \text{MPC}_\tau(a) = 1 + \tau r$. Using (79) we obtain (35).□

A.4 Proof of Proposition 3

Integrating (36), 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 (38). 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.4.5 below.

[59]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.

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 (38), 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 (38) will be identical to the behavior of this limit. Assume that the *leading 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 (40).

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 (38) and the fact that s_1, s_2 are continuous and differentiable.

A.4.4 Part 4: 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.

A.4.5 Constants of Integration for Stationary Distribution (38)

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 (37), i.e. we can pin down the constants of integration (κ_1, κ_2) in (38). In the case with the Dirac mass (if Assumption 1 holds), this condition is

$$m_1 + \lim_{\varepsilon \to 0} \int_{\underline{a}+\varepsilon}^{a_{\max}} g_1(a) da = \frac{\lambda_2}{\lambda_1 + \lambda_2}, \qquad \int_{\underline{a}}^{a_{\max}} g_2(a) da = \frac{\lambda_1}{\lambda_1 + \lambda_2} \tag{80}$$

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:*

$$0 = -\lim_{\varepsilon \to 0} s_1(\underline{a} + \varepsilon) g_1(\underline{a} + \varepsilon) - \lambda_1 m_1 \tag{81}$$

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 (80), we have (81).□

Equation (81) can be used as a boundary condition for (38). From (38) 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

$$g_1(a) = -\frac{\lambda_1 m_1}{s_1(a)} \exp\left(-\int_{\underline{a}}^{a} \left(\frac{\lambda_1}{s_1(x)} + \frac{\lambda_2}{s_2(x)}\right) dx\right), \tag{82}$$

$$g_2(a) = +\frac{\lambda_1 m_1}{s_2(a)} \exp\left(-\int_{\underline{a}}^{a} \left(\frac{\lambda_1}{s_1(x)} + \frac{\lambda_2}{s_2(x)}\right) dx\right). \tag{83}$$

Substituting (83) into (80) 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:

$$m_1 = \frac{\lambda_2}{\lambda_1 + \lambda_2}\tilde{m}_1, \quad \frac{1}{\tilde{m}_1} = \lambda_2 \int_{\underline{a}}^{a_{\max}} \left\{\frac{1}{s_2(a)} \exp\left(-\int_{\underline{a}}^{a} \left(\frac{\lambda_1}{s_1(x)} + \frac{\lambda_2}{s_2(x)}\right) dx\right)\right\} da. \tag{84}$$

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 (80) with $m_1 = 0$ and these two equations pin down the constants of integration in (38).

A.5 Proof of Proposition 4

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 (38) 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 (63) and (70) 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 (70). As $r \downarrow -\infty$, the right-hand side of(70) 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

$$\lim_{r \downarrow -\infty} S(r) = \underline{a}. \tag{85}$$

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 (70) 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 (38) that

$$\lim_{r \uparrow \rho} S(r) = \infty \tag{86}$$

Given that S is continuous in r and satisfies (85) and (86), it must intersect zero at least once, proving the existence of a stationary equilibrium.□

A.6 Proof of Proposition 5

As mentioned in Section 2.6, our uniqueness result not only applies to the two-state income process analyzed in Sections 1 and 2 but also to any stationary Markovian income process, e.g. the diffusion process of Section 4.1. To treat the general case, we write the HJB and KF equations as

$$\rho v = \max_{c} \; u(c) + (y + ra - c)\partial_a v + \mathcal{A}v, \tag{87}$$

$$0 = -\partial_a(s(a, y)g) + \mathcal{A}^* g, \tag{88}$$

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 $(\mathcal{A}v)(a, y_j) = \eta_j(v(a, y_{-j}) - v(a, y_j))$. Or if y_t is a continuous diffusion as in Section 4.1, then $\mathcal{A}v = \mu(y)\partial_y v + \frac{\sigma^2(y)}{2}\partial_{yy}v$. Readers who are not familiar with infinitesimal generators and so on, 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.6.1 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 (45) holds. The proof combines two Lemmas. The first Lemma is due to Olivi (2017).

Lemma 6 (Olivi, 2017) *Consider the HJB equation (87). The corresponding consumption policy function satisfies*

$$\frac{\partial c(a, y)}{\partial r} = \frac{1}{u''(c_0)}\mathbb{E}_0 \int_0^\tau e^{-\int_0^t \xi_s ds}\{u'(c_t) + u''(c_t)(\partial_a c_t)a_t\}dt \tag{89}$$

with $\xi_t := \rho - r + \partial_a c_t > 0$ and where $\tau := \inf\{t \geq 0 | a_t = 0\}$ 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 6: Define $\eta(a, y) := \partial_a v(a, y)$. Differentiating (87) 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

$$(\rho - r + \partial_a c)\partial_r \eta = \eta + a\partial_a \eta + \partial_a[\partial_r \eta]s + \mathcal{A}\partial_r \eta. \tag{90}$$

We next evaluate $\partial_r \eta$ in (90) 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[60]

$$\mathbb{E}_t[d(\partial_r \eta_t)] = [\partial_a(\partial_r \eta(a_t, y_t))s(a_t, y_t) + \mathcal{A}\partial_r \eta(a_t, y_t)]\, dt. \tag{91}$$

and hence (90) is

$$\xi_t \partial_r \eta_t = \eta_t + a_t \partial_a \eta_t + \frac{1}{dt}\mathbb{E}_t[d(\partial_r \eta_t)], \qquad \xi_t := \rho - r + \partial_a c_t > 0. \tag{92}$$

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 $\tau := \inf\{t \geq 0 | a_t = 0\}$ the first time the process a_t hits the borrowing constraint $\underline{a} = 0$. Note that τ is a stopping time and itself a random variable. Integrating (92), we have that for any $T < \tau$

$$\partial_r \eta_0 = \mathbb{E}_0\left[\int_0^T e^{-\int_0^t \xi_s ds}\{\eta_t + a_t \partial_a \eta_t\}dt + e^{-\int_0^T \xi_s ds}\partial_r \eta_T\right]. \tag{93}$$

Now consider the limit as $T \to \tau$. First, recall that from the state constraint boundary condition (15) we have $\eta(a_\tau, y_\tau) = u'(y_\tau + ra_\tau)$ and therefore $\partial_r \eta(a_\tau, y_\tau) < \infty$. On the other

[60] First consider the case when y_t follows a diffusion process (54). 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 (91). 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 (91) still holds.

hand, as $T \to \tau$, we have $\partial_a c(a_T, y_T) \to \infty$ and therefore $\xi_T \to \infty$. Hence for any sample path (a_t, y_t) and corresponding stopping time τ, we have

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

Therefore (93) implies

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

and from the first-order condition $\eta_t = u'(c_t)$ we immediately obtain (89).$\square$

Lemma 7 *Assume that the IES is weakly greater than one, i.e.* (45) *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 7: 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 (45). The first weak inequality uses that the IES is greater than one from (45). 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 $a \geq 0$.[61] The strict inequality at the end uses that $c(0,y) > 0$ for all y. $\square$

Conclusion of Proof of Proposition 5, Part 1: The proof of Part 1 concludes by combining Lemmas 6 and 7. From Lemma 6 we see that $c(a,y)$ is strictly decreasing in r if (i) $\tau > 0$ so that the integral in (89) is different from zero, and if (ii) $u'(c_t) + u''(c_t)(\partial_a c_t)a_t > 0$ point-by-point in the integral in (89) over sample paths $(a_t, y_t)_{0 \leq t \leq \tau}$. Requirement (i) that $\tau > 0$ 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 7 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 $a \geq 0$: because $u''(c) < 0$, $u'(c(a,y)) +$

[61]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.

$u''(c(a,y))\partial_a c(a,y)a > 0$ for all $a < 0$, independently of Lemma 7. Hence consumption is strictly decreasing in r even if we allow for borrowing, $a < 0$.$\square$

A.6.2 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.6.3 Proof of Proposition 5, Part 3: First-order Stochastic Dominance

Integrating the stationary Kolmogorov Forward equation (8) and using that $s_j(0)g_j(0) = 0$ for $j = 1, 2$, we have

$$0 = -s_j(a)\partial_a G_j(a) - \lambda_j G_j(a) + \lambda_{-j}G_{-j}(a), \quad j = 1, 2. \tag{94}$$

for $a > 0$. We have further shown in Proposition 3 that for each $r < \rho$ there is an $a_{\max}(r) < \infty$ such that the stationary distribution has no mass above $a_{\max}(r)$, $g_j(a) = 0$ for all $a \geq a_{\max}(r)$ and $j = 1, 2$. Or in terms of the CDF $G_j(a) = \lambda_{-j}/(\lambda_1 + \lambda_2)$ for all $a \geq a_{\max}(r)$ and $j = 1, 2$. For the purpose of the proof, it is convenient to analyze (94) on a bounded domain and one that is independent of r. Because $a_{\max}(r)$ is the point at which $s_2(a;r)$ intersects zero, it is increasing in r. There is then an $\bar{a} < \infty$ such that $a_{\max}(r) \leq \bar{a}$ for all $r < \rho$ (the relevant range of r). For all r, we therefore consider (94) on the bounded domain $(0, \bar{a})$ with boundary condition $G_j(\bar{a}) = \lambda_{-j}/(\lambda_1 + \lambda_2)$ for $j = 1, 2$.

Next define

$$f_j(a) := \frac{\partial G_j(a)}{\partial r}, \quad j = 1, 2.$$

The goal is to prove that $f_j(a) \leq 0$ for all $a \in (0, \bar{a})$ and $j = 1, 2$.[62] Differentiating (94) with respect to r, we have

$$0 = -\frac{\partial s_j}{\partial r} g_j - s_j f_j' - \lambda_j f_j + \lambda_{-j} f_{-j}, \quad a \in (0, \bar{a}), \quad j = 1, 2, \tag{95}$$

with boundary conditions $f_j(\bar{a}) = \partial G_j(\bar{a})/\partial r = 0, j = 1, 2$ and $f_2(0) = \partial G_2(0)/\partial r = 0$. In

[62] As an aside, another way of stating Part 3 is: if $s_j(a; r_H) > s_j(a; r_L)$ for all a and $j = 1, 2$, and $G_j(a; r_H), G_j(a; r_L)$ solve (94) on $(0, \bar{a})$ then $G_j(a; r_H) < G_j(a; r_L)$ for all $a \in (0, \bar{a})$. In mathematics, this type of theorem is known as a "comparison theorem". The strategy we follow in this proof is a relatively general strategy for establishing such comparison theorems.

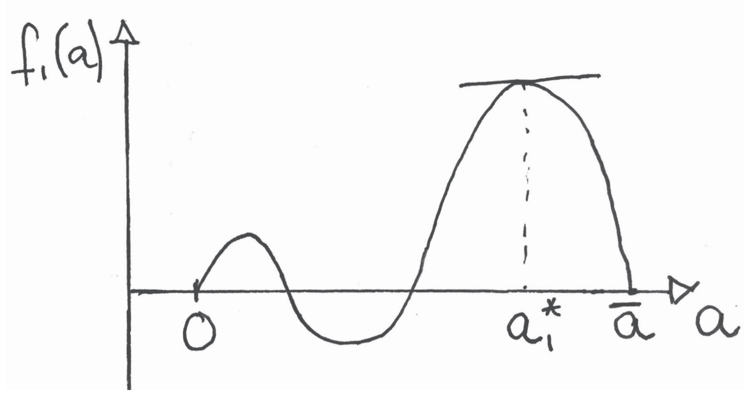


Figure 12: The function $f_1(a) = \partial G_1(a)/\partial r$

contrast, $f_1(0) = \partial G_1(0)/\partial r$ may be non-zero because there is, in general, a Dirac mass of low income types at the borrowing constraint, $G_1(0) > 0$. There are two cases: (i) $f_1(0) \leq 0$ and (ii) $f_1(0) > 0$.

Case 1: $f_1(0) \leq 0$**:** To obtain a contradiction, assume that $f_1(a) > 0$ or $f_2(a) > 0$ for some a. First note that these can only occur together, i.e. if $f_1(a) > 0$ for some a, then also $f_2(a) > 0$ for some a and vice versa. To see this, note that if $f_1(a) > 0$ for some a, then because $f_1(0) \leq 0$ and $f_1(\bar{a}) = 0$, f_1 has a global maximum for some interior a_1^*. The situation is depicted in Figure 12. At this a_1^*, $f_1'(a_1^*) = 0$ and $f_1(a_1^*) > 0$. Therefore from (95)

$$\lambda_1 f_1(a_1^*) - \lambda_2 f_2(a_1^*) = -\frac{\partial s_1(a_1^*)}{\partial r} g_1(a_1^*) < 0. \tag{96}$$

Hence we also have $f_2(a_1^*) > 0$. There must then be interior points a_1^* and a_2^* such that f_1 has a global maximum at a_1^* and f_2 has a global maximum at a_2^*. At these points $f_1'(a_1^*) = f_2'(a_2^*) = 0$ and $f_1(a_1^*) > 0, f_2(a_2^*) > 0$. Hence, in addition to (96), we also have the symmetric condition

$$\lambda_2 f_2(a_2^*) - \lambda_1 f_1(a_2^*) = -\frac{\partial s_2(a_2^*)}{\partial r} g_2(a_2^*) < 0. \tag{97}$$

But from (96) and because the f_j attain global maxima at a_j^* we have

$$\lambda_2 f_2(a_2^*) \geq \lambda_2 f_2(a_1^*) > \lambda_1 f_1(a_1^*) \geq \lambda_1 f_1(a_2^*).$$

This contradicts (97).

Case 2: $f_1(0) > 0$**:** In this case, the maximum could, in principle, be at the boundary, i.e. $f_1(0) \geq f_1(a)$ for all a but $f_1'(0) \neq 0$. But we have $s_1(0) = 0$ and hence again (95) implies

(96) with $a_1^* = 0$. The rest of the argument is identical.

This concludes the proof.□

A.6.4 Proof of Proposition 5, Part 4: $S(r)$ is increasing in r

Part 3 immediately implies uniqueness of the stationary equilibrium. First, note that first-order stochastic dominance implies that $S(r)$ is increasing in r: integrating (11) by parts we have

$$S(r) = a_{\max}(r) - \int_0^\infty (G_1(a;r) + G_2(a;r))da$$

which is strictly increasing in r because $G_j(a;r)$ is strictly decreasing in r for all a and $j = 1, 2$ and $a_{\max}(r)$ is strictly increasing in r (because it is the point at which $s_2(a;r)$ intersects zero). Since $S(r)$ is strictly increasing, there can be at most one r solving $S(r) = B$.□

A.7 Proof of Proposition 6

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

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

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.

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 (55) 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:

$$v_j(a_t) = \max_c \; u(c)\Delta + \beta(\Delta)\left(p_j(\Delta)v_j(a_{t+\Delta}) + (1 - p_j(\Delta))v_{-j}(a_{t+\Delta})\right) \quad \text{s.t.} \tag{98}$$

$$a_{t+\Delta} = \Delta(y_j + ra_t - c) + a_t \tag{99}$$

$$a_{t+\Delta} \geq \underline{a} \tag{100}$$

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 (98) 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 (102) and (100). 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 (102) and (100). 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 (100) 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

$$v(a_t, y_t) = \max_c u(c)\Delta + \beta(\Delta)\mathbb{E}[v(a_{t+\Delta}, y_{t+\Delta})] \quad \text{s.t.} \tag{101}$$

$$a_{t+\Delta} = \Delta(y_t + ra_t - c) + a_t \tag{102}$$

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

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

$$\rho v(a_t, y_t) = \max_c \; u(c) + \frac{\mathbb{E}[dv(a_t, y_t)]}{dt} \tag{103}$$

Now assume y_t follows a diffusion process (54) 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 (103) yields (55). 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 (101) 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

$$d\tilde{a}_t = s_j(\tilde{a}_t, t)dt \tag{104}$$

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 (104)

$$\tilde{a}_{t+\Delta} = \tilde{a}_t + \Delta s_j(\tilde{a}_t) \tag{105}$$

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)

$$G_j(a,t) = \Pr(\tilde{a}_t \le a, \tilde{y}_t = y_j). \tag{106}$$

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 (105) but with another (equally correct) discrete-time analogue of (104):[63]

$$\tilde{a}_t = \tilde{a}_{t+\Delta} - \Delta s_j(\tilde{a}_{t+\Delta}) \tag{107}$$

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 $s_j(a) \le 0$ (the case with $s_j(a) > 0$ is symmetric), we have

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

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

$$\begin{aligned}\Pr(\tilde{a}_{t+\Delta} \le a, \tilde{y}_{t+\Delta} = y_j) = (1-\Delta\lambda_j)\Pr(\tilde{a}_t \le a - \Delta s_j(a), \tilde{y}_t = y_j)\\ +\Delta\lambda_{-j}\Pr(\tilde{a}_t \le a - \Delta s_{-j}(a), \tilde{y}_t = y_{-j})\end{aligned} \tag{108}$$

[63]Note that from (104) $\tilde{a}_{t+\Delta} = \int_t^{t+\Delta} s_j(\tilde{a}_\tau)d\tau + \tilde{a}_t$. The integral is approximately equal to both $\Delta s_j(\tilde{a}_t)$ and $\Delta s_j(\tilde{a}_{t+\Delta})$ and therefore both (105) and (107) are meaningful discrete-time analogues. The difference is that the former looks forward in time and the latter looks backward in time.

Using the definition of G_j in (106), 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

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

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

$$\partial_t G_j(a,t) = -s_j(a)\partial_a G_j(a,t) - \lambda_j G_j(a,t) + \lambda_{-j}G_{-j}(a,t), \tag{109}$$

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 (109) 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 positive $s_j(a) < 0$. 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 (56). One relatively accessible derivation is provided by Kredler (2014).

C Weak Solutions of the HJB and KF equations

As stated in Section 1.4, we generally look for certain *weak solutions* of the HJB and KF equations. These are solutions that may not be continuously differentiable or even continuous but still satisfy these equations in some sense. First, Appendix C.1 explains the notion of viscosity solution to an HJB equation and Appendix C.2 that of a measure-valued solution to a KF equation (with appropriate adjustment for the Dirac mass at the boundary).

C.1 Viscosity Solutions to HJB Equations

See http://www.princeton.edu/~moll/viscosity_slides.pdf. In particular, slides 5 to 15 define the notion of viscosity solution and provide some intuition. Slides 20 to 28 discuss problems with state constraints, derive the state constraint boundary condition (10), and make the connection to the notion of constrained viscosity solution of an HJB equation with state constraints. Slides 29 to 41 discuss an important feature of viscosity solutions, namely that an HJB equation typically has a *unique* viscosity solution. This property means that there exists a unique v_j that solves the HJB equation (7) with the state constraint (3). Using another theorem from the literature, one can also show that (as expected) this v_j coincides with the solution to the "sequence problem", the value of maximizing (1) subject to (2), (3) and the process for y_t. Analogous statements are true for (12), (55) and all other HJB equations in the paper.

As noted at several points, most of our paper does not really make use of the powerful theory of viscosity solutions. An important exception is Section 4.3 in which we examine a model in which a non-convexity results in a kinked value function so that its derivative does not exist at the kink points. Importantly, viscosity solutions are *designed* to handle exactly such kinked value functions.

C.2 Measure-Valued Solutions to KF Equations, Extended to Allow for Mass on Boundary

We explain the notion of a measure-valued solution and how to deal with a potential Dirac mass at the borrowing constraint with a simplified version of (7) (or the counterpart with a continuum of income types (56)). 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):

$$0 = -(s(a)g(a))' \quad \text{on } (\underline{a}, \infty), \tag{110}$$

$$\partial_t g(a,t) = -\partial_a (s(a)g(a,t)), \quad \text{on } (\underline{a}, \infty) \times (0, \infty) \quad \text{with } g(a,0) = g_0(a). \tag{111}$$

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 (110) and (111).

C.2.1 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 solutions g. The goal is to obtain a more general equation that also has meaning if this is not the case so that (111) 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 (111), and then integrate by parts, to get[64]

$$\frac{d}{dt}\int_{\underline{a}}^{\infty} \varphi(a)d\mu_t(a) = \int_{\underline{a}}^{\infty} \varphi'(a)s(a)d\mu_t(a). \tag{112}$$

Now comes the key observation: we have shown that the KF equation (111) implies that (112) holds for any test function φ. But the converse is not true, i.e. (112) is a weaker requirement than the KF equation (111). In particular (112) still has an interpretation if the KF equation (111) 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 (111) if the corresponding measure μ_t satisfies (112) 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 (110) if the corresponding measure μ satisfies

$$0 = \int_{\underline{a}}^{\infty} \varphi'(a)s(a)d\mu(a) \tag{113}$$

for all test functions φ. This solution concept is entirely intuitive: after all, the eco-

[64] That is, we follow these steps:

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

.

nomic/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."

C.2.2 Measure-Valued Solution with No Mass on Boundary: Derivation

In fact, (112) 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 (112) 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 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 (112). To derive the KF equation (110) 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 64 in reverse and using 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 (110). The derivation of a measure-valued solution of (110) basically stops half-way through this derivation – at (112) – and therefore does not require the assumption that μ admits a density.

C.2.3 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 (112) and (113) *must also hold* for test functions φ that *do not vanish at* $a = \underline{a}$, i.e. even if $\varphi(\underline{a}), \varphi'(\underline{a}) \neq 0$.

C.2.4 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 2, 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

$$d\mu(a) = g(a)d\mathcal{L}(a) + m\delta_{\underline{a}}, \tag{114}$$

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

$$\int_{\underline{a}}^{\infty} \varphi(a)d\mu(a) = \varphi(\underline{a})m + \int_{\underline{a}}^{\infty} \varphi(a)g(a)d\mathcal{L}(a) \tag{115}$$

for all φ and, in particular, $1 = m + \int_{\underline{a}}^{\infty} g(a)d\mathcal{L}(a)$. Further (113) becomes

$$0 = \int_{\underline{a}}^{\infty} \varphi'(a)s(a)g(a)d\mathcal{L}(a) + \varphi'(\underline{a})s(\underline{a})m. \tag{116}$$

Integrating the first term by parts, we can again see that, in $(\underline{a}, \infty)$, g is a measure-valued solution to (110). The difference is the second term.[65]

Summarizing, whenever we make statements of the sort, "g is a solution to (110)", the precise meaning is that there is a non-negative real-valued function g and a non-negative real-valued scalar m that satisfy (116) for any test function φ, including ones that do not vanish at $a = \underline{a}$.

[65]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 $\varphi(\underline{a})$").

Application 2: Dirac Mass in Interior The model with soft borrowing constraints in Section 2.7 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 9(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 $a = 0$.[66] The analogue of (116) is then

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

C.2.5 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 Section 4.1. The statement "g is a solution to (56)" 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.

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 $X_t \in \bar{\Omega}$.

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

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.

D Accuracy of Finite Difference 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 finite difference scheme explained in Section 3.4 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.

D.1 Finite Differences and Dirac Masses: Theoretical Considerations

Appendix C.2 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 3.4 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 C.2. 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 3.4, 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:

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

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 (115) 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 3.4 and the discrete counterpart of the weak formulation of the KF equation (113). When discretizing the term $\varphi'(a)s(a)$ we upwind it as explained in Section 3.3

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

where $s_i := s(a_i)$. The discrete counterpart to (113) is then

$$
0 = \sum_{i \geq 0} \left(\frac{\varphi_{i+1} - \varphi_i}{\Delta a} s_i^+ + \frac{\varphi_i - \varphi_{i-1}}{\Delta a} s_i^- \right) g_i \Delta a. \tag{117}
$$

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

$$0 = -\frac{s_i^+ g_i - s_{i-1}^+ g_{i-1}}{\Delta a} - \frac{s_{i+1}^- g_{i+1} - s_i^- g_i}{\Delta a}, \qquad i = 1, ..., I. \tag{118}$$

Note that this is exactly the discretization of the KF equation advocated in Section 3.4. In particular, note that (118) 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 3.3.

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 (117) is consistent with the weak formulation of the continuous KF equation (116). 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.

Finite Difference 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 2.7) and as briefly discussed in Section C.2.4 (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

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

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 9 (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.

D.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 finite difference 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 (38) 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 (38) 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

$$s_1(a) = -\sqrt{2\nu_1}\sqrt{a-\underline{a}}, \qquad s_2(a) = -\zeta_2(a - a_{\max}) \tag{119}$$

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

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

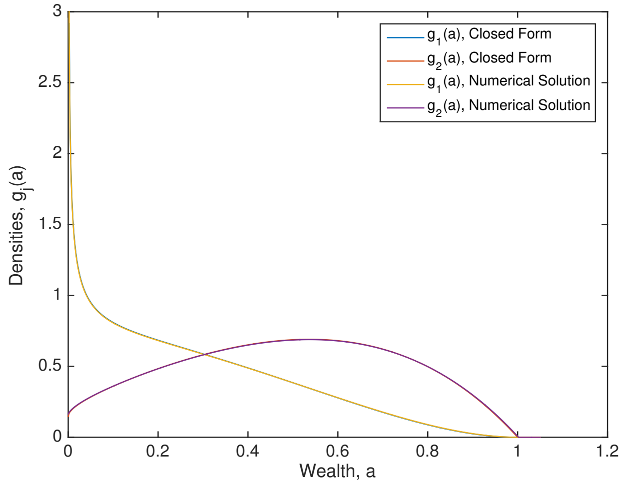
Therefore (38) and the Dirac point mass m_1 defined in Proposition 3 become

$$\begin{aligned} g_1(a) &= \frac{\kappa_1}{-\sqrt{2\nu_1}\sqrt{a-\underline{a}}}\exp\left(\frac{\lambda_1}{\sqrt{\nu_1/2}}\sqrt{a-\underline{a}}\right)\left(\frac{a_{\max}-a}{a_{\max}-\underline{a}}\right)^{\lambda_2/\zeta_2} \\ g_2(a) &= \frac{\kappa_2}{-\zeta_2(a-a_{\max})}\exp\left(\frac{\lambda_1}{\sqrt{\nu_1/2}}\sqrt{a-\underline{a}}\right)\left(\frac{a_{\max}-a}{a_{\max}-\underline{a}}\right)^{\lambda_2/\zeta_2} \\ m_1 &= \frac{\lambda_2}{\lambda_1+\lambda_2}\tilde{m}_1, \quad \frac{1}{\tilde{m}_1} = \frac{\lambda_2}{\zeta_2}(a_{\max}-\underline{a})^{-\lambda_2/\zeta_2}\int_{\underline{a}}^{a_{\max}}\exp\left(\frac{\lambda_1}{\sqrt{\nu_1/2}}\sqrt{a-\underline{a}}\right)(a_{\max}-a)^{\lambda_2/\zeta_2-1}da. \end{aligned} \tag{120}$$

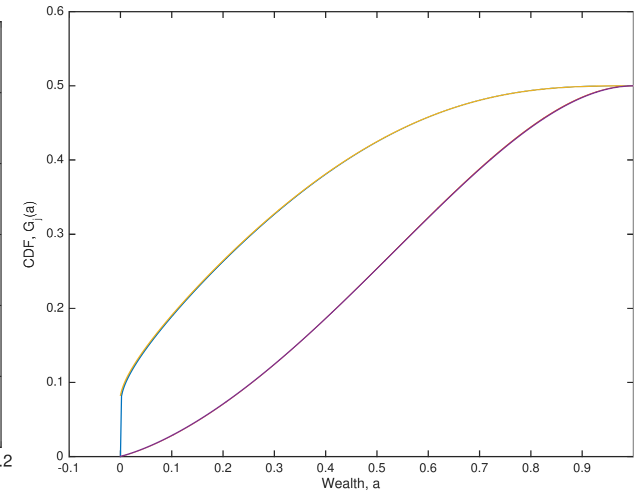
Finally, as explained in Appendix A.4.5, we have $\kappa_1 = -\lambda_1 m_1$ and $\kappa_2 = \lambda_2 m_2$.

Figure 13 plots this solution and compares it to the numerical solution computed using the algorithm laid out in Section 3.4 with $I = 500$ wealth grid points (of course, also assuming that s_1 and s_2 are given by (119)). 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).[67]

[67] To obtain the CDFs corresponding to the closed-form solution (120), we numerically integrate g_1 and

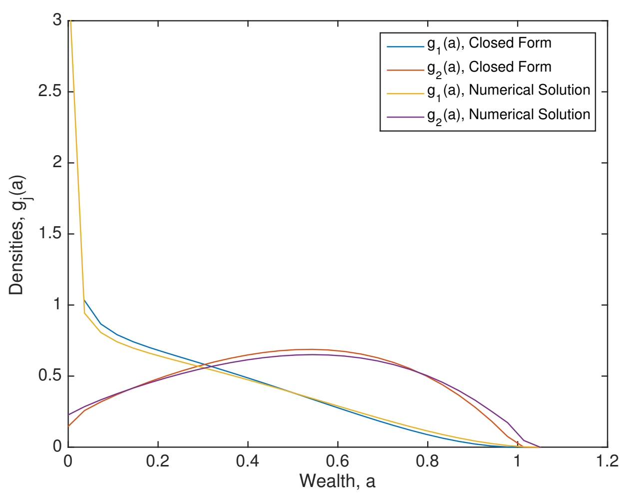


(a) Densities

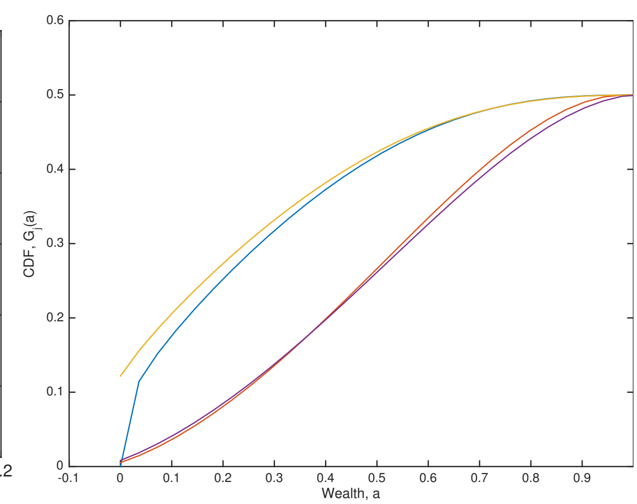


(b) Cumulative Distribution

Figure 13: A fine grid with $I = 500$ points results in the finite difference scheme being highly accurate



(a) Densities



(b) Cumulative Distribution

Figure 14: A coarse grid with only $I = 30$ points results in the finite difference scheme being relatively inaccurate

Figure 14 repeats the exercise but with only $I = 30$ wealth grid points. With this much low 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 http://www.princeton.edu/~moll/HACTproject/HACT_Numerical_Appendix.pdf for a discussion on how to do this.

The code used to generate Figures 13 and 14 are available online at http://www.princeton.edu/~moll/HACTproject/KFE_accuracy_check.m. The interested reader can try out herself how varying the number of grid points and parameter values affects the accuracy of the numerical solution.

E Appendices for Section 4

As in the proof of Proposition 5 in Appendix A.6, we work with a general income process from the get-go and write the HJB equation as (87) where $\mathcal{A}$ is the infinitesimal generator ("infinite-dimensional transition matrix") of the stochastic process for y_t. The HJB equation with a diffusion process (55) is the special case with $\mathcal{A}v = \mu(y)\partial_y v + \frac{\sigma^2(y)}{2}\partial_{yy}v$. But the notation also allows for other stationary processes on $[\underline{y}, \bar{y}]$.

E.1 Generalization of Results from Section 2 to Other Income Processes

As stated in the text, Propositions 1, 2, 4 and 5 generalize to the other income processes.

Proposition 7 (Generalization of Proposition 1 to Other Income Processes) *Assume that $r < \rho$ and that Assumption 1 holds with y_1 replaced by $\underline{y}$. Then the solution to the HJB equation (87) 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*

g_2 in (120). In this regard, a difficulty is that $g_1(a) \to \infty$ as $a \to \underline{a}$. We therefore compute the cumulative 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.

consume satisfy

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

where $\underline{c}(y) = c(\underline{a},y)$ is consumption at the borrowing constraint. For the special case with a diffusion process (54) so that $\mathcal{A}v = \mu(y)\partial_y v + \frac{\sigma^2(y)}{2}\partial_{yy}v$:

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

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.

Proof of Proposition 7: Analogous to Proposition 1, we start by differentiating (55) with respect to a (envelope condition) and using the FOC $u'(c(a,y)) = \partial_a v(a,y)$ to obtain the "Euler equation"

$$
(\rho-r)u'(c) = u''(c)(\partial_a c)s + \mathcal{A}u'(c) \tag{121}
$$

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 (121) 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 $y \leq y^*$ 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 $s(a,y) \sim -\sqrt{2\nu(y)}\sqrt{a - \underline{a}}$.□

Proposition 8 (Generalization of Proposition 2 to Other Income Processes) *Consider the HJB equation (87) 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*

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

The proof follows identical steps as in the proof of Proposition 2 and is available upon request.

We do not state the extended Propositions 4 and 5 here because the wording is unchanged. The proof of Proposition 4 is available upon request. The proof of Proposition 5 in Appendix 5 already covered the case of a general income process.

E.2 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.[68] 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

[68] 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.

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.[69] 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

$$dk_t + db_t = (z_t + \tilde{R}_t k_t + r_t b_t - c_t)dt \tag{122}$$

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

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

where R_1 and R_2 are parameters 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 R_t is the return from owning and running a private firm.[70] 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 $a_t = b_t + k_t$:

$$da_t = (z_t + ra_t + (R - r)k_t - c_t)dt + \sigma k_t dW_t \tag{123}$$

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

[70] 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

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

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.

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

$$k_t \leq a_t + \phi \tag{124}$$

Individuals maximize (1) subject to (123), (124) 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

$$\begin{aligned}\rho v_j(a) = \max_{c, 0 \leq k \leq a+\phi} \; & u(c) + v_j'(a)(z_j + ra + (R-r)k - c) \\ & + \frac{1}{2} v_j''(a)\sigma^2 k^2 + \lambda_j (v_{-j}(a) - v_j(a))\end{aligned} \tag{125}$$

$$0 = -\frac{d}{da}[s_j(a) g_j(a)] + \frac{1}{2}\frac{d^2}{da^2}[\sigma^2 k_j(a)^2 g_j(a)] - \lambda_j g_j(a) + \lambda_{-j} g_{-j}(a). \tag{126}$$

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 (125) 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 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:

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

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*

$$c_j(a) \sim \left(\frac{\rho - (1-\gamma) r}{\gamma} - \frac{1-\gamma}{2\gamma} \frac{(R-r)^2}{\gamma \sigma^2} \right) a \tag{127}$$

$$s_j(a) \sim \left(\frac{r - \rho}{\gamma} + \frac{1+\gamma}{2\gamma} \frac{(R-r)^2}{\gamma \sigma^2} \right) a \tag{128}$$

$$k_j(a) \sim \frac{R-r}{\gamma \sigma^2} a. \tag{129}$$

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*

$$\zeta = \gamma \left(\frac{2\sigma^2(\rho - r)}{(R-r)^2} - 1 \right). \tag{130}$$

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 (129)) which is a force towards lower top wealth inequality. Formula (130) 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 (129) we have

$$\sigma k_j(a) \sim \frac{R-r}{\gamma\sigma} a \tag{131}$$

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,

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

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 (130) 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.

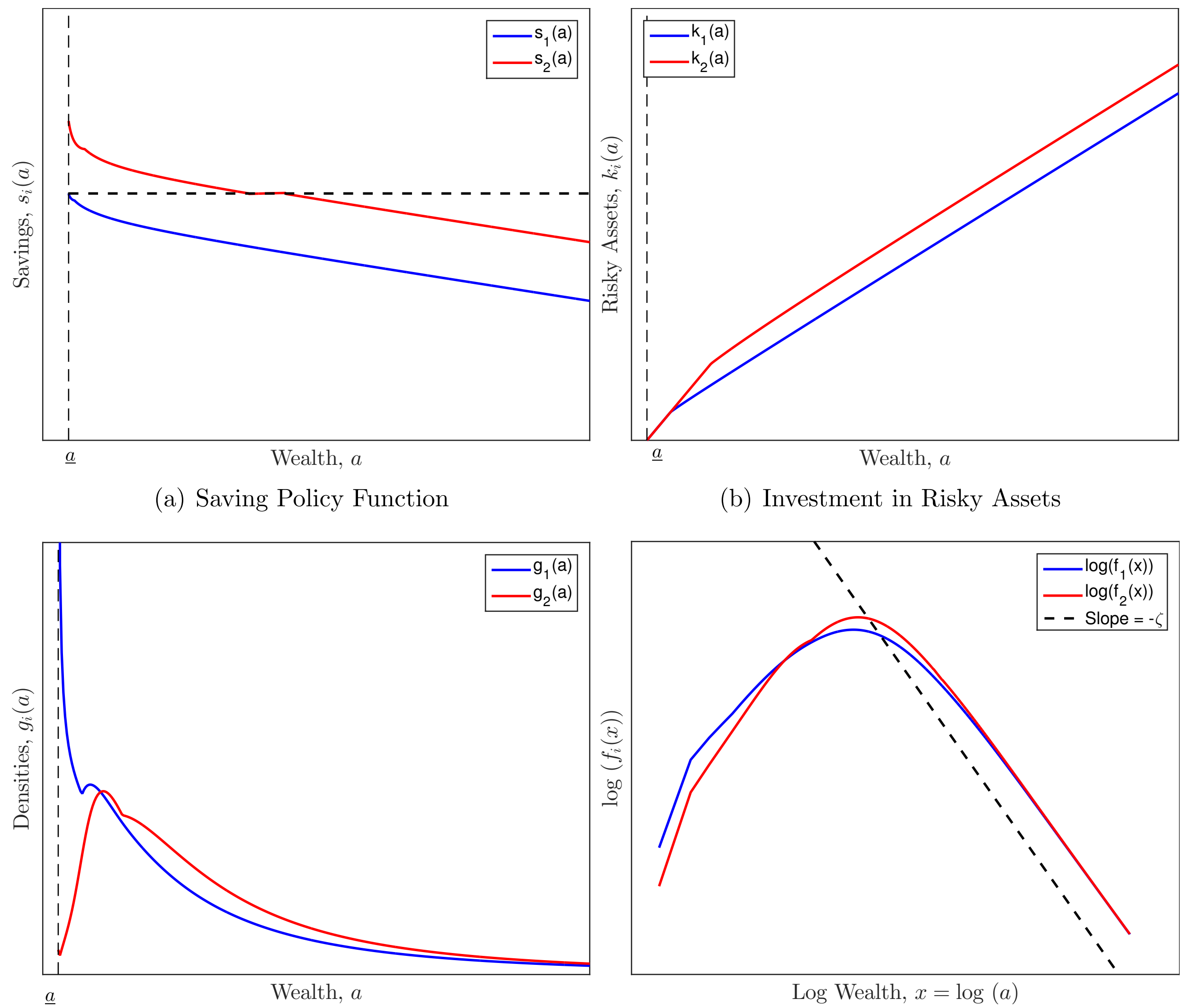
Figure 15 plots individuals' optimal choices and the resulting wealth distribution for the model with both a risky and a riskless asset. Panel (d) in particular shows that the distribution behaves asymptotically like a Pareto distribution by showing that the logarithm of the density of log wealth $f_i(x)$ is asymptotically linear in the logarithm of wealth $x = \log(a)$.[71] Table 1 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 distribution of the United States quite well, particularly at the top.[72]

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 (122) to

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

[71] 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.

[72] 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.



(a) Saving Policy Function

(b) Investment in Risky Assets

(c) Stationary Densities

(d) Pareto Tail of Stationary Distribution

Figure 15: Optimal Choices and Pareto Tail of Wealth Distribution in Two-Asset Model

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 (130) 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 (131). 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 (131).

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 (125) satisfies a certain homogeneity property.

Lemma 8 *Consider the problem*

$$\rho v(a) = \max_{c,k} \; u(c) + v'(a)(ra + (R-r)k - c) + \frac{1}{2}v''(a)\sigma^2 k^2 \tag{132}$$

where $u(c) = c^{1-\gamma}/(1-\gamma), \gamma > 0$. The optimal policy functions that solve (132) are linear in wealth and given by

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

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

$$k(a) = \frac{R-r}{\gamma\sigma^2}a \tag{135}$$

Proof of Lemma 8 Grouping terms by the relevant maximization problems and solving these, we can write

$$\begin{aligned} \rho v(a) &= H(v'(a)) + G(v'(a), v''(a)) + v'(a)ra \\ H(p) &= \max_c \{u(c) - pc\} = \frac{\gamma}{1-\gamma}p^{\frac{\gamma-1}{\gamma}} \\ G(p,q) &= \max_k \left\{p(R-r)k + \frac{1}{2}q\sigma^2k^2\right\} = \frac{1}{2}\frac{p^2}{-q}\frac{(R-r)^2}{\sigma^2} \end{aligned} \tag{136}$$

and from the first-order conditions

$$u'(c(a)) = v'(a), \quad k(a) = -\frac{v'(a)}{v''(a)}\frac{R-r}{\sigma^2} \tag{137}$$

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

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

Substituting into (136) and dividing by $Ba^{1-\gamma}$, we have

$$\rho = \gamma((1-\gamma)B)^{-\frac{1}{\gamma}} + \frac{1}{2}\frac{(R-r)^2}{\sigma^2}\frac{1-\gamma}{\gamma} + (1-\gamma)r. \tag{138}$$

From (137) $c(a) = ((1-\gamma)B)^{-\frac{1}{\gamma}} a$ and hence using (138) we obtain (133), (134) and (135).□

Lemma 9 *Consider the problem* (125). *For any* $\xi > 0$,

$$v_j(\xi a) = \xi^{1-\gamma} v_{\xi,j}(a) \tag{139}$$

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

$$\begin{aligned} \rho v_{\xi,j}(a) = \max_{c, 0 \le k \le a + \phi/\xi} \quad & u(c) + v_j'(a)(y_j/\xi + ra + (R-r)k - c) \\ & + \frac{1}{2} v_{\xi,j}''(a)\sigma^2 k^2 + \lambda_j(v_{\xi,-j}(a) - v_{\xi,j}(a)) \end{aligned} \tag{140}$$

Proof of Lemma 9 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 (130). Adding the two Kolmogorov Forward (Fokker-Planck) equations (126)

$$0 = -\frac{d}{da}[s_1(a)g_1(a) + s_2(a)g_2(a)] + \frac{\sigma^2}{2}\frac{d^2}{da^2}[k_1(a)^2 g_1(a) + k_2(a)^2 g_2(a)]. \tag{141}$$

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

$$\bar{s} = \frac{r-\rho}{\gamma} + \frac{1+\gamma}{2\gamma}\frac{(R-r)^2}{\gamma\sigma^2}, \quad \bar{k} = \frac{R-r}{\gamma\sigma^2} \tag{142}$$

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 (141) for large enough a (all other terms go to zero as $a \to \infty$). We here present a more rigorous and constructive proof. Integrating (141)

$$\frac{\sigma^2}{2}\frac{d}{da}[k_1(a)^2 g_1(a) + k_2(a)^2 g_2(a)] = [s_1(a)g_1(a) + s_2(a)g_2(a)] + C. \tag{143}$$

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 (143) as

$$y_1'(a) + y_2'(a) = \frac{2s_1(a)}{\sigma^2 k_1(a)^2} y_1(a) + \frac{2s_2(a)}{\sigma^2 k_2(a)^2} y_2(a). \tag{144}$$

Define $y(a) = y_1(a) + y_2(a)$. After collecting the leading term, (144) is written as

$$\begin{aligned} y'(a) &= \frac{\theta}{a} y(a) + h_1(a)y_1(a) + h_2(a)y_2(a), \\ \theta &= \frac{2\bar{s}}{\sigma^2 \bar{k}^2}, \quad h_j(a) = \frac{2}{\sigma^2}\left(\frac{\tilde{s}_j + \bar{s}a}{(\tilde{k}_j + \bar{k}a)^2} - \frac{\bar{s}}{\bar{k}^2 a}\right), \ j = 1, 2. \end{aligned} \tag{145}$$

Dividing (145) by $y(a)$ and integrating both sides from a_1 to a_2 where $a_1 < a_2$ are large enough, we have

$$\ln\left(\frac{y(a_2)}{a_2^\theta}\right) - \ln\left(\frac{y(a_1)}{a_1^\theta}\right) = \int_{a_1}^{a_2} \frac{h_1(x)y_1(x)}{y(x)} dx + \int_{a_1}^{a_2} \frac{h_2(x)y_2(x)}{y(x)} dx. \tag{146}$$

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

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

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

$$\lim_{a\to\infty} \frac{g(a)}{a^{\theta-2}} = \frac{2\exp(\bar{\xi})}{\sigma^2 \bar{k}^2}. \tag{147}$$

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 (142) into the expression for ζ yields (130).□

Appendix References

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

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.

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.

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," Working papers, University of Southern California.