Redistribution and Social Insurance:
Online Appendix
Mikhail Golosov Maxim Troshkin Aleh Tsyvinski

A Proofs and Additional Details

A.1 Proof of Lemma 1

Given any solution \( u^*(\theta) \), following a sequence of reports \( (\theta^i, \hat{\theta}) \), to maximization problem (6) and (11), we can construct

\[
\omega \left( \hat{\theta}|\theta \right) = \int_0^{\infty} u^*(\theta^i, \hat{\theta}, s) f_{t+1}(s|\theta) \, ds.
\]

We can re-write (5) as

\[
\max_{\hat{\theta}} \mathcal{V} \left( \hat{\theta}; \theta \right) \equiv \max_{\hat{\theta}} U \left( c(\hat{\theta}), y(\hat{\theta}); \theta \right) + \beta \omega(\hat{\theta}|\theta).
\]

Since \( c(\cdot) \) and \( \omega(\cdot|\theta) \) are piecewise \( C^1 \), they are differentiable except at a finite number of points. Then for all \( \theta \) where they are differentiable,

\[
U_c \left( c(\theta), y(\theta); \theta \right) \dot{c} \left( \theta \right) + U_y \left( c(\theta), y(\theta); \theta \right) \dot{y} \left( \theta \right) + \beta \omega_1(\theta|\theta) = 0, \quad (25)
\]

where \( \dot{c} \) and \( \dot{y} \) are derivatives of \( c \) and \( y \). Optimality requires that \( y(\cdot) \) and \( \mathcal{V}(\cdot; \theta) \) are piecewise \( C^1 \) and \( c(\cdot) \) and \( \omega(\cdot|\theta) \) are.

Suppose that the global incentive constraint is violated, i.e. \( \mathcal{V} \left( \hat{\theta}; \theta \right) -
\( \mathcal{V}(\theta; \theta) > 0 \) for some \( \hat{\theta} \). Suppose \( \hat{\theta} > \theta \) is a point of differentiability. Then

\[
0 < \int_{\theta}^{\hat{\theta}} \frac{\partial \mathcal{V}(x; \theta)}{\partial x} \, dx = \int_{\theta}^{\hat{\theta}} \left[ U_c(x; \theta) \dot{c}(x) + U_y(x; \theta) \dot{y}(x) + \beta \frac{d\omega(x|\theta)}{dx} \right] \, dx.
\]

Since all of the objects under the integral are piecewise differentiable, it can be represented as a finite sum of the terms

\[
\int_{\theta_j}^{\theta_{j+1}} \left[ \dot{c}(x) + \dot{y}(x) \frac{U_y(x; \theta)}{U_c(x; \theta)} + \beta \frac{\omega_1(x|\theta)}{U_c(x; \theta)} \right] \, dx
\]

for some finite number of intervals \((\theta_j, \theta_{j+1})\).

If \( x > \theta \), \( \frac{U_y(x; \theta)}{U_c(x; \theta)} \leq \frac{U_y(x; x)}{U_c(x; x)} \) and \( U_c(x; \theta) \geq U_c(x; x) \) (from the single crossing property in Assumption 2 and \( U_{cl} \geq 0 \) in Assumption 3) and \( \omega_1(x|x) \geq \omega_1(x|\theta) \) from Assumption 3. Therefore

\[
\int_{\theta_j}^{\theta_{j+1}} \left[ \dot{c}(x) + \dot{y}(x) \frac{U_y(x; \theta)}{U_c(x; \theta)} + \beta \frac{\omega_1(x|x)}{U_c(x; x)} \right] \, dx \\
\leq \int_{\theta_j}^{\theta_{j+1}} \left[ \dot{c}(x) + \dot{y}(x) \frac{U_y(x; x)}{U_c(x; x)} + \beta \frac{\omega_1(x|x)}{U_c(x; x)} \right] \, dx \\
= 0
\]

where the last equality follows from (25). Therefore, \( \int_{\theta}^{\hat{\theta}} \frac{\partial \mathcal{V}(x; \theta)}{\partial x} \, dx \leq 0 \), a contradiction. If \( \hat{\theta} < \theta \) the arguments are analogous. Finally, since \( \mathcal{V}(\hat{\theta}; \theta) \) is continuous in \( \hat{\theta} \), taking limits establishes that \( \mathcal{V}(\hat{\theta}; \theta) \leq \mathcal{V}(\theta; \theta) \) at the points of non-differentiability.
A.2 Decomposition in equation (17)

We omit explicit time subscripts $t$ whenever it does not lead to confusion. The Hamiltonian to problem (6) and (11) is

$$H = (c - \theta l + R^{-1}V_{t+1}(w, w_2, \theta)) f_t + \psi \left[ -U_l(c, l) \frac{l}{\theta} + \beta w_2 \right]$$

$$-\lambda_1 \bar{\alpha}_t u(\theta) f_t + \lambda_2 u(\theta) f_{2,t} + \varphi [u - U(c, l) - \beta w],$$

where $f_{2,t} = 0$ if $t = 0$. The envelope conditions are

$$\frac{\partial V_t}{\partial w} = \lambda_1, \frac{\partial V_t}{\partial w_2} = -\lambda_2. \quad (26)$$

The first-order conditions are

$$\varphi - \lambda_1 \bar{\alpha}_t f + \lambda_2 f_2 = -\dot{\psi} \quad (27)$$

$$-U_l \varphi - \theta f = -\frac{1}{\theta} \psi \left[ \frac{U_\ell l + U_l}{U_l} \right] (-U_l) \quad (28)$$

$$f - \psi U_\alpha \frac{l}{\theta} = \varphi U_c \quad (29)$$

$$\frac{1}{R} \frac{\partial V_{t+1}}{\partial w} f = \varphi \beta \quad (30)$$

$$\frac{1}{R} \frac{\partial V_{t+1}}{\partial w_2} f = -\psi \beta$$
Use (29) to substitute away for $\varphi$

\[
\frac{1}{U_c} \dot{f} - \lambda_1 \dot{\alpha} t f + \lambda_2 f_2 - \psi \frac{U_{dl}}{\theta U_c} = -\psi \\
- \frac{U_t}{U_c} f - \psi \frac{U_{dl} (-U_t)}{U_c} - \theta f = -\frac{1}{\theta} \psi \left[ \frac{U_{dl} + U_t}{U_t} \right] (-U_t) \\
\frac{1}{\beta R} \frac{\partial V_{t+1}}{\partial w} = \frac{1}{U_c} - \frac{\psi U_{dl}}{\theta f U_c} \\
\frac{1}{\beta R} \frac{\partial V_{t+1}}{\partial w_2} = -\frac{\psi}{f} \\
\tag{31}
\tag{32}
\tag{33}
\tag{34}
\]

Use definitions of $\varepsilon, \gamma$ to write (32) as

\[
\left( \frac{U_i}{\theta U_c} + 1 \right) \theta f = \frac{1}{\theta} \psi (1 + \varepsilon - \gamma) (-U_i) .
\]

Since $\tau^y = 1 + \frac{U_t}{\theta U_c}$ this can be equivalently written as

\[
\frac{\tau^y}{1 - \tau^y} = \psi \frac{U_c}{\theta f} (1 + \varepsilon - \gamma).
\]

This expression together with (34) implies

\[
\lambda_{2,t+1} = - \frac{\partial V_{t+1}}{\partial w_2} = \beta R \frac{\tau^y (\theta)}{1 - \tau^y (\theta) U_{c,t} (\theta)} \frac{\theta}{(1 + \varepsilon_t (\theta) - \gamma_t (\theta))^{-1}}. \\
\tag{35}
\tag{36}
\]

To find $\psi$ we integrate (31)

\[
\psi (\theta) = \int_{\theta}^{\infty} \exp \left( - \int_{\theta}^{x} \gamma (\tilde{x}) \frac{d\tilde{x}}{\tilde{x}} \right) \left( \frac{1}{U_c (x)} f (x) - \lambda_1 \dot{\alpha} t f (x) - \lambda_2 f_2 (x) \right) dx. \\
\]

iv
From boundary condition $\psi(0) = 0$ we get

$$\lambda_{1,t} = \frac{\int_0^\infty \exp\left( - \int_0^x \gamma_t(\bar{x}) \frac{d\bar{x}}{\bar{x}} \right) \left( \frac{1}{U_{c,t}(x)} f_t(x) + \lambda_{2,t} f_{2,t}(x) \right) dx}{\int_0^\infty \exp\left( - \int_0^x \gamma_t(\bar{x}) \frac{d\bar{x}}{\bar{x}} \right) \tilde{\alpha}_t(x) f_t(x) dx} \quad (37)$$

and $\lambda_{2,t}$ is given by (36). If $U$ is separable, then $\gamma = 0$ and from our assumption on Pareto weights that implies that $\int_0^\infty \tilde{\alpha}_t(x) f_t(x) dx = 1$ for all $t$, we get $\lambda_{1,t} = \int_0^\infty f_t(x) dx$ for all $t$.

Use the expression for $\psi(\theta)$ and (36) for $t - 1$ to substitute into (35):

$$\frac{\tau_t^y(\theta)}{1 - \tau_t^y(\theta)} = (1 + \varepsilon_t(\theta) - \gamma_t(\theta)) \frac{1}{\theta f_t(\theta)} \int_0^\infty \frac{U_{c,t}(\theta t)}{U_{c,t}(x)} \exp\left( - \int_0^x \gamma_t(\bar{x}) \frac{d\bar{x}}{\bar{x}} \right) (1 - \lambda_{1,t} \tilde{\alpha}_t U_{c,t}(x)) f_t(x) dx$$

$$+ \beta R \frac{\tau_{t-1}^y}{1 - \tau_{t-1}^y} \frac{1 + \varepsilon_t(\theta) - \gamma_t(\theta)}{1 + \varepsilon_{t-1} - \gamma_{t-1}} \frac{U_{c,t}(\theta t)}{U_{c,t-1}(x)} \frac{\theta f_t(\theta)}{\theta f_{t-1}(\theta)} \int_0^\infty \exp\left( - \int_0^x \gamma_t(\bar{x}) \frac{d\bar{x}}{\bar{x}} \right) f_{2,t}(x) dx.$$

Finally note that

$$\frac{U_{c,t}(\theta_t)}{U_{c,t}(x)} \exp\left( - \int_0^x \gamma_t(\bar{x}) \frac{d\bar{x}}{\bar{x}} \right) = \exp\left( \ln \frac{U_{c,t}(\theta_t)}{U_{c,t}(x)} - \int_0^x \gamma_t(\bar{x}) \frac{d\bar{x}}{\bar{x}} \right)$$

$$= \exp\left( - \int_0^x \frac{dU_{c,t}(\bar{x})}{U_{c,t}(\bar{x})} - \int_0^x \gamma_t(\bar{x}) \frac{d\bar{x}}{\bar{x}} \right)$$

$$= \exp\left( \int_0^x \left( \sigma_t(\bar{x}) \frac{\sigma_t(\bar{x})}{\sigma_t(\bar{x})} - \gamma_t(\bar{x}) \frac{\nu_t(\bar{x})}{\nu_t(\bar{x})} \right) d\bar{x} \right)$$

which is the same expression as (17) in the general, non-separable case.
A.3 Proofs of Proposition 1, Corollary 1, equation (22)

A.3.1 Preliminary results

We first prove some preliminary results about the speed of convergence of $c_t(\theta), y_t(\theta),$ and $l_t(\theta)$, provided that limits exist, distortions remain finite, and elasticities are bounded. These arguments are the same for both separable and non-separable preferences, so we present them for the general case.

Let $U$ be a utility function that satisfies Assumption 2, let $\sigma, \varepsilon$ be as defined in (15) and $\gamma \equiv \frac{U_{y}\bar{y}}{U_{c}}$. Preferences are separable if $\gamma = 0$ for all $(c, l)$. Preferences are GHH if $U(c, l) = \frac{1}{1-\nu} \left( c - \frac{1}{1+1/\zeta} \right)^{1-\nu}$ for $\zeta, \nu > 0$.

We use notation $x_t(\theta)$ to represent the optimal value of variable $x_t(\theta^{t-1} \theta)$ for a given $\theta^{t-1}$. We make the following assumption.

**Assumption 6.** $\varepsilon_t(\theta), \sigma_t(\theta), \gamma_t(\theta), \frac{\alpha_t(\theta)}{y_t(\theta)}$ have finite, non-zero limits; $\frac{\tau_t^y(\theta)}{1-\tau_t^x(\theta)}$ has a finite limit; $\frac{\dot{c}_t(\theta)/c_t(\theta)}{y_t(\theta)/y_t(\theta)}$ has a limit as $\theta \to \infty$. $\alpha(\theta)$ is bounded and $\alpha(\theta), \varepsilon_t(\theta), U_{c,t}(\theta)$ have finite limits as $\theta \to 0$.

Note in particular that when preferences are separable, then Assumption 4 implies Assumption 6. Let $\sigma_t(\theta) \to \bar{\sigma}, \gamma_t(\theta) \to \bar{\gamma}, \varepsilon_t(\theta) \to \bar{\varepsilon}$, $\tau_t(\theta) / (1 - \tau_t(\theta)) \to \bar{\tau} / (1 - \bar{\tau})$ for some $\bar{\sigma}, \bar{\gamma}, \bar{\varepsilon}, \bar{\tau} / (1 - \bar{\tau})$, and let $X_t(\theta) \equiv \frac{\alpha_t(\theta)}{(1-\tau_t(\theta))\theta l_t(\theta)} \to \bar{X}$ as $\theta \to \infty$. If Assumption 6 is satisfied, these limits are well defined, finite and, with the exception of $\bar{\gamma}$ and $\bar{\tau} / (1 - \bar{\tau})$, are non-zero.

**Lemma 2.** Suppose that Assumption 6 is satisfied. If $\lim_{\theta \to \infty} \theta l_t / l_t$ and $\lim_{\theta \to \infty} \theta c_t/c_t$
are finite, then
\[
\lim_{\theta \to \infty} \frac{\dot{l}_t}{l_t} = \frac{1 - \sigma + \bar{\gamma} X}{\bar{\sigma} + \bar{\varepsilon} - \bar{\gamma} (X + 1)}, \quad \lim_{\theta \to \infty} \frac{\dot{y}_t}{y_t} = \lim_{\theta \to \infty} \frac{\dot{c}_t}{c_t} = \frac{1 + \bar{\varepsilon} - \bar{\gamma}}{\sigma + \bar{\varepsilon} - \gamma (X + 1)}.
\]
(39)

If \( U \) is separable or GHH, then these limits exist and finite. In separable case, \( \bar{\sigma}, \bar{\varepsilon}, \bar{\gamma} \) generically depend only on \( U : \bar{\gamma} = 0, \bar{\sigma} = \lim_{c \to \infty} -\frac{U_{cc}}{U_c}, \bar{\varepsilon} = \lim_{l \to \infty} \frac{U_{ll}}{U_l} \) if \( \bar{\sigma} < 1, \bar{\varepsilon} = \lim_{l \to 0} \frac{U_{ll}}{U_l} \) if \( \bar{\sigma} > 1 \). In GHH case, \( \lim_{\theta \to \infty} \frac{\dot{l}_t}{l_t} = \xi, \lim_{\theta \to \infty} \frac{\dot{y}_t}{y_t} = \lim_{\theta \to \infty} \frac{\dot{c}_t}{c_t} = 1 + \zeta, \).

Proof. Since \( \frac{\dot{c}_t}{y_t} = \frac{\dot{c}_t}{y_t} \) and the limit of the right hand side exists as \( \theta \to \infty \), \( \lim_{\theta \to \infty} \frac{\dot{c}_t}{y_t} \) exists. We must have \( c_t(\theta), y_t(\theta) \to \infty \) as \( \theta \to \infty \), otherwise
\[
1 - \tau_t^y(\theta) = \frac{-U_{1,t}(\theta)}{\theta U_{1,t}(\theta)} \to 0,
\]
contradicting the assumption that \( \lim_{\theta \to \infty} \frac{\tau_t^y(\theta)}{1 - \tau_t^y(\theta)} < \infty \). Therefore the L'Hospital's rule implies
\[
\lim_{\theta \to \infty} \frac{c_t(\theta)}{y_t(\theta)} = \lim_{\theta \to \infty} \frac{\dot{c}_t(\theta) / c_t(\theta) \cdot c_t(\theta)}{\dot{y}_t(\theta) / y_t(\theta) \cdot y_t(\theta)}
\]
or
\[
1 = \lim_{\theta \to \infty} \frac{\dot{c}_t \theta}{y_t \theta} = \lim_{\theta \to \infty} \frac{\dot{c}_t \theta}{y_t \theta}.
\]
(40)

Since \( \bar{\tau} < 1 \), applying L'Hospital's rule,
\[
1 = \frac{\lim_{\theta \to \infty} \frac{U_{1,t}(\theta)}{\theta U_{1,t}(\theta)}}{1 - \bar{\tau}} = \lim_{\theta \to \infty} \frac{\varepsilon_t(\theta) \frac{\dot{l}_t}{l_t} - \gamma_t(\theta) X_t(\theta) \frac{\dot{c}_t}{c_t} \theta}{1 - \sigma_t(\theta) \frac{\dot{c}_t}{c_t} \theta + \gamma_t(\theta) \frac{\dot{l}_t}{l_t} \theta}.
\]
(41)

When \( \lim_{\theta \to \infty} \theta \dot{l}_t / l_t, \lim_{\theta \to \infty} \theta \dot{c}_t / c_t \) are finite, we can use (40) and (41) to get (39).

We verify that \( \lim_{\theta \to \infty} \theta \dot{l}_t / l_t, \lim_{\theta \to \infty} \theta \dot{c}_t / c_t \) are finite when preferences are
Suppose that Assumption 6 is satisfied. Then \( c_t (\theta) = o \left( \theta^k \right) \) \((\theta \to \infty)\) for any \( \hat{k} > \frac{1 + \hat{\varepsilon} - \hat{\gamma}}{\hat{\sigma} + \hat{\varepsilon} - \hat{\gamma} (X+1)} \) and there exists \( \kappa > 0 \) such that \( U_{c,t} = o \left( \theta^{-\kappa} \right) \ (\theta \to \infty) \). If preferences are separable, this holds for any \( \kappa < \frac{(1 + \varepsilon) \theta}{\sigma + \varepsilon} \).

Proof. We first show that for any \( \hat{k} > \frac{1 + \hat{\varepsilon} - \hat{\gamma}}{\hat{\sigma} + \hat{\varepsilon} - \hat{\gamma} (X+1)} \) there exist \( \hat{K}, \hat{\theta} \) such that \( c_t (\theta) \leq \hat{K} \theta^{\hat{k}} \) for all \( \theta \geq \hat{\theta} \). By Lemma 2 for any \( \hat{k} > \frac{1 + \hat{\varepsilon} - \hat{\gamma}}{\hat{\sigma} + \hat{\varepsilon} - \hat{\gamma} (X+1)} \) we can pick \( \hat{\theta} \) such that \( \theta \hat{c}_t / c_t < \hat{k} \) for all \( \theta \geq \hat{\theta} \). Let \( \hat{K} = c_t (\hat{\theta}) / \theta^{\hat{k}} \). Consider a function \( G (\theta) \equiv \hat{K} \theta^{\hat{k}} - c_t (\theta) \), which is continuous for \( \theta \geq \hat{\theta} \) with \( G (\hat{\theta}) = 0 \). For any \( \theta > \hat{\theta} \) we have

\[
G (\theta) = \int_{\hat{\theta}}^{\theta} G' (x) \, dx = \int_{\hat{\theta}}^{\theta} \left[ \hat{K} \hat{k} x^{\hat{k}} - \frac{\hat{c}_t (x) x}{c_t (x) c_t (x)} \right] \frac{dx}{x}.
\]

If \( G (\theta) = 0 \) for some \( \theta \geq \hat{\theta} \), then \( G' (\theta) = \left[ \hat{K} \hat{k} \theta^{\hat{k}} - \frac{\hat{c}_t (\theta) \theta}{c_t (\theta) c_t (\theta)} \right] \frac{1}{\theta} > \frac{1}{\theta} \frac{\hat{c}_t (\theta) \theta}{c_t (\theta) c_t (\theta)} G (\theta) = 0 \). Since \( G (\hat{\theta}) = 0 \), this implies that for all \( \theta \geq \hat{\theta} \), \( G (\theta) \) never crosses zero.
from above and is weakly positive. This establishes that \( c_t(\theta) = O\left(\theta^k\right) \).

Since \( c_t(\theta) = O\left(\theta^k\right) \) for any \( k \in \left(\frac{1+\varepsilon-\gamma}{\sigma+\varepsilon-\gamma(X+1)}, \hat{k}\right) \) and \( \theta^k = o\left(\theta^k\right) \) it also implies that \( c_t(\theta) = o\left(\theta^k\right) \).

If preferences are separable, we can use the same arguments to show that 
\( U_c(c) = o\left(c^{-\hat{k}}\right) \) for any \( \hat{k} < \bar{\sigma} \). We then define \( \kappa = \hat{k}\hat{k} \) to show that \( U_{c,t} = o\left(\theta^{-\kappa}\right) \) for any \( \kappa < \frac{(1+\varepsilon)\bar{\theta}}{\sigma+\varepsilon} \). For all other preferences

\[
U_c(c_t(\theta), l_t(\theta)) - U_c(c_t(\bar{\theta}), l_t(\bar{\theta})) = \int_\theta^\bar{\theta} \left[ -\sigma_t(x) \frac{\hat{c}_t(x)x}{c_t(x)} + \gamma_t(x) \frac{\hat{l}_t(x)x}{l_t(x)} \right] \frac{U_{c,t}(x)dx}{x}
\]

and the bounds are established analogously to the bounds for \( c_t(\theta) \).

\[ \square \]

**Lemma 4.** Suppose that Assumptions 1, 2 and 6 are satisfied and \( \lim_{\theta \to \infty} D_t(\theta) = 0 \). Then \( C_t(\theta) \geq 0 \) for sufficiently large \( \theta \) and

\[
\lim_{\theta \to \infty} C_t(\theta) = 1 + \frac{\bar{r}}{1 - \bar{r}} \frac{\bar{\sigma} - \bar{\gamma}}{\bar{\sigma} + \bar{\varepsilon} - \bar{\gamma} (X + 1)}.
\]

(42)

Assumption 6 is satisfied only if equation

\[
\frac{\bar{r}}{1 - \bar{r}} = (1 + \bar{\varepsilon}) \left(1 + \frac{\bar{r}}{1 - \bar{r}} \frac{\bar{\sigma} - \bar{\gamma}}{\bar{\sigma} + \bar{\varepsilon} - \bar{\gamma} (X + 1)} \right) \lim_{\theta \to \infty} \frac{1 - F_t(\theta)}{\theta f_t(\theta)}
\]

(43)

holds for a non-negative \( \frac{\bar{\tau}}{1 - \bar{r}} \).

**Proof.** Let \( g_t(\bar{x}) \equiv \left[ \sigma_t(\bar{x}) \frac{\hat{c}_t(\bar{x})}{c_t(\bar{x})} + \gamma_t(\bar{x}) \left(\frac{\hat{l}_t(\bar{x})}{l_t(\bar{x})} + 1\right) \right] \) and re-write \( C_t \) as

\[
C_t(\theta) = \frac{\exp\left(-\int_0^\theta g_t(\bar{x})d\bar{x}\right) \int_\theta^\infty \exp\left(\int_\theta^x g_t(\bar{x})d\bar{x}\right) \left(1 - \lambda_{1,t}(x) U_{c,t}(x)\right) f_t(x) dx}{1 - F_t(\theta)}.
\]

(44)
Since $\frac{\tau^\nu(\theta)}{1-\tau^\nu(\theta)}$, $A_t(\theta)$, $B_t(\theta)$ and $D_t(\theta)$ all tend to finite limits as $\theta \to \infty$ by Assumptions 1 and 6, equation (17) implies that the limit of $C_t(\theta)$ also exists and is finite. Since $U_{c,t}(\theta) \to 0$ ($\theta \to \infty$) from Lemma 3 and $\tilde{\alpha}_t(\theta)$ is bounded, $C_t(\theta)$ is positive for sufficiently high $\theta$.

Apply L’Hospital’s rule and substitute for $C_t(\theta)$ from (17)

$$
\lim_{\theta \to \infty} C_t(\theta) = \lim_{\theta \to \infty} \frac{-(1 - \lambda_{1,t}\tilde{\alpha}_t(\theta) U_{c,t}(\theta)) f_t(\theta)}{-f_t(\theta)} + \lim_{\theta \to \infty} \frac{-g_t(\theta) C_t(\theta) (1 - F_t(\theta))}{-\theta f_t(\theta)}
$$

$$
= 1 + \lim_{\theta \to \infty} g_t(\theta) \left\{ \left[ \frac{\tau_t(\theta)}{1 - \tau_t(\theta)} - \beta R \frac{\tau_{t-1}}{1 - \tau_{t-1}} D_t(\theta) \right] \frac{1}{A_t(\theta)} \right\}
$$

$$
= 1 + \lim_{\theta \to \infty} \left\{ \sigma_t(\theta) \frac{\hat{c}_t}{c_t} - \gamma_t(\theta) \frac{\hat{y}_t}{y_t} \right\} \left[ \frac{\tau_t(\theta)}{1 - \tau_t(\theta)} - \beta R \frac{\tau_{t-1}}{1 - \tau_{t-1}} D_t(\theta) \right] \frac{1}{A_t(\theta)}
$$

Equation (42) follows from substituting (39) and $\lim_{\theta \to \infty} D_t(\theta) = 0$ into the expression above.

Since $U$ satisfies Assumption 2, $A_t(\theta) \geq 0$ for all $\theta$ and therefore $\lim_{\theta \to \infty} \frac{\tau^\nu(\theta)}{1-\tau^\nu(\theta)} = \lim_{\theta \to \infty} A_t(\theta) B_t(\theta) C_t(\theta) \geq 0$. Therefore equation (43) should be satisfied for a non-negative $\frac{\tau}{1-\tau}$.

A.3.2 Proof of Proposition 1

Proof. We first show that there are real $k_1, k_2$ such that $A_t(\theta) B_t(\theta) C_t(\theta) \sim k_1 \frac{F_t(\theta)}{f_t(\theta)}$, $D_t(\theta) \sim k_2 \varphi_t(\theta)$ ($\theta \to 0$). Use (38) to write $C_t$ as

$$
C_t(\theta) = \int_\theta^\infty \frac{U_{c,t}(\theta)}{U_{c,t}(x)} (1 - \lambda_{1,t}\tilde{\alpha}_t(x) U_{c,t}(x)) \frac{f_t(x)}{1 - F_t(\theta)} dx.
$$
Note that since $U_{c,t}(0)$ is well-defined and finite by Assumption 4,

$$
\lim_{\theta \to 0} C_t(\theta) = U_{c,t}(0) \int_0^\infty \left( \frac{1}{U_{c,t}(x)} - \lambda_{1,t} \bar{a}_t(x) \right) f_t(x) \, dx \\
= U_{c,t}(0) \left[ \int_0^\infty \frac{1}{U_{c,t}(x)} f_t(x) \, dx - \lambda_{1,t} \right] = 0
$$

from the definition of $\lambda_{1,t}$ and the fact that $\int_0^\infty \bar{a}_t(x) f_t(x) \, dx = 1$ for all $t$. Applying L’Hospital’s rule,

$$
\lim_{\theta \to 0} \frac{C_t(\theta) {1-F_t(\theta)}}{F_t(\theta)} = - \left( \frac{1}{U_{c,t}(0)} - \lambda_{1,t} \bar{a}_t(0) \right),
$$

since limits of $U_{c,t}(\theta)$ and $\bar{a}_t(\theta)$ are well defined. Let $k_1 = -(1 + \varepsilon_t(0)) (1 - \lambda_{1,t} \bar{a}_t(0) U_{c,t}(0))$, which is well-defined by Assumption 4. We have

$$
\lim_{\theta \to 0} \frac{A_t(\theta) B_t(\theta) C_t(\theta)}{k_1 F_t(\theta)/\theta f_t(\theta)} = \lim_{\theta \to 0} \frac{1}{k_1} A_t(\theta) U_{c,t}(\theta) \frac{C_t(\theta) {1-F_t(\theta)}}{F_t(\theta)} = 1.
$$

The result for $D_t(\theta)$ follows immediately by setting $k_2 = \frac{A_t(0) U_{c,t}(0)}{A_{t-1} U_{c,t-1}}$, which is well-defined by Assumption 4.

We next show that $D_t(\theta) = o\left(\frac{1}{\theta^{k_4}}\right)$ ($\theta \to \infty$) and $k_4 > 0$ generically depends only on $U$. Since $\varphi_t(\theta)$ is bounded by Assumption 1 and $A_t(\theta)$ is bounded for $\theta$ sufficiently high by Assumption 4, $|D_t(\theta)| \leq K_{t-1} U_{c,t}(\theta)$ for some $K_{t-1} > 0$. Lemma 3 yields the result.

Finally we show that $A_t(\theta) B_t(\theta) C_t(\theta) \sim k_3 \frac{1-F_t(\theta)}{\theta f_t(\theta)}$ as $\theta \to \infty$ and $k_3 > 0$. 

xi
depends generically on \( U \) and \( f \). Using Lemma 4,

\[
\frac{\bar{r}}{1 - \bar{r}} = \lim_{\theta \to \infty} A_t(\theta) B_t(\theta) C_t(\theta) = (1 + \bar{e}) \left( 1 + \frac{\bar{r}}{1 - \bar{r}} \frac{\bar{\sigma}}{\bar{\sigma} + \bar{e}} \right) \lim_{\theta \to \infty} \frac{1 - F_t(\theta)}{\theta f_t(\theta)}.
\]

(46)

If \( \lim_{\theta \to \infty} \frac{1 - F_t(\theta)}{\theta f_t(\theta)} = 0 \), then \( A_t(\theta) B_t(\theta) C_t(\theta) \sim (1 + \bar{e}) \frac{1 - F_t(\theta)}{\theta f_t(\theta)} (\theta \to \infty) \). If \( \lim_{\theta \to \infty} \frac{1 - F_t(\theta)}{\theta f_t(\theta)} > 0 \), then (46) defines \( \frac{\bar{e}}{1 - \bar{r}} \) as a function of \( f_t, \bar{e}, \bar{\sigma} \). Then setting \( k_3 = (1 + \bar{e})(1 + \frac{\bar{e}}{1 - \bar{r} \bar{\sigma} + \bar{e}}) \) we obtain the result for \( \theta \to \infty \). Note that \( \bar{e}, \bar{\sigma} \) generically depend only on \( U \) by Lemma 2.

\[\square\]

### A.3.3 Proof of Corollary 1

We first prove a preliminary lemma about the properties of \( f \).

**Lemma 5.** Suppose \( f_t \) satisfies Assumption 5. Then

1. \( \varphi_t(\theta) = \rho \) for all \( \theta \); If \( \rho \geq 0 \) then there is \( \hat{\theta} \) such that \( f_{2,t}(\theta) \geq 0 \) for all \( \theta \geq \hat{\theta} \).

2. \( i \frac{F_t(\theta)}{\theta f_t} \sim \frac{f_t}{\theta f_t} \sim -\frac{\nu^2}{\ln \theta} (\theta \to 0) \);

3. If \( f_t \) is lognormal/mixture, then \( \frac{1 - F_t(\theta)}{\theta f_t} \sim \frac{f_t}{\theta f_t} \sim -\frac{\nu^2}{\ln \theta} (\theta \to \infty) \); if \( f_t \) is Pareto-lognormal then \( \lim_{\theta \to \infty} \frac{1 - F_t(\theta)}{\theta f_t} = \frac{1}{a} \) and \( \lim_{\theta \to \infty} \frac{f_t}{\theta f_t} = -\frac{1}{a+1} \).

**Proof.** Let \( \Phi(\cdot), \phi(\cdot) \) be standard normal cdf and pdf. Direct calculations yield

\[
\lim_{x \to \infty} \frac{\phi(x)}{\Phi(x)} = 0, \quad \lim_{x \to -\infty} \frac{\phi(x)}{\Phi(x)} = \infty, \quad \lim_{x \to -\infty} \frac{\phi(x)}{-x\Phi(x)} = 1.
\]

(47)

When \( f_t \) is lognormal, it is given by \( f_t(\theta) = \frac{1}{\theta v} \phi \left( \frac{\ln \theta - \mu_t}{v} \right) \) where \( \mu_t = b_t + \rho \ln \theta_{t-1} \); when \( f_t \) is a mixture then \( f_t(\theta) = \sum_{i=1}^T \frac{p_i}{\bar{v}_i} \phi \left( \frac{\ln \theta - \mu_{t,i}}{v_i} \right) \) where \( \mu_t = x_{ii} \).
Suppose \( f_t \) is lognormal. Then \( f_{t-1}f_2; t = f_t \phi \left( \frac{\ln \theta - \mu_t - a v^2}{v} \right) + a \rho f_t (\theta) \).

Note that using integration by parts

\[
\int_{\theta}^{\infty} \hat{A}_t \frac{\rho}{v} x^{-a-1} \phi \left( \frac{\ln x - \mu_t - a v^2}{v} \right) dx = \rho a \int_{\theta}^{\infty} \hat{A}_t x^{-a-1} \Phi \left( \frac{\ln x - \mu_t - a v^2}{v} \right) dx - \rho \hat{A}_t \theta^{-a} \Phi \left( \frac{\ln \theta - \mu_t - a v^2}{v} \right) = a \rho (1 - F_t(\theta)) - \rho \theta f_t(\theta).
\]

Therefore

\[
\int_{\theta}^{\infty} \theta_{t-1} f_{2,t}(x) dx = -a \rho (1 - F_t(\theta)) + \rho \theta f_t(\theta) + a \rho (1 - F_t(\theta)) = \rho \theta f_t(\theta)
\]

and hence \( \varphi_t(\theta) = \rho \) for all \( \theta \). The second part of (i) follows by inspection of expressions for \( f_{2,t}(\theta) \).

(ii) and (iii). Suppose \( f_t \) is lognormal. Then \( \theta f_t' = -f_t \left( 1 + \frac{\ln \theta - \mu_t}{v^2} \right) \), and
therefore $\theta f_t' / f_t \sim -\frac{\ln \theta}{v_1^2} (\theta \to 0, \infty)$. By L'Hospital's rule,

$$\lim_{\theta \to 0} \frac{F_t(\theta)}{\theta f_t(\theta)} = \lim_{\theta \to 0} \frac{1}{\theta f_t(\theta) / f_t(\theta) + 1} = 0,$$

$$\lim_{\theta \to 0} \frac{1 - F_t(\theta)}{\theta f_t(\theta)} = \lim_{\theta \to 0} \frac{-1}{\theta f_t(\theta) / f_t(\theta) + 1} = 0,$$

and

$$\lim_{\theta \to 0} \frac{-F_t \ln \theta / v^2}{\theta f_t} = \lim_{\theta \to 0} -\frac{\ln \theta / v^2 - F_t / (\theta f_t v^2)}{(\theta f_t / f_t + 1)} = 1,$$

$$\lim_{\theta \to \infty} \frac{(1 - F_t) \ln \theta / v^2}{\theta f_t} = \lim_{\theta \to \infty} -\frac{\ln \theta / v^2 + (1 - F_t) / (\theta f_t v^2)}{(\theta f_t / f_t + 1)} = 1.$$

This implies that $F_t(\theta) / f_t \sim -\frac{\ln \theta}{v_1^2} (\theta \to 0)$ and $(1 - F_t(\theta)) / f_t \sim -\frac{\ln \theta}{v_1^2} (\theta \to \infty)$.

If $f_t$ is a mixture, assume without loss of generality that $v_1 \geq v_i$ for all $i$. Then

$$\frac{\theta f_t'}{f_t} = -\left(\frac{\ln \theta - \hat{\mu}_{1,t}}{v_1^2} + 1\right) \left(\frac{p_{1}}{v_1} + \sum_{i=2}^{I} \frac{p_{i}}{v_i} \frac{\phi(\frac{\ln \theta - \hat{\mu}_{i,t}}{v_i} / \phi(\frac{\ln \theta - \hat{\mu}_{i,t}}{v_i}))}{v_i^2} + 1\right).$$

Since $v_1 \geq v_i$, $\phi(\frac{\ln \theta - \hat{\mu}_{i,t}}{v_i}) / \phi(\frac{\ln \theta - \hat{\mu}_{1,t}}{v_1}) \to 0$ as $\ln \theta \to \pm \infty$ and therefore the last term in the expression above converges to 0 as $\ln \theta \to \pm \infty$. This implies that $\theta f_t'/f_t \sim -\ln \theta / v_1^2 (\theta \to 0, \infty)$. The rest follows by analogy with the lognormal case.

If $f_t$ is Pareto-lognormal, then $\theta f_t' = (-a - 1) f_t + f_t \frac{1}{v} \phi(\frac{\ln \theta - \hat{\mu}}{v}) / \Phi(\frac{\ln \theta - \hat{\mu}}{v})$.

xiv
which immediately implies that $\theta f_t' / f_t \to -(a + 1) (\theta \to \infty)$. Also

$$\lim_{\theta \to 0} \frac{\theta f_t'}{f_t - \ln \theta} = \lim_{\theta \to 0} \frac{\phi \left( \frac{\ln \theta - \mu}{\sigma \sqrt{\ln \theta}} \right)}{\Phi \left( \frac{\ln \theta - \mu}{\sigma \sqrt{\ln \theta}} \right)}$$

From (47), $\phi \left( \frac{\ln \theta - \mu}{\sigma \sqrt{\ln \theta}} \right) / \Phi \left( \frac{\ln \theta - \mu}{\sigma \sqrt{\ln \theta}} \right) \sim \frac{\ln \theta - \mu}{\sigma \sqrt{\ln \theta}} \left( \ln \theta - \mu \to -\infty \right)$, therefore $\theta f_t' / f_t \sim \frac{-\ln \theta - \mu}{\sigma^2} \sim \frac{-\ln \theta}{\sigma^2} (\theta \to 0)$. The rest follows by analogy with the lognormal case.

With this lemma we can prove Corollary 1.

**Proof (of Corollary 1).** If $f_t$ satisfies assumption 5, then Lemma 5 and Proposition 1 show that $A_t(\theta) B_t(\theta) C_t(\theta) \sim k_1 \frac{a^2}{-\ln \theta} (\theta \to 0)$, $D_t(0) = \rho \frac{U_{c,t}(0)}{U_{c,t-1}} > 0$. This establishes (20). They also establish that $\lim_{\theta \to \infty} D_t(\theta) = 0$. Therefore from Lemma 4 it follows that $\lim_{\theta \to \infty} C_t(\theta) = 1 + \frac{a}{\sigma + \bar{\epsilon}} \frac{\bar{\tau}}{1 - \bar{\tau}}$ and the expressions for $\bar{\sigma}$ and $\bar{\epsilon}$ in terms of limits of $\frac{U_{c,t}}{U_{c,t}}$ and $\frac{U_{m,t}}{U_{l,t}}$ follows from Lemma 2. This establishes (18).

Finally, to show (19) we first suppose that $f_t$ is Pareto-lognormal. Then from Lemma 5 $\lim_{\theta \to \infty} B_t(\theta) = a^{-1}$, and taking limits of (17) yields

$$\frac{\bar{\tau}}{1 - \bar{\tau}} = \frac{1 + \bar{\epsilon}}{a} \left[ 1 + \frac{\bar{\tau}}{1 - \bar{\tau}} \frac{\bar{\sigma}}{\bar{\sigma} + \bar{\epsilon}} \right].$$

Re-arranging the terms, we obtain $\frac{\bar{\tau}}{1 - \bar{\tau}} = \left[ a \frac{1}{1 + \bar{\epsilon}} - \frac{\bar{\sigma}}{\bar{\sigma} + \bar{\epsilon}} \right]^{-1}$. By Lemma 4 this limit must be non-negative, therefore a necessary condition for the distortions to be finite is that $a \frac{1}{1 + \bar{\epsilon}} - \frac{\bar{\sigma}}{\bar{\sigma} + \bar{\epsilon}} > 0$.

If $f_t$ is lognormal/mixture, then Lemma 5 and Proposition 1 imply that $B_t(\theta) \sim \left[ \frac{\ln \theta}{\sigma^2} \right]^{-1}$, $D_t(\theta) = \rho \left[ \left( \frac{\ln \theta}{\sigma^2} \right)^{-1} \right] (\theta \to \infty)$, the latter follows from the
fact that $\lim_{\theta \to \infty} \theta^{-\kappa} \ln \theta = 0$ for any $\kappa > 0$. Since $\lim_{\theta \to \infty} A_t(\theta) = 1 + \bar{\varepsilon}$
and $C_t(\theta)$ are bounded, this implies that $\lim_{\theta \to \infty} \frac{\tau^*_t(\theta)}{1 - \tau^*_t(\theta)} = 1$ and therefore
$\lim_{\theta \to \infty} C_t(\theta) = 1$. Therefore
$\frac{\tau^*_t(\theta)}{1 - \tau^*_t(\theta)} \sim A_t(\theta) B_t(\theta) C_t(\theta) \sim \left[ \frac{\ln \theta}{\ln (1 + \bar{\varepsilon})} \right]^{-1}$.

\end{proof}

\section*{A.3.4 Proofs in Section 2.2}

We first show equation (22).

\begin{lemma}
Suppose that Assumptions 5 and 6 are satisfied, $\rho \geq 0$, $U_{cl} \geq 0$. Then (22) holds.
\end{lemma}

\begin{proof}
We first show that $D_t(\theta) = o \left( \theta^{-\kappa} \right)$ ($\theta \to \infty$) where $\kappa$ as defined in Lemma 3. If $f_t$ satisfies Assumption 5 and $\rho \geq 0$, then by Lemma 5 there
exists $\hat{\theta}$ such that $f_{2,t}(\theta) \geq 0$ for all $\theta \geq \hat{\theta}$. Therefore if $U_{cl} \geq 0$ then $\gamma \geq 0$ and
$\exp \left( - \int_{\theta}^{\bar{x}} \gamma(\bar{x}) \frac{d\bar{x}}{\bar{x}} \right) f_{2,t}(x) \leq f_{2,t}(x)$ for all $x, \theta$ such that $x \geq \theta \geq \hat{\theta}$. Using Lemma 5,

$$\frac{\theta_{t-1} \int_{\theta}^{\infty} \exp \left( - \int_{\theta}^{\bar{x}} \gamma(\bar{x}) \frac{d\bar{x}}{\bar{x}} \right) f_{2,t}(x) dx}{\theta f_t(\theta)} \leq \rho$$ \text{ for all } \theta \geq \hat{\theta}.

Therefore $D_t(\theta) \leq K_{t-1} U_{c,t}(\theta)$ for some $K_{t-1}$ and then Lemma 3 yields the
result that $D_t(\theta) = o \left( \theta^{-\kappa} \right)$. Since $\lim_{\theta \to \infty} D_t(\theta) = 0$, Lemma 4 implies that
the limit $\bar{\tau}$ satisfies (43). The rest of the steps are identical to the proof of
Corollary 1.
\end{proof}

We now show the remaining results discussed in Section 2.2.
Lemma 7. Suppose that Assumption 6 is satisfied. Then

\[ \frac{\beta w_{1,t} (\theta) \theta}{U_{c,t} (\theta) c_t (\theta)} = (1 - \tau_t^\theta (\theta)) \frac{\dot{y}_t}{c_t} - \frac{\dot{c}_t}{c_t}. \]  

(48)

In the limit

\[ \lim_{\theta \to \infty} \frac{\beta w_{1,t} (\theta) \theta}{U_{c,t} (\theta) c_t (\theta)} = \frac{1 - X}{X} \lim_{\theta \to \infty} \frac{\dot{y}_t}{y_t}. \]  

(49)

Proof. Differentiating (10), we get

\[ \dot{u}_t (\theta) = U_{c,t} (\theta) \dot{c}_t (\theta) + U_{l,t} (\theta) \dot{l}_t (\theta) + \beta (w_{1,t} (\theta) + w_{2,t} (\theta)). \]

Substitute into (7) to get

\[ U_{c,t} (\theta) \dot{c}_t (\theta) + U_{l,t} (\theta) \dot{l}_t (\theta) + \beta w_{1,t} (\theta) = -U_{l,t} (\theta) \frac{l_t (\theta)}{\theta}. \]  

(50)

Re-arrange to get (48). Note that \( \theta \dot{y}_t / y_t = 1 + \dot{\theta} \dot{l}_t / l_t. \) Then use (39) to obtain the limit. \( \square \)

Compensated and uncompensated elasticities holding savings fixed coincide with compensated and uncompensated elasticities in the static model, where they are given by (see p. 227 in Saez (2001))

\[ \zeta^u = \frac{U_{l,l} - (U_{l} / U_{c})^2 U_{c c} + (U_{l} / U_{c}) U_{l c}}{U_{l l} + (U_{l} / U_{c})^2 U_{c c} - 2 (U_{l} / U_{c}) U_{l c}}, \]

\[ \zeta^c = \frac{U_{l,l}}{U_{l l} + (U_{l} / U_{c})^2 U_{c c} - 2 (U_{l} / U_{c}) U_{l c}}, \]

\[ -\eta = \zeta^c - \zeta^u. \]

Note that normality of leisure implies \( \eta < 0. \) We use \( \zeta^u (\theta), \zeta^c (\theta), \eta (\theta) \) to denote the elasticities evaluated at the optimum and \( \bar{\zeta}^u, \bar{\zeta}^c, \bar{\eta} \) their limits as \( \theta \to \infty. \)
Lemma 8. \(A_t(\theta)\) and \(C_t(\theta)\) can be written as

\[
A_t(\theta) = \frac{1 + \zeta_t^u(\theta)}{\zeta_t^c(\theta)},
\]

\[
C_t(\theta) = \frac{\int_\theta^\infty \exp \left( \left\{ \int_\theta^x \frac{-\eta_t(\tilde{\theta}) \hat{y}_t}{\tilde{\eta}_t} + \sigma_t(\tilde{x}) \frac{(1-\tau_t^c(\tilde{\theta}))\hat{y}_t - \hat{c}_t}{c_t} \right\} \tilde{x} d\tilde{x} \right) (1 - \lambda_{1,t} \hat{\eta}_t U_{c,t}(x)) f_t(x) \, dx}{1 - F_t(\theta)}.
\]

If preferences are GHH, then (24) holds.

Proof. The proof for \(A_t(\theta)\) follows from the definition of elasticities. To rewrite \(C_t(\theta)\) let \(g_t(\cdot)\) be as defined in the proof of Lemma 4. Using (48) it can be written as

\[
g_t(\theta) = -\sigma_t(\theta) \frac{(1 - \tau_t^c(\theta))(\hat{y}_t - \hat{c}_t)}{c_t} - \frac{\eta_t(\theta) \hat{y}_t}{\zeta_t^c(\theta) y_t}.
\]

Substitute into (44) to get the expression for \(C_t\). When preferences are GHH, \(\eta_t(\theta) = 0\) and \(\lim_{\theta \to \infty} \frac{\hat{y}_t}{\hat{y}_t} = 1 + \zeta\) by Lemma 2. Use this fact together with Lemma 7 to show that \(g_t(\theta) \to \frac{1-x}{X} (1 + \zeta) (\theta \to \infty)\) in this case. Since \(A_t(\theta) = (1 + \zeta)/\zeta\), this together with (45) implies (24).

Proof (of Proposition 2). We can express \(\varphi\) using (30) rather than (29), in which case the differential equation for \(\psi\), (31), becomes \(\frac{1}{\beta R} \frac{\partial V_{t+1}}{\partial w} f - \lambda_1 f + \lambda_2 f_2 = -\dot{\psi}\). Integrate this expression from 0 to infinity, use the boundary conditions \(\psi(0) = \psi(\infty) = 0\) and \(\int_0^\infty f_2 \, dx = 0\) to obtain\(^{19}\)

\[
\lambda_{1,t} = \frac{1}{\beta R} \int_0^\infty \frac{\partial V_{t+1}}{\partial w} (x) f(x) \, dx.
\]

\(^{19}\)To see that \(\int_0^\infty f_2(x|\theta_-) \, dx = 0 \) for all \(\theta_-\), differentiate both sides of \(\int_0^\infty f(x|\theta_-) \, dx = 1\) with respect to \(\theta_-\).
Combine this expression with (26) to get \( \frac{\partial V_t}{\partial \omega} = \frac{1}{\beta R} \mathbb{E}_t \frac{\partial V_{t+1}}{\partial \omega} \) and by the law of iterated expectations

\[
\frac{\partial V_t}{\partial \omega} = \left( \frac{1}{\beta R} \right)^{T-t-1} \mathbb{E}_t \frac{1}{U_c} \frac{1}{U_c}.
\] (51)

When \( F_T(0|\theta) = 1 \) for all \( \theta \), \( \frac{\partial V_T}{\partial \omega}(\theta^T) = \frac{1}{U_c(\theta^T)} \), which, from (51), implies

\[
\lambda_{1,t} = \frac{\partial V_t}{\partial \omega} > 0 \text{ for all } t
\] (52)

and, in combination with (33),

\[
\frac{1}{U_c(\theta^t)} - \psi \frac{\partial \theta}{\partial \gamma} = \left( \frac{1}{\beta R} \right)^{T-t} \mathbb{E}_t \frac{1}{U_c} \frac{1}{U_c}.
\]

Note that \( \frac{\partial U_c(c|y/\theta)}{\partial \gamma} \geq 0 \) from Assumption 2 implies that \( 1 + \varepsilon - \gamma \geq 0 \), therefore from equation (35) the sign of \( \psi \) is equal to the sign of \( \tau_v \). Thus if \( \tau_v \geq 0 \) then

\[
\frac{1}{U_c(\theta^t)} \geq \left( \frac{1}{\beta R} \right)^{T-t} \mathbb{E}_t \frac{1}{U_c} \frac{1}{U_c} \geq \left( \frac{1}{\beta R} \right)^{T-t} \frac{1}{\mathbb{E}_t U_c(\theta^t)}.
\]

where the last expression follows from Jensen’s inequality. This expression implies that \( \tilde{\tau}_t^y(\theta^t) \geq 0 \). This inequality is strict if \( \text{var}_{\theta^t}(c_T) > 0 \). \( \square \)

**Lemma 9.** Suppose that preferences are \( U \left( c - \frac{1}{1+1/\zeta} t^{1+1/\zeta} \right) \), where \( U \) is concave, \( U''/U' \) is bounded away from zero, \( f_t \) satisfies Assumption 5 with \( \rho \geq 0 \) and \( F_T(0|\theta) = 1 \) for all \( \theta \). If \( \tau_v^y(\theta) \) is positive and bounded away from 1 for high \( \theta \), then \( \tau_v^y(\theta) \to 0 \) as \( \theta \to \infty \). A sufficient condition for \( \tau_v^y(\theta) \) to be
bounded is that \( U \) is exponential: 
\[
U(x) = -\exp \left( -\hat{k} x \right) \text{ for some } \hat{k} > 0.
\]

**Proof.** The first order conditions (29) and (30) can be written as
\[
\frac{1}{U_c} - \frac{\psi U_{cl}}{\theta f U_c} = \frac{1}{\beta R} \frac{\partial V_{t+1}}{\partial w}.
\]

From (35) we have
\[
\frac{\psi}{\theta f} = \frac{\tau^y}{1 - \tau^y} (1 + \varepsilon - \gamma)^{-1} \frac{1}{U_c}
\]

which implies
\[
\frac{1}{\beta R} \frac{\partial V_{t+1}(w)}{\partial w} U_c = 1 - \frac{\tau^y}{1 - \tau^y} (1 + 1/\zeta)^{-1} \frac{U_{cl}}{U_c} = 1 - (1 + 1/\zeta)^{-1} \frac{-U''}{U} \tau^y \theta l
\]

Since \( F_T(0|\theta) = 1 \), both sides of this expression must be positive by (52).

Suppose that \( \tau^y \) does not converge to 1. Take any sequence \( \tau^y(\theta_n) \) and since \( \tau^y(\theta_n) \in [0,1] \) it must have a convergent subsequence. We will show that any such subsequence that does not converge to 1 must converge to 0.

Suppose \( \tau^y(\theta_n) \to \bar{\tau}^y < 1 \). Then the FOCs \( l^{1/\zeta} = \theta (1 - \tau^y) \) implies that \( l \to \infty \) (\( \theta \to \infty \)) and, since \( \frac{-U''}{U} \) is bounded away from 0, the right hand side of (53) converges to \( -\infty \). The left hand side is positive, leading to a contradiction.

Under our assumptions \( \tau^y(\theta) \) diverges to 1 only if either \( C_t(\theta) \) or \( D_t(\theta) \) diverge to \( +\infty \). Either of these cases would imply that \( U_t(\theta) \to -\infty \) as \( \theta \to \infty \).

If \( U \) is exponential, it is bounded above in all periods, and therefore \( U_t(\theta) \to -\infty \) in any \( t \) would violate the incentive compatibility. In particular, to see
that $C_t (\theta) \to \infty$ implies $U_t (\theta) \to -\infty$, note that with exponential $U$ there is some $\hat{k} > 0$ so that

\[
C_t (\theta) = \int_{\theta}^{\infty} \exp \left( \int_{\theta}^{x} \beta \frac{U''(\tilde{x})}{U'_t(\tilde{x})} \tilde{c}_t (\tilde{x}) d\tilde{x} \right) \frac{U'_t(x)}{1 - \lambda_{1,t}\tilde{\alpha}_t (x) U'_t (x)} \frac{f_t (x) dx}{1 - F_t (\theta)}
\]

= \int_{\theta}^{\infty} \exp \left( \beta \hat{k} (c_t (\theta) - c_t (x)) \right) (1 - \lambda_{1,t}\tilde{\alpha}_t (x) U'_t (x)) \frac{f_t (x) dx}{1 - F_t (\theta)}.

Since $\lambda_{1,t} > 0$ by (52), $1 - \lambda_{1,t}\tilde{\alpha}_t (x) U'_t (x)$ is bounded from above and therefore $C_t (\theta)$ can diverge to infinity only if the exponent diverges to infinity, which is possible only if $c_t (x) \to -\infty$ and therefore $U_t (x) \to -\infty$.

A.4 Additional details for Section 3

We first describe further details of the analysis in Section 3 and then provide additional illustrations and robustness checks.

To make the numerical solution feasible we exploit the recursive structure of the dual formulation of the planning problem that we discussed in Section 1. The recursive problem is (6) together with (11) and $V_0 (\hat{w}_0) = 0$, which is a finite-horizon discrete-time dynamic programming problem with a three-dimensional continuous state vector: $\hat{w}$ is the promised utility associated with the promise-keeping constraint (8); $\hat{w}_2$ is the state variable associated with the threat-keeping constraint (9); $\theta_-$ is the type in the preceding period. In the initial period the state is $\hat{w}_0$, given by the solution to $V_0 (\hat{w}_0) = 0$.

We proceed in stages. First, we implement a value function iteration for problems (6) and (11). We start from the last working period, $\hat{T} - 1$, and proceed by backward induction. Since $F_t (0|\theta) = 1$ for all $\theta$ for $t \geq \hat{T}$, the
Figure 6: Optimal average labor (Panel A) and savings (Panel B) distortions as functions of current shock realization at selected periods.

The planner sets \( w_2(\theta) = 0 \) for all \( \theta \) in period \( \hat{T} - 1 \) and we replace the value function \( V_{\hat{T}}(w(\theta), 0, \theta) \) in problem (6) for period \( \hat{T} - 1 \) with the discounted present value of resources required to provide promised utility \( w \) over the remaining \( T - \hat{T} + 1 \) periods.

We approximate value functions with tensor products of orthogonal polynomials evaluated over the state space. We use Chebyshev polynomials of degrees 1 through 10 and check in the baseline case that value function differences do not exceed 1 percent of original values after doubling the degrees to 20. The evaluation nodes are allocated over the state space at the roots of the polynomials, given by \( r_n = -\cos(\pi (2n - 1)/2N) \), where \( n = 1, \ldots, N \) indexes the nodes. This gives the roots on the interval \([-1, 1]\) and a change of variables is needed to adjust the root nodes. We let \( N = 11 \) for both the promise, \( \hat{w} \), and for the threat, \( \hat{w}_2 \). For the skill, we set 30 logarithmically spaced nodes to better capture the more complex U-shapes in the left tail. The polynomial coefficients are computed by minimizing the sum of squared distances from the computed values at the nodes. The approximation provides each period-\( t \)
Table 2: Simulated earnings and consumption moments of the constrained optima and the earnings moments in the data.

<table>
<thead>
<tr>
<th>Stochastic process</th>
<th>Mean</th>
<th>SD</th>
<th>Kurtosis</th>
<th>Kelly’s Skewness</th>
<th>P10</th>
<th>P90</th>
<th>P50</th>
<th>P90</th>
<th>P99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data earnings moments ($y_{t, data}^t$):</td>
<td>0.009</td>
<td>0.52</td>
<td>11.31</td>
<td>-0.21</td>
<td>-0.44</td>
<td>0.47</td>
<td>10.06</td>
<td>10.76</td>
<td>11.71</td>
</tr>
<tr>
<td>Lognormal constrained-optimum earnings moments ($y_{t, log normal}^t$):</td>
<td>-0.005</td>
<td>1.09</td>
<td>2.51</td>
<td>-0.03</td>
<td>-1.02</td>
<td>0.90</td>
<td>13.46</td>
<td>13.79</td>
<td>14.48</td>
</tr>
<tr>
<td>Mixture constrained-optimum earnings moments ($y_{t, mixture}^t$):</td>
<td>0.004</td>
<td>0.79</td>
<td>11.48</td>
<td>-0.37</td>
<td>-0.90</td>
<td>0.51</td>
<td>13.03</td>
<td>13.73</td>
<td>14.29</td>
</tr>
<tr>
<td>Lognormal constrained-optimum consumption moments ($c_{t, log normal}^t$):</td>
<td>0.001</td>
<td>0.75</td>
<td>3.95</td>
<td>-0.18</td>
<td>-0.13</td>
<td>1.07</td>
<td>10.52</td>
<td>11.92</td>
<td>12.66</td>
</tr>
<tr>
<td>Mixture constrained-optimum consumption moments ($c_{t, mixture}^t$):</td>
<td>-0.001</td>
<td>0.13</td>
<td>19.07</td>
<td>0.15</td>
<td>-0.33</td>
<td>0.55</td>
<td>11.89</td>
<td>12.07</td>
<td>13.43</td>
</tr>
</tbody>
</table>

problem (6) with a continuously differentiable function approximating $V_{t+1}$.

We use the trigonometric form of the polynomials in the evaluation of the tensor products, $P_d (r) = \cos (d \arccos (r))$, to be able to apply an implementation of algorithmic (chain rule) differentiation.

It is a familiar property of the state space in such problems that no constrained optimal allocations may exist for some nodes (see, e.g., the discussion in subsection 3.2. in Abraham and Pavoni (2008)). To deal with this while maintaining large enough number of computed nodes, we follow the procedure in Kapička (2013) in subsections 7.1 and 7.2.\textsuperscript{20}

For computational feasibility it is essential to use an efficient and robust optimization algorithm for the minimization problems at each node. We use an implementation of the interior-point algorithm with conjugate gradient it-

\textsuperscript{20}Generally one has a choice to implement a state space restriction procedure, to eliminate such nodes, or a procedure assigning sufficiently large penalties. For discussions of both and examples of implementation in closely related problem see, e.g., Abraham and Pavoni (2008) and Kapička (2013) and references therein.
Figure 7: The decomposition of optimal labor distortions as functions of current earnings: only intratemporal forces ($A_t, B_t, C_t$) in Panels A and C; both intra- and intertemporal forces ($\tau^y_t / (1 - \tau^y_t)$) in Panels B and D. Panels A and B have a history of $\bar{\theta}$ shocks chosen so that an individual with a lifetime of $\bar{\theta}$ shocks will have the average lifetime earnings approximately equal to the average U.S. male earnings in 2005; Panels C and D are the analogues with $\bar{\theta}$ chosen so that the average lifetime earnings approximately equal twice the U.S. average.

eration to compute the optimization step.\footnote{See, for example, Su and Judd (2007).} It uses a trust-region method to solve barrier problems; the acceptance criterion is an $l_1$ barrier penalty function. To improve the accuracy of the solution estimates, including multipliers, we proceed to active-set iterations that use the output of the interior-point algorithm as its input. The implementation of the active-set algorithm is based on the sequential linear quadratic programming.
Table 3: Calibrated parameters of the shock process for selected Frisch elasticity parameter values.

<table>
<thead>
<tr>
<th>Stochastic process</th>
<th>Initial distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>$\mu_3$</td>
</tr>
<tr>
<td>The higher elasticity case of $\varepsilon = 1$:</td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td>-0.45</td>
</tr>
<tr>
<td>The baseline case of $\varepsilon = 2$:</td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td>-0.47</td>
</tr>
<tr>
<td>The lower elasticity case of $\varepsilon = 4$:</td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>-0.51</td>
</tr>
</tbody>
</table>

We check at this stage the increasing properties of Assumption 3 used in Lemma 1 (Assumption 2 is satisfied analytically given the choices of preferences). At each node, we compute relative forward differences in policies $c(\cdot)$, $\omega(\cdot|\theta)$, and $\omega_1(\cdot)$, i.e. the differences in a policy at $\theta''$ and at $\theta' < \theta''$ relative to the value of the policy at $\theta'$. To verify the nodes with numerical errors (the largest relative error is one one-thousands of 1 percent of the policy at the lower type), we then follow the procedure in subsection 7.2.2 in Kapíčka (2013) as an additional check of global incentive constraints, which amounts to letting the agent re-optimize with respect to reported type given the policies and verifying that the true type is a solution.\footnote{Abraham and Pavoni (2008) in subsection 3.3 and Farhi and Werning (2013) in subsection 2.2.4 describe applications in related settings.}

The next stage computes $\hat{w}_0$ such that $V_0(\hat{w}_0) = 0$ using binary search given $V_0$ computed in the first stage.

In the final stage, we simulate the optimal labor and savings distortions described in Section 3. Given $V_t$’s computed in the first stage and $\hat{w}_0$ solved for in the second stage, we generate optimal allocations by forward induction, starting from policy functions produced by $V_0(\hat{w}_0)$ from (11). Optimal distor-
Figure 8: An illustration of the typical effects on the optimal labor distortions of the changes in the Frisch elasticity parameter.

The distortions can then be computed from the policy functions using definitions (13) and (14). To compute the average distortions in Section 3 we do $5 \times 10^5$ Monte Carlo simulations. As a robustness check, Figure 6 here provides the analogue of Figure 4 in the main text plotted against the shock realizations. In addition, Table 2 summarizes the changes in aggregate earnings and consumption moments in the simulations discussed in the main text.

At this stage we also compute the objects whose limiting behavior is required by Assumption 4: $U_{c,t}(\theta)$, $\frac{c_t(\theta)}{y_t(\theta)}$, and $\frac{\hat{c}_t(\theta)/c_t(\theta)}{y_t(\theta)/y(\theta)}$. In the Monte Carlo histories we find that these expressions have finite numerical values of the same order of magnitude as the terms in Figure 3 in the main text, both in the left and right tails of the distribution. In a given period, the terms $\frac{c_t(\theta)}{y_t(\theta)}$ and $\frac{\hat{c}_t(\theta)/c_t(\theta)}{y_t(\theta)/y(\theta)}$ asymptote fairly quickly as $\theta \to \infty$, to virtually constant values at earnings above $300,000$. Relatedly, Figure 7 here further quantifies the intertemporal forces in Figure 3 in the main text. For the history of low earnings, Panel A in Figure 7 isolates intratemporal forces, displaying them
without the intertemporal terms, and Panel B provides an illustration of the effect of including intertemporal forces; Panels C and D illustrate the same for the history of high earnings.

We provide several further robustness checks and additional illustrations. First, we summarize the robustness checks with respect to a key fundamental, the Frisch elasticity of labor supply. We follow the same procedure we described for the baseline case of parameter \( \varepsilon = 2 \) in the main text, calibrating the same setup except with \( \varepsilon = 4 \) and then with \( \varepsilon = 1 \), which correspond to Frisch elasticities of 0.25 and 1 respectively. Table 3 compares the calibrated parameters for the initial distribution and the stochastic process for the shock in the three cases. The parameters are chosen to match the moments from the data displayed in Table 1 in the main text. In particular, lower Frisch
elasticities of labor supply (which correspond to higher values of \( \varepsilon \)) require lower maximum variance in the mixture, but drawn with higher probability to match the same data moments we discussed in the main text, particularly the high kurtosis.

We simulate the optimal distortions in the economies with \( \varepsilon = 4 \) and with \( \varepsilon = 1 \) and compare them to the baseline distortions: Figure 8 displays the typical effects, shown here for a representative history of twice the average earnings. Lower elasticities result in generally higher distortions, especially for the left part of the earnings distribution around the U-shapes. The right tail of the distribution displays the same pattern but with smaller differences.
because the effects of the higher parameter $\varepsilon$ are offset by the effects of the lower maximum variance in the mixture.

Next, to supplement the comparison with the static results of Saez (2001) in the main text, we illustrate here an experiment where a static model is simulated with the shock distribution given by our calibrated initial distribution, $F_0$. Figure 9 reproduces the labor distortions from our baseline simulation, analyzed in the main text with Figure 2, and compares them to the static distortions in the experiment. It is important to keep in mind, however, that the static model in which shocks are drawn from an initial-period Pareto-lognormal distribution understates the actual cross-sectional dispersion of shocks and leads to lower distortions, as Figure 9 indicates.

Finally, we make here transparent the role of kurtosis explored in the main text and illustrated with Figures 1 and 5. Figure 10 here provides an analogue of Figure 1 where we vary the kurtosis in the mixture distribution. The three distribution examples in Figure 10 illustrate the effects of increasing the level of kurtosis from 3 in the case of normal shocks (reproduced in Panels A and D from Figure 1) to the kurtosis of 6 (Panels B and E) and finally to 12 (Panels C and F). The rest of the parameters are kept unchanged compared to Figure 1.

References


