NBER WORKING PAPER SERIES

TRUST IN RISK SHARING: A DOUBLE-EDGED SWORD

Harold L. Cole
Dirk Krueger
George J. Mailath
Yena Park

Working Paper 26667
http://www.nber.org/papers/w26667

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2020, Revised August 2022

Cole, Krueger, and Mailath thank the National Science Foundation for research support (grants #SES-112354, 1260753, 1326781, 1559369, 1757084, 1851449), and Park thanks the New Faculty Startup Fund support from Seoul National University. Earlier versions were circulated under the titles “Coalition-Proof Risk Sharing Under Frictions” and “Social Capital: A Double-Edged Sword.” We thank the editor, referees, Loretta Mester, and the participants in the Wharton Macro lunch, European Summer Symposium in Economic Theory (Gerzensee), One World Mathematical Game Theory Seminar, 12th World Congress of the Econometric Society, 20th Annual Society for the Advancement of Economic Theory Conference, and many departmental seminars for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Trust in Risk Sharing: A Double-Edged Sword
Harold L. Cole, Dirk Krueger, George J. Mailath, and Yena Park
NBER Working Paper No. 26667
January 2020, Revised August 2022
JEL No. D15,D16,E20

ABSTRACT

We analyze efficient risk-sharing arrangements when the value from deviating is determined endogenously by another risk sharing arrangement. Coalitions form to insure against idiosyncratic income risk. Self-enforcing contracts for both the original coalition and any coalition formed (joined) after deviations rely on a belief in future cooperation which we term "trust". We treat the contracting conditions of original and deviation coalitions symmetrically and show that higher trust tightens incentive constraints since it facilitates the formation of deviating coalitions. As a consequence, although trust facilitates the initial formation of coalitions, the extent of risk sharing in successfully formed coalitions is declining in the extent of trust and efficient allocations might feature resource burning or utility burning: trust is indeed a double-edged sword.

Harold L. Cole
Economics Department
University of Pennsylvania
3718 Locust Walk
160 McNeil Building
Philadelphia, PA 19104
and NBER
colehl@sas.upenn.edu

Dirk Krueger
Economics Department
University of Pennsylvania
The Ronald O. Perelman Center
133 South 36th Street
Philadelphia, PA 19104
and NBER
dkrueger@econ.upenn.edu

George J. Mailath
University of Pennsylvania
Department of Economics
133 South 36th Street
Philadelphia, PA 19104-6297
gmailath@econ.upenn.edu

Yena Park
Seoul National University
1 Gwanak-ro, Gwanak-gu, 16-636
Department of Economics
Seoul 08826
South Korea
parkye82@gmail.com

Contents

1 Introduction

A large literature in economics, political science and sociology argues that *trust* is a critical determinant of communities' ability to cooperate. This literature provides evidence that trust differs systematically across cultures, countries, and time, and is correlated with economic prosperity. In economics, Tabellini (2008a) shows that measures of generalized trust towards others from the World Value Surveys predict well-functioning economic institutions at the level of countries or regions, and Tabellini (2010) demonstrates that trust is correlated with regional economic development across European regions. In political science, Fukuyama posits that "a nation's well-being, as well as its ability to compete, is conditioned by a single, pervasive cultural characteristic: the level of trust inherent in society" (Fukuyama, 1995, p. 7).[1] Fukuyama (2001, p. 7) provides the following definition: "trust is an instantiated informal norm that promotes co-operation between two or more individuals." Fukuyama's definition is consistent with the view that agents behave cooperatively because they expect their cooperation to be reciprocated in the future. In other words, cooperation requires a *shared* belief in future cooperation.

In this paper, we model *trust* as the ability to generate this shared belief and analyze theoretically its impact on the efficiency of social arrangements. The value of agents' outside options are endogenous, reflecting this shared belief, and the notion of efficiency is *second best*, capturing these endogenously determined incentive constraints. Consequently, higher trust need not imply greater social welfare and is a double-edged sword: agents in societies with more trust are both more likely to enter into beneficial arrangements but also more likely to find ways to circumvent such arrangements when profitable to do so, by forming cooperative coalitions with other members of the society at large after deviating, thus undermining the original arrangement.[2]

Central to our modeling is the assumption that deviations do not preclude agents from reaching beneficial arrangements that are exactly as attractive as the original arrangement. This symmetry assumption captures two distinct ideas. First, in a large society with significant anonymity, an agent who behaves opportunistically and deviates within one arrangement may be able to easily join another similar arrangement after being excluded from the original group following the deviation. For example, a worker who shirks at one firm may be able to obtain a similar job at a new firm after being fired. Second, a coalition of agents

[1] Fukuyama (1995) argues that "trust" is fundamental to the formation of large corporations and through this mechanism explains economic differences both across time and across countries.

[2] The premise of this paper that the notion of trust should accommodate its negative consequences has also been recently stressed in political science and sociology, see, for example, Portes and Landolt (1996), Putnam (2000), Woolcock (1998), Woolcock and Narayan (2000), and Woolcock (2001).

may deviate as a group, understanding that after the deviation, as a group they may be able to implement a beneficial arrangement. With these beneficial arrangement opportunities for deviating agents, supporting the original arrangement with the threat of adverse outcomes in case of defection is not credible.

To make this idea concrete, we study risk sharing in an infinite-horizon economy with idiosyncratic income risk.[3] By pooling income in each period, a coalition of agents can achieve higher ex ante utility for each agent. Such a cooperative agreement, however, requires currently rich agents to sacrifice current consumption. In the absence of commitment, the standard incentive device to induce cooperation by the rich is to exclude defectors from future insurance. But if agents were able to reach the original cooperative agreement, then there is also the possibility that rich agents deviate by leaving the current arrangement in the hope of, for example, replicating the current arrangement with other deviating rich agents.[4]

Since we are interested in the comparative statics of risk-sharing allocations and welfare with respect to the trust present in a society, we need to model how trust impacts the ability of a group of agents to reach cooperative agreements, based on a shared belief in future cooperation. We parameterize this ability in a stark fashion as the probability $\pi \in [0, 1]$ that a coalition coordinates beliefs on the most efficient allocation. With complementary probability, the absence of belief in coordination is permanent and there is no risk sharing.[5] In Section 2, building on Dixit (2004) and Tabellini (2008b), we show that this parameterization of trust emerges naturally in a simple matching model in which agents only cooperate with others that have sufficiently similar characteristics (which can be interpreted as sharing a similar culture, language or ethnicity).

What dynamic risk-sharing allocations emerge in society when trust in cooperation is modeled in this way? We say that an allocation is a *social norm* if it is robust to the possibility that an agent could defect, not contribute in the current period and (with probability π) "reinitialize" risk-sharing using the *same* allocation with other agents.[6] In order for an allocation to be credible as an agreement, it is clearly necessary that it be a social norm,

[3]The risk sharing application is chosen for concreteness. The modeling of trust and its impact on the efficiency of economic cooperation extends seamlessly to other applications, such as firms or other production networks (see Section S.2.2 in the appendix) or multinational organizations.

[4]Genicot and Ray (2003) also study the stability of risk sharing when joint deviations are possible. In contrast to our setting, Genicot and Ray (2003) study a finite society where it is impossible for a deviation to replicate the current arrangement. We discuss their paper in more detail in Section 3.

[5]The precise specification after a failure to coordinate beliefs is not important; it is important that the failure is costly. As illustrated by our analysis in Section S.2.1, our results are robust to alternative specifications.

[6]In the simple matching model of Section 2, a deviating agent can "reinitialize" using the same allocation with a currently unmatched agent. In the general model of Sections 4, if a deviation is profitable, a positive measure of agents will find it profitable, and so the original allocation is also feasible for the deviating set.

since if an agreed allocation is not a social norm, after some history, an agent will find it optimal to deviate, and after the deviating period follow that same *original* consumption allocation, undermining the credibility of the allocation.

The robustness requirement in a social norm is very weak, as illustrated by autarky (no risk sharing) being a social norm. A natural strengthening of the requirement is to require the allocation to be robust to the possibility that an agent could defect, not contribute in the current period and (with probability π) "reinitialize" risk-sharing using *any* similarly robust allocation. We call such an allocation a *strong social norm*. While natural, for high levels of trust π, this robustness requirement can be too demanding, in that strong social norms do not necessarily exist. We characterize the *constrained-efficient* social norms (i.e., the ex ante utility maximizing social norms), which always exist and coincide with the strong social norm when the latter exists.

A critical feature of the incentive constraints, and thus constrained efficient allocations, is that both sides of the constraint depend upon the allocation. As a consequence, the constraint set for the program determining constrained-efficient allocation is not convex, necessitating an indirect approach to characterizing these allocations.[7] Section 6 describes this general indirect approach, which focuses on maximizing ex ante utility subject to exogenous outside options. For low degrees of trust π, there is a fixed point characterization of the constrained-efficient social norm relating the value of the outside options and the maximized value of ex ante utility. In that case, efficient social norms are necessarily unique and equal the strong social norm (Proposition 2).

Proposition 3 characterizes general properties of strong social norms and Proposition 4 contains the central comparative statics results of constrained efficient allocations with respect to the degree of trust π in society. As long as households are sufficiently patient (the discount factor satisfies $\beta > \underline{\beta}$), there is a critical value of trust, $\overline{\pi}(\beta) \in (0, 1]$, such that for values of trust below this threshold, $\pi \leq \overline{\pi}(\beta)$, the fixed point characterization applies, and strong social norms can be determined using standard techniques.[8] A larger value of π reduces risk-sharing and lowers expected utility from a successfully formed coalition, strictly so if first-best insurance cannot be sustained. A critical ingredient in this comparative static is the characterization of the sense in which, when an exogenous outside option is binding, increasing the value of that outside option necessarily reduces risk sharing. Nonetheless, ex ante utility, the weighted sum of a successfully formed coalition (weight π) and an unsuccessful attempt at coordinating beliefs (weight $1 - \pi$), is strictly increasing in π.

[7] Since ex ante utility is continuous, and the set of social norms is compact (in the product topology), existence of a constrained-efficient allocation is immediate.

[8] If $\beta \leq \underline{\beta}$, the only social norm is autarky.

For values of trust above the threshold $\overline{\pi}(\beta)$ (whose determination we study in Section 7), the value of the outside option is so attractive that no allocations satisfy the stronger notion of robustness discussed above; that is, no strong social norms exist. We show in Section 8 that to prevent deviations, utility must be "burnt", either through introducing further inefficiencies in risk sharing (Proposition 7) or by burning resources (Proposition 8). The need for utility burning is strictly increasing in π, and ex ante utility remains at its maximal sustainable level as π rises from $\overline{\pi}(\beta)$ to 1.

The paper proceeds by developing first, in Section 2, a simple matching model of trust in risk sharing in which the key theoretical concepts and main results of the paper can be stated in its simplest form. Equipped with these results in Section 3 we then place the theory into the literature and discuss its empirical predictions. The theoretical analysis of the complete model laid out in Section 4 then proceeds by defining social norms in Section 5, characterizing strong social norms in Sections 6 and 7, and analyzing utility burning when strong social norms do not exist in Section 8. Section 9 concludes, and the appendix contains detailed theoretical derivations and proofs (Appendices A-D and S.1) as well as extensions (Appendix S.2) and numerical examples (Appendix S.3).

2 A Simple Matching Model of Trust

In this section we present a simple matching model to motivate our notion of trust in the full model, as well as to introduce our concept of a constrained efficient social norm and its fixed point characterization in what we think is simplest possible environment.[9]

There is a continuum of risk averse agents, $i \in [0,1]$, facing income risk in a discrete-time infinite-horizon environment. All agents have the same strictly concave period utility function u. An agent in each period has low income $y = \ell > 0$ and high income $y = h > \ell$ with equal probability. We write $Y := \{\ell, h\}$ and denote by $\overline{y} := \frac{1}{2}(\ell + h)$ per capita income.

For this section only, we make a critical simplifying assumption: At the beginning of each period (before income is realized), all unmatched agents are matched randomly into pairs, and within a matched pair, income shocks are perfectly negatively correlated (as in Krueger and Perri (2006) and Ábrahám and Laczó (2017)).[10] This assumption dramatically simplifies the analysis by allowing us to introduce the main ideas of this paper using stationary

[9] We thank a referee for encouraging us to study this simple model.

[10] The negative correlation can be obtained by, for example, assuming that agents can, at the beginning of the period choose a red or green endowment. The outcome of these endowments are determined by a common coin flip, with a heads giving the red endowment the high realization and the green the low; and a tails reversing this. Note that an agent alone does not care which technologies she chooses, and a matched pair simply wants to choose the opposite of each other, giving rise to the assumed perfectly negatively correlated incomes.

allocations in which the consumption of agents only depends on their current income realizations (and not on entire income histories or calendar time). An additional implication of the assumptions in the simple model is that the relevant incentive constraints only concern unilateral deviations, since risk sharing occurs within a pair. The full model in Section 4 drops these restrictions, and the end of this section explains why, even in the context of the simple model, the restriction to stationary allocations can be unduly restrictive.

Each pair of matched agents attempts to reach an agreement on risk sharing where the currently rich agent forgoes current consumption in exchange for the promise of future income insurance. The difficulty is that continued participation in the agreement is voluntary and risk sharing requires a belief in future cooperation. In our view, this belief is not guaranteed. Agents are more likely to trust family members, neighbors, attendees of the same church, and less likely to trust people they have little in common with (such as strangers or foreigners). Moreover, agents are less likely to trust partners who had betrayed an earlier trust.

Trust is a social phenomenon, reflecting the behavior of all members of society, and so is a property of society. We model trust as the likelihood that any matched pair of agents coordinates on the most cooperative allocation rather than behaving opportunistically. While it is beyond the scope of this paper to microfound trust in the general model with large coalitions, we can provide a specific microfoundation for the model in this section. Following Dixit (2004) and Tabellini (2008b), we assume each agent has a permanent idiosyncratic characteristic $\theta_i \in \mathbb{R}$ and introduce a parameter Θ quantifying the level of trust in society. If the matched pair has characteristics θ_i and θ_j satisfying

$$|\theta_i - \theta_j| \leq \Theta,$$

then the partners agree to share risk, and we call such a pair *compatible*. But if the inequality fails, then the agents do not trust each other, no agreement is reached, the pair remains unmatched, and both agents consume their incomes and try again next period with new partners. In evaluating the benefits of cooperating, a θ_i-agent assigns a probability of reaching an agreement with a new draw from the unmatched pool of

$$\pi(\theta_i) = \Pr\{|\theta_i - \theta_j| \leq \Theta \mid \theta_i\}.$$

To aid interpretation and exposition, assume θ_i is uniformly distributed on the circle, so that $\pi' := \pi(\theta_i)$ is independent of θ_i.[11] We interpret this common probability π' of trusting each other and reaching an agreement as society's (level of) *trust*.

[11] Alternatively, we could assume that θ_i is drawn from an improper uniform distribution with support $\mathbb{R}$; this again implies the conditional probability is independent of θ_i. While nonstandard, this assumption has

To keep the share of unmatched agents in the population constant, assume that each agent survives to the next period with some probability and dies with complementary probability. Also assume that if a partner within an ongoing risk-sharing coalition dies, so does the other partner. Finally, a mass of new unmatched agents equal to those who died is born each period and enters the unmatched pool. The combination of impatience and death implies an effective discount factor of $\beta \in (0,1)$.

Consider first the problem facing a compatible pair. While the agents are initially willing to trust each other, this trust is not blind. It is natural to assume that a failure to cooperate results in the breakdown of the agreement, with both agents being returned to the unmatched pool. This implies that any agreement must have the property that the lifetime continuation utility of a currently rich agent (after any history) must be at least the lifetime utility from failing to cooperate and then receiving F, the expected lifetime utility of an unmatched agent.[12]

Suppose a compatible pair treat F as exogenous. If risk sharing is feasible, then it is well known that the optimal allocation constrained by an exogenous outside option in the simple model with perfectly negatively correlated income shocks is income-history independent (see Appendix A for a proof) and so characterized by a single transfer x with stationary consumption given by:

$$c(y^t) = \begin{cases} h - x, & \text{if } y_t = h, \\ \ell + x, & \text{if } y_t = \ell, \end{cases} \tag{1}$$

and expected lifetime utility

$$V(x) := \tfrac{1}{2}[u(h-x) + u(\ell + x)]. \tag{2}$$

The first-best allocation is stationary and given by $x^{FB} := \frac{1}{2}(h-\ell)$, with an expected lifetime utility of $V^{FB} := u(\overline{y})$. Autarky is given by $x^A := 0$, with an expected lifetime utility of $V^A := Eu(y) = \frac{1}{2}(u(\ell) + u(h))$.

Given the constrained optimality of stationary consumptions, we *temporarily* restrict attention to stationary allocations. The constrained optimal transfer $x(F)$ is the transfer x maximizing $V(x)$ subject to

$$\Gamma(x) := (1-\beta)u(h-x) + \beta V(x) \geq (1-\beta)u(h) + \beta F. \tag{3}$$

previously been used in the global games literature to simplify exposition (see, in particular, the discussion in Morris and Shin (2003)).

[12] In the current setting, the currently poor agent will never fail to cooperate. This is not true in the general model: The constraint that currently poor agents wish to cooperate does bind for some poor agents.

Observe that the infinite collection of constraints that, after any income history, the lifetime continuation utility of a currently rich agent is at least the lifetime utility from failing to cooperate and then receiving F is replaced by the single stationary constraint (3).

Of course, F is not exogenous, but is determined by the agreements reached by other compatible pairs. If all compatible pairs agree to share risk via the same stationary transfer x, then F depends on x and satisfies

$$\begin{aligned} F &= \pi' V(x) + (1-\pi')\left[(1-\beta)V^A + \beta F\right] \\ &= \frac{1}{1-\beta(1-\pi')}\left[\pi' V(x) + (1-\pi')(1-\beta)V^A\right] \\ &= \pi V(x) + (1-\pi)V^A =: F(x), \end{aligned} \tag{4}$$

where $\pi := \pi'/[1-\beta(1-\pi')]$. Note that π is a strictly increasing function of π' mapping $[0,1]$ onto $[0,1]$, and so we also refer to π as *trust* henceforth.[13] If $V(x) > V^A$, then $F(x) \in (V^A, V(x))$ for $\pi \in (0,1)$. If instead $\pi = 0$, then the outside option is exogenous and equal to autarkic utility, i.e., $F = V^A$, and our model effectively collapses to the standard limited commitment model studied in much of the literature, which is discussed in the next section.

If all compatible pairs agree to share risk via the same stationary transfer x, then such an agreement must provide sufficient incentives for the currently rich agent to cooperate, i.e., $\Gamma(x) = (1-\beta)u(h-x) + \beta V(x) \geq \Psi(x;\pi)$, where

$$\Psi(x;\pi) := (1-\beta)u(h) + \beta[\pi V(x) + (1-\pi)V^A], \tag{5}$$

We call such agreements *social norms*.

Definition 1 *A stationary transfer x is a* social norm *if $\Gamma(x) \geq \Psi(x;\pi)$.*

A social norm x is a stationary transfer that satisfies an *internal* notion of incentive compatibility: If all compatible pairs reach the same agreement x, then it must be that case that the transfer x does not induce an agent to deviate when confident that the same agreement x would be reached with any new compatible partner.[14] Trivially, autarky ($x = 0$) is always a social norm. As we will see, typically there is a plethora of social norms.

[13] Strictly speaking, π is the *normalized* trust, but $\pi = 0$ if and only if $\pi' = 0$, $\pi = 1$ if and only if $\pi' = 1$, and increases in one imply increases in the other.

[14] The one-shot deviation principle holds here: If x is a social norm, then it is also not optimal for a rich agent to deviate, planning when matched compatibly to always similarly deviate if rich and consume $\ell + x$ if poor.

The requirement that the stationary transfer be a social norm is clearly necessary for everyone to agree to that risk sharing arrangement, but it may be too weak. In particular, the notion of a social norm leaves open the question of whether a compatible pair could do even better. Our next notion requires that compatible pairs cannot do better.

Definition 2 *The transfer x^* is a* strong social norm *if x^* maximizes $V(x)$ subject to* (3) *when $F = F(x^*)$.*

A strong social norm x is a stationary transfer that satisfies an *external* notion of incentive compatibility: A transfer is a strong social norm if, when every compatible pair agrees to share risk via that transfer (which determines the outside option F), then that agreement is the best risk sharing arrangement respecting the outside option determined by F, and so is the agreement that would be reached by any compatible pair. Note that the strict concavity of u implies that if a strong social norm exists, it must be unique.

We now turn to the characterization and existence of strong social norms. When all other compatible pairs reach the agreement x, the constraint (3) on possible agreements $\tilde{x}$ can be rewritten as

$$\Gamma(\tilde{x}) \geq \Psi(x;\pi). \tag{6}$$

The strong social norm is the unique fixed point $x(\pi)$ of the mapping

$$\mathcal{T}^X(x;\pi) = \underset{\{\tilde{x}:\Gamma(\tilde{x})\geq\Psi(x;\pi)\}}{\arg\max} V(\tilde{x}). \tag{7}$$

If the mapping $\mathcal{T}^X(\,\cdot\,;\pi)$ does not have a fixed point, then there is no strong social norm (at that level of trust π). If $\pi = 0$, the probability of a compatible pair being formed is zero. *But*, if such a pair were to form, one would expect the agreement reached by patient agents would involve significant risk sharing; the case $\pi = 0$ is the focus of much of the literature on risk sharing with limited commitment. Finally, observe that $\mathcal{T}^X(\,\cdot\,;0)$ trivially always has a fixed point.

2.1 Risk Sharing in the Strong Social Norm

Since the expected utility $V(x)$ inside a coalition is strictly increasing in the transfer $x \leq x^{FB}$, (7) immediately implies that either x^{FB} satisfies the incentive constraint (6) at $x = x^{FB}$ and full risk sharing (achieving the value of first-best insurance $V^{FB} = u(\overline{y})$) is the strong social norm, or the constraint is binding at some $x < x^{FB}$ and the strong social norm x equates $\Gamma(x)$ and $\Psi(x;\pi)$.

We first turn to the question of when first-best risk sharing constitutes a strong social norm. A straightforward calculation determines bounds on β and π that imply (6) is satisfied at $\tilde{x} = x = x^{FB}$ (i.e., $\Gamma(x^{FB}) \geq \Psi(x^{FB}, \pi)$):

Proposition 1 *First-best insurance x^{FB} with value $V^{FB} = u(\overline{y})$ is a strong social norm if and only if*

$$\pi \leq \pi^{FB} := 1 - \frac{(1-\beta)[u(h) - V^{FB}]}{\beta\,[V^{FB} - V^{A}]} < 1. \tag{8}$$

and the largest outside option F^{FB} consistent with first-best insurance satisfies

$$F^{FB} := \pi^{FB} V^{FB} + (1 - \pi^{FB}) V^{A} = \frac{V^{FB} - (1-\beta)u(h)}{\beta}. \tag{9}$$

Moreover, $\pi^{FB} < 0$ and first-best insurance is not a strong social norm for any $\pi \in [0,1]$ if and only if

$$\beta < \beta^{FB} := \frac{u(h) - V^{FB}}{u(h) - V^{A}}. \tag{10}$$

Finally, $F^{FB} > V^{A}$ if and only if $\beta > \beta^{FB}$.

The requirement that trust not be too large for first-best insurance to be a strong social norm should not be surprising as the currently h-income agents sacrifice current consumption to insure the currently ℓ-income agents. If π is close to one, deviating and then attempting to rematch incurs almost no loss in insurance and so is attractive, preventing first-best insurance from being a social norm.

At the other extreme, if agents are sufficiently impatient, the autarkic allocation ($x = 0$) is the only strong social norm (irrespective of the value of π): It is straightforward to verify that the incentive constraint (6) at $x = 0$ is violated for every positive $\tilde{x}$ if and only if

$$\beta \leq \frac{2u'(h)}{u'(\ell) + u'(h)} =: \underline{\beta}. \tag{11}$$

In this case, it is immediate that autarky is the strong social norm.

We now analyze the intermediate case $\beta \in (\underline{\beta}, \beta^{FB})$. The functions V, Γ, as well as $\Psi(\,\cdot\,;\pi)$ for different values of π are depicted in Figure 1.[15] Starting with $\pi = 0$, the strong social norm $x^0 > 0$ is determined by the intersection of the $\Gamma(x)$ curve and the $\Psi(x;0)$ curve. This is the largest transfer x that satisfies the constraint $\Gamma(x) \geq \Psi(x,0)$, it is positive (as $\beta > \underline{\beta}$) but smaller than first-best insurance (as $\beta < \beta^{FB}$). It is worth noting that while

[15]The calculations are in Appendix S.1. The appendix also includes figures of the case $\beta \geq \beta^{FB}$ and thus $\pi^{FB} \geq 0$, as well as for the case $\beta \leq \underline{\beta}$ (in which case $\Gamma(x)$ is downward sloping at $x = 0$, and thus intersects the $\Psi(x;\pi)$ curves only at $x = 0$).

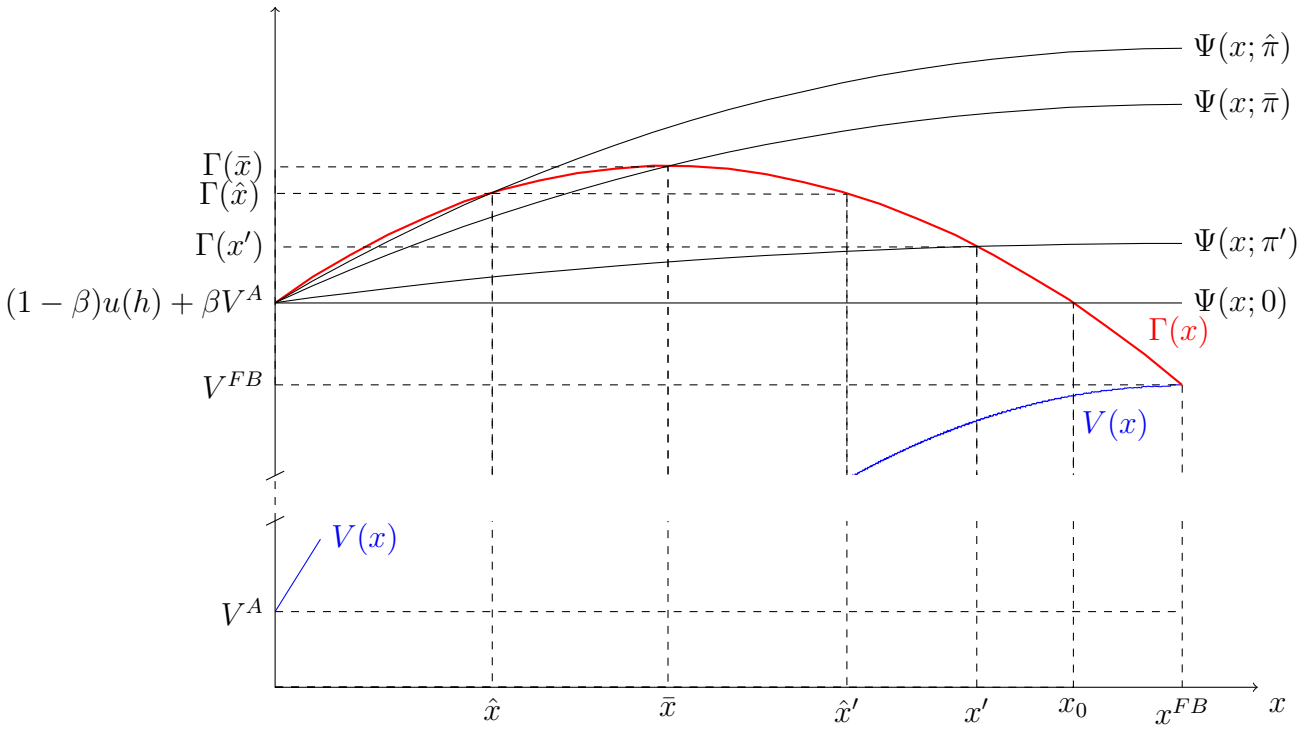

Figure 1: The functions V, Γ, and Ψ when $\underline{\beta} < \beta < \beta^{FB}$, for the parameterization $u(c) = \log(c)$, $\ell = 1-\epsilon$, and $h = 1+\epsilon$, and $0 < \pi' < \bar{\pi} < \hat{\pi} < 1$. Since V^A is much smaller than V^{FB}, the middle of the y-axis has been omitted.

$x = 0$ also equates $\Gamma(x)$ and $\Psi(x, 0)$, it is not the strong social norm because $\tilde{x} = 0$ does not maximize $V(\tilde{x})$ over those transfers satisfying $\Gamma(\tilde{x}) \geq \Psi(0, 0)$. The resulting allocation (implied by x_0) is exactly the one emerging in the standard limited commitment literature (since at $\pi = 0$ the outside option is exogenous), and thus fully mirrors the results in the two agent economies of, e.g., Kehoe and Levine (2001) and Krueger and Perri (2006).

Now consider larger values of trust π. As π increases, $\Psi(x;\pi)$ becomes steeper, with the value of $\Psi(0;\pi)$ unchanged. The strong social norm $x = x(\pi)$ (which is the strictly positive transfer equating $\Gamma(x)$ and $\Psi(x;\pi)$) strictly decreases as π increases. In other words, as π increases from 0, risk sharing falls (i.e., $V(x(\pi))$ is strictly decreasing in π). Nonetheless, the expect payoff of an unmatched agent ($F(x(\pi)) = (\Gamma(x(\pi)) - (1-\beta)u(h))/\beta$) is strictly increasing in π. This is the sense in which trust is a double-edged sword: an increase in π tightens the incentive constraint and thus reduces the ability of compatible pairs to share risk in coalitions that have successfully formed by reducing x, but raises the ability of society to successfully form coalitions in the first place.

Denote by $\overline{F}$ the largest value of F for which the set of transfers x satisfying (3) is nonempty, so that

$$(1-\beta)u(h)+\beta\overline{F}=\max_x \Gamma(x).$$

Set $\overline{x}=\arg\max_x \Gamma(x)$ and denote by $\overline{\pi}$ the value of trust that satisfies $\Psi(\overline{x};\overline{\pi})=(1-\beta)u(h)+\beta\overline{F}$. A strong social norm exists and the comparative statics in the previous paragraph on π apply for $\pi\leq\overline{\pi}$.

For $\hat{\pi}>\overline{\pi}$, a strong social norm does not exist (in Appendix S.1 we fully characterize $(\overline{F},\bar{x},\overline{\pi})$ for the parametric example of Figure 1; crucially, we show that $\overline{\pi}<1$). While there is a strictly positive transfer $\hat{x}$ at which $\Gamma(\hat{x})=\Psi(\hat{x};\hat{\pi})$, as is clear from Figure 1, that value of x does not maximize $V(x)$ subject to $\Gamma(x)\geq\Psi(\hat{x},\hat{\pi})$; the maximizing value of x is $\hat{x}'$.[16]

2.2 Nonexistence of the Strong Social Norm

What are we to make of the nonexistence of strong social norms for $\pi>\overline{\pi}$? We believe that this nonexistence reflects a problem with the notion of strong social norm. Indeed, such nonexistence is to be expected from an overly demanding notion of social norm, since this notion requires robustness against *incredible* deviations. Consider again $\hat{\pi}>\overline{\pi}$. The allocation $\hat{x}$ fails to be a strong social norm because it does not maximize $V(\tilde{x})$ subject to (6) (at $x=\hat{x}$). Suppose a compatible pair does agree on the allocation $\hat{x}'$ that maximizes $V(x)$ subject to the same constraint. This implies that the agents believe that the stationary transfer $\hat{x}'$ will be implemented in the future. But when the current income uncertainty is resolved, surely the rich agent will now deviate (consuming h this period), return to the unmatched pool with the expectation that any agreement reached with a new compatible partner will be at $\hat{x}'$, not $\hat{x}$ (as required by the putative strong social norm). Since both partners in a compatible pair understand this, the belief that the stationary transfer $\hat{x}'$ will be implemented in the future has been undermined.

The previous paragraph motivates a key point: If a strong social norm exists, then the strong social norm is a compelling description of social behavior. But if a strong social norm does not exist, we still have the persuasive notion of a social norm. As is clear from Figure 1, for $\pi>\overline{\pi}$, there are many social norms. We focus on the constrained efficient social norm.

Definition 3 *The* constrained-efficient stationary social norm *is the stationary transfer, $\hat{x}$, that maximizes $V(x)$ subject to $\Gamma(x)\geq\Psi(x;\pi)$.*

[16]Note that if $\beta\leq\underline{\beta}$, then such an allocation $\hat{x}>0$ does not exist ($\Gamma(x)$ is monotonically declining in x) and autarky is indeed the unique fixed point, and trivially the strong social norm, as asserted in the text. This is consistent with Krueger and Uhlig (2006), who show that in an economy equivalent to our $\pi=1$ case but with storage, if the storage return is sufficiently low (which amounts to the assumption $\beta\leq\underline{\beta}$ in our model), then there exists a fixed-point social norm and it is autarkic.

While the restriction to stationary transfers (and so consumption allocations) is without loss of generality when analyzing strong social norms, it is restrictive when considering the constrained-efficient social norm. If we restrict attention to stationary allocations, ex ante welfare (the expected utility of an unmatched agent) is uniquely maximized at $\overline{\pi}$, and is strictly *decreasing* in π for $\pi > \overline{\pi}$ (this is immediate from Figure 1). In particular, for $\hat{\pi} > \overline{\pi}$, the constrained efficient stationary social is given by the transfer $\hat{x}$, and as $\hat{\pi} \to 1$, $\hat{x} \to 0$.

2.3 Nonstationary Allocations

The restriction to stationary allocations gives a misleading picture of the comparative statics of ex ante welfare for high trust. In particular, we will informally argue here that for the simple model, welfare in the constrained efficient social norm is necessarily nondecreasing in trust. We provide a formal proof of this result for the general model in Section 8.

We first construct a nonstationary allocation for $\hat{\pi} > \overline{\pi}$ that has the same ex ante welfare as the strong social norm at $\overline{\pi}$. The idea (familiar from the efficiency wage literature discussed in the next section) is that by postponing risk sharing, the nonstationary allocation makes deviations less attractive, since a deviation only delays the start of risk sharing. Recalling that $\bar{x}$ maximizes $V(\tilde{x})$ subject to $\Gamma(\tilde{x}) \geq \Psi(\bar{x}; \overline{\pi})$, the inequality constraint holds as an equality at $\overline{x}$ and so, since $\hat{\pi} > \overline{\pi}$ and $V(\overline{x}) > V^A$,

$$\hat{\pi} V(\overline{x}) + (1 - \hat{\pi}) V^A > \overline{F}.$$

The nonstationary allocation defers risk sharing for a number of periods T followed by a transition period and then risk sharing with transfers $\bar{x}$. The delay $T \geq 0$ and transition transfer $x^\dagger$ in period $T + 1$ are chosen so that

$$\hat{\pi} \left((1 - \beta^T) V^A + \beta^T (1 - \beta) \tfrac{1}{2} [u(h - x^\dagger) + u(\ell + x^\dagger)] + \beta^{T+1} V(\overline{x}) \right) + (1 - \hat{\pi}) V^A = \overline{F}. \quad (12)$$

By construction, the expected utility of unmatched agent is $\overline{F}$, for any $\hat{\pi} > \overline{\pi}$.

Informally, a *nonstationary* consumption allocation is a social norm, if after every history, no agent has an incentive to deviate when after the deviation, a new compatible match results in restarting the same allocation (see Definition 5). From (12), after any history of more than T periods within the risk sharing arrangement, the rich agent is indifferent between continuing with the arrangement and deviating. In the first $T + 1$ periods, the benefit from deviating is less (indeed, if $T > 0$, in the first T periods there is no benefit from deviating), so the allocation is a social norm.

We claim this consumption allocation with delay is the constrained efficient social norm: Note that any social norm must satisfy the nonstationary version of (3) for F given by (4) (replacing $V(x)$ by the expected discounted utility of the social norm). The set of consumption allocations that satisfy the nonstationary version of (3) must then be nonempty at F. As noted earlier, the maximum of expected utility over this set is achieved by a stationary consumption allocation. But since $\overline{F}$ is the largest value of F for which the constraint set with stationary consumptions is nonempty, we have $F \leq \overline{F}$. Since F is strictly increasing in the expected discounted utility of the social norm, maximizing the latter is equivalent to maximizing the former, and so any social norm with the property $F = \overline{F}$ is constrained efficient.

In Section 4 we demonstrate these results in the general model of risk sharing with large coalitions in which the assumption of perfectly negatively correlated incomes and implied stationarity of social norms is relaxed. First, however, equipped with the concepts and results from the simple model we now relate our paper to the theoretical literature and discuss how the predictions of the model square with the existing empirical evidence.

3 Relation to the Literature and Empirical Predictions

3.1 Related Theoretical Literature

Apart from the general literature on the impact of trust in society on economic outcomes discussed in the introduction, our paper more specifically builds on the theoretical literature in macroeconomics that derives imperfect consumption insurance from limited commitment. As explained in the Introduction, and in contrast to that literature, we explicitly allow all agents to have access to the same risk sharing possibilities, both in forming the original agreement and after a deviation. This distinction in turn underlies the stark differences in resulting outcomes when trust is high, i.e., the nonexistence of a fixed point and utility burning when $\pi > \bar{\pi}$. Note also that that literature focuses on the two boundary cases of $\pi = 0$ and $\pi = 1$.

The classic limited commitment literature pioneered by Harris and Holmström (1982), Thomas and Worrall (1988), Kehoe and Levine (1993), Kocherlakota (1996), and Alvarez and Jermann (2000) assumes, explicitly or implicitly, that deviators have less opportunities than members of the originally formed coalition. Krueger and Perri (2006) extend this literature to a risk sharing economy with as a continuum of households exactly of the form studied in the remainder of this paper and Ábrahám and Laczó (2017) incorporate a private storage technology. These papers share our focus on self-enforcing arrangements, but take the outside

option for those that share risk as exogenously given, and typically equal to autarky. In the context of our paper, this amounts to assuming $\pi = 1$ in the original coalition and, crucially, $\pi = 0$ for deviating coalitions, resulting in permanent autarky for its members. *Given* this outside option, the qualitative properties of the constrained-efficient allocation in Section 6 when $\pi \leq \overline{\pi}$ will be similar to the literature: high-income agents receive high consumption to deter defection, and consumption drifts down a ladder with each subsequent low-income realization until it hits a lower bound. One contribution of our paper to this literature is to provide a theoretical convergence result: the constrained-efficient allocation converges over time to a stationary ladder, with declining risk sharing over time.

Our paper is most closely related to the limited commitment literature with *endogenous* outside option. Krueger and Uhlig (2006) assume that the outside option is determined by the best insurance contract offered by a competing financial intermediary, who has long term commitment.[17] The only punishment for deviating countries in Hellwig and Lorenzoni (2009) and its extension in Martins-da Rocha and Santos (2019) is the denial of future credit; they are allowed to save with a "Swiss banker", in the tradition of limited commitment in the sovereign debt literature pioneered by Bulow and Rogoff (1989).[18] These two papers also define equilibrium as a fixed point, but unlike in our paper, the nonexistence issue does not emerge. The central difference to our paper is that they assume asymmetric contracting conditions between the original agreement and after a deviation. With such an asymmetric treatment, there is more room for relaxing incentive constraints through adjustments of endogenous variables that have differential effects on the payoff before and after a deviation. In Krueger and Uhlig (2006), this is the effective loss of storage upon deviation, and in Hellwig and Lorenzoni (2009), default removes the possibility of borrowing (which pre-default was limited by an endogenous borrowing constraint) but not of savings. This asymmetry in Hellwig and Lorenzoni (2009) supports insurance because endogenously low equilibrium interest rates make switching from borrowing to savings after a default unattractive. With the symmetric treatment of contracting conditions in our paper, the ex ante value and the deviation payoff are themselves related by a fixed point, which generates a strong feedback on risk sharing opportunities from the outside option of a coalition deviation.

[17] Phelan (1995) also endogenizes the outside option. His timing assumptions imply full commitment for one period, and private information about income limits consumption insurance.

[18] In a separate literature, the outside option is endogenous from the perspective of a policy maker whose choice of policies (unemployment insurance, progressive taxation, disability insurance, monetary policy) impacts the outside option and thus equilibrium private insurance, see e.g. Aiyagari and Williamson (2000), Krueger and Perri (2011), and Park (2014). A related literature which endogenizes the outside option by assuming that private noncontingent intertemporal trades can be enforced and examines how this interacts with insurance (see, for example, Allen, 1985) or government taxation (see Farhi, Golosov, and Tsyvinski (2009)). In the context of private information, Cole and Kocherlakota (2001) endogenize the outside option with hidden storage, and Golosov and Tsyvinski (2006) analyze optimal disability insurance.

Our findings stand in contrast to the original Bulow and Rogoff (1989) result in that we obtain risk sharing even when $\pi = 1$, an outcome that is not feasible in the Bulow and Rogoff (1989) environment. In our model, when $\pi \geq \bar{\pi}$, and given the symmetric contracting conditions, the mechanism supporting insurance is utility burning at the beginning of the coalition. This can be best understood by comparing the results with the stationarity assumption (which precludes utility burning) to those without it. When $\pi = 1$, the constrained efficient *stationary* social norm is the Bulow-Rogoff zero transfer result, whereas the nonstationary allocation with initial utility burning achieves positive insurance. In fact, in the simple stationary example of Hellwig and Lorenzoni (2009) and in our simple model at $\pi = \bar{\pi}$, the risk sharing achieved with the transfer $\bar{x}$ is exactly the same in the two economies despite the very different mechanisms of supporting this insurance. At $\pi = 1$, the insurance with the transfer $\bar{x}$ in our model is postponed to achieve the necessary utility burning in the *nonstationary* constrained-efficient allocation.

The previously reviewed literature does not allow the value from deviating to be determined endogenously by another risk sharing arrangement, thereby limiting the extent of insurance that can be obtained after deviating. An exception is Genicot and Ray (2003) and its application in Bold and Broer (2021), who study the formation and stability to *joint deviations* of risk sharing coalitions in economies with finite populations. In their model coalitions must be stable against deviations of smaller sub-coalitions of the original group, and the main purpose of the paper is to determine the endogenous *size* of stable coalitions.[19] Since larger coalitions are more prone to successful deviation, an optimal size of the original coalition emerges. This result stems from their assumption that the deviating coalition can only make an arrangement with the original coalition members, while in the formation of the original coalition, all members of the population could be considered as potential members. Bold and Broer (2021) quantitatively evaluate this model by estimating it on Indian village data.[20] We share with these papers the basic notion that any risk-sharing agreement must be robust to possible future risk sharing by any set of deviating agents. In contrast to Genicot and Ray (2003), however, we allow the deviating agents to have precisely the same insurance capabilities as the original coalition.

[19]Genicot and Ray (2003) builds on a more abstract game theoretic literature on coalition deviations pioneered by Bernheim, Peleg, and Whinston (1987) and Greenberg (1990), and extended and unified by Kahn and Mookherjee (1992, 1995) to infinite games and adverse selection insurance economies. This abstract literature shares with Genicot and Ray (2003) the assumption that coalition formation is "easy", i.e., $\pi = 1$.

[20]They find that stable risk sharing coalitions are typically small, and that the resulting consumption allocations—especially the symmetric response of consumption to income shock—accord better with the Indian village data than those generated by the standard limited commitment model with an autarkic outside option.

Conceptually, our notions of social norms and strong social norms are reminiscent of notions in both cooperative and noncooperative game theory. A social norm is "free of internal contradictions" and so is similar to von Neumann and Morgenstern's (1944) *internal stability* notion; see the discussion in Greenberg (1990, Section 2.3). It also resembles, in the theory of repeated games, Farrell and Maskin's (1989) notion of *weakly renegotiation proof* and Bernheim and Ray's (1989) notion of *internal consistency*. The strong social norm cannot be "discredited" or "dominated" by any other risk-sharing arrangement and so this notion is analogous to von Neumann and Morgenstern's (1944) *external stability*; again see the discussion in Greenberg (1990, Section 2.3). It is also analogous to *strongly renegotiation proof* in Farrell and Maskin (1989) or *strong consistency* in Bernheim and Ray (1989). Note that these authors all effectively assume $\pi = 1$.

The nonexistence of a strong social norm under the stronger robustness notion and the associated need for utility burning is a general phenomenon. The use of utility and money burning at the beginning of the allocation is reminiscent of efficiency wage (Shapiro and Stiglitz, 1984, MacLeod and Malcomson, 1989) and gift-exchange and related models (Carmichael and MacLeod, 1997, Kranton, 1996a,b, Ghosh and Ray, 1996). In particular, the idea that if it is too easy to start a new relationship (worker-firm, principal-agency, partnership, etc) after opportunistic behavior (shirking for example), then it is impossible to deter opportunistic behavior. In order to deter deviations, it is therefore necessary to impose some form of friction (such as delays in joining a new firm, involuntary unemployment, or engaging in inefficient actions in the beginning of the new relationship, exchange of inefficient gifts).

3.2 Empirical Predictions and Relation to the Applied Literature

The main predictions of the model state that a larger level of trust Θ and, thus, a larger belief in cooperation π leads to better ex-ante outcomes and welfare (up to a point, $\bar{\pi}$ in the model), but results in less ex-post risk sharing.

Our results for low values of trust Θ (respectively, π) are consistent with the results from Alesina and Giuliano (2014) indicating that trust systems based on narrow kinship or tribal relationships are negatively correlated with generalized trust and therefore are detrimental to overall cooperation, economic efficiency and development.[21] In a similar vein racial fragmentation has been empirically shown to hinder growth in a cross-section of countries (Easterly and Levine (1997)) because it impedes generalized trust in society.

[21]This explanation has also been suggested for the observation that the basic institutions of modern life, such as local government, schools, courts, and hospitals function better in the North of Italy relative to the South despite the common administrative and legal structure (see Edward (1958), Leonardi et al. (2001)).

The implication of our model that agents with a small capacity to trust can achieve remarkable cooperation within a successfully formed coalition but are not able to extend cooperation outside of a narrow group is consistent with Greif (1993)'s classic study of the Maghribi traders. This tightly knit group of Jewish medieval merchants was able to form long-lasting cooperative arrangements over large distances. Their trading was largely restricted to members of their own community, not even extending readily to other Jewish merchants, despite seemingly profitable reasons for doing so.[22]

Our simple model predicting that a reduction in trust reduces ex-ante cooperation and welfare is also consistent with the evidence that factors which destroy such trust can have long-term negative economic consequences. Nunn and Wantchekon (2011) show that measures of high slave exports in African countries between 1400–1900 correlate with micro-based surveys of trust, and Nunn (2008) shows that these measures correlate with low growth and poor public policies.[23]

The previous papers provided supporting qualitative evidence for the predictions of our model from specific case studies. Furthermore, the prediction that ex ante welfare is increasing in π is also consistent with systematic empirical evidence showing that trust is a key component of economic differences across time and across countries. The most commonly used trust indicator from the World Values Survey (WVS) measures trust of overall people in the society. Knack and Keefer (1997) show that indicators of trust and civic norms from the WVS are positively correlated with income and negatively correlated with the dispersion in income in a sample of 29 market economies. Zak and Knack (2001) confirm these results even after controlling for measures of quality of government. For review of the impact of trust on various economic outcomes, see Algan and Cahuc (2014).

One key prediction of our model is that trust π is a double-edged sword when it comes risk sharing and welfare. Higher level of trust implies smaller amount of risk sharing conditional on successfully forming a coalition, and when the level of trust is very high (above $\bar{\pi}$), higher level of trust no longer implies higher ex ante welfare. Roth (2009) provides empirical evidence for this prediction. He extends the regression analysis of Knack and Keefer (1997) and Zak and Knack (2001) to include a quadratic trust term in his panel of 41 countries over the time period from 1980 to 2004. His estimates implies a concave statistical relationship between trust and economic performance, leading him to conclude that "an increase in trust

[22]Within the Maghrib community of traders, only descendants of the original community could become members. To generate cooperative outcomes within this group, the Maghribi traders developed a communal system of information sharing and punishment, with exclusion from trade being the standard punishment.

[23]Nunn and Wantchekon (2011) note that: "This finding is consistent with the historical fact that by the end of the slave trade, it was not uncommon for individuals to be sold into slavery by neighbors, friends and family members" (p. 3222).

is crucial for countries with low levels of trust, but can likely be neglected by countries with sufficient levels of trust and may even hamper economic performance in countries with high levels of trust" (abstract, p. 141).[24]

The key mechanism through which higher trust in our model reduces cooperation in successfully formed coalitions works through an endogenously tightened outside option. Evidence for this comes from the empirical literature on endogenous risk-sharing in poor agricultural villages pioneered by Townsend (1994) and Udry (1994). The paper by Bramoullé and Kranton (2007) finds that when risk sharing relationships across villages are possible, the extent of risk-sharing within a village is reduced, leading to potentially ambiguous effects on welfare. Morten (2019) shows that the option to temporarily migrate and work in the city is also detrimental to risk sharing within the village.[25]

Finally, our model is also consistent with other features of risk sharing in the real world. First, the optimal allocation in our full model features front-loading of consumption, which is consistent with Morten (2019)'s finding that transfers in the village risk sharing depend negatively on the history of past transfers. Second, when the level of trust is low, the economy shows a relatively symmetric consumption response to income shocks, with very high level of risk sharing in the successfully formed coalition (occurring with low probability) and the autarkic allocation in case of failed formation (realized with high probability). This is consistent with the symmetric response observed in the village data studied by Bold and Broer (2021). Third, our model predicts that even when the level of trust is very high and defection is "easy" (e.g., for $\pi = 1$), coalitions still achieve some risk sharing, consistent with real world risk-sharing, and in contrast to the Bulow-Rogoff prediction.

Although our simple model provides a useful framework to analyze the role of trust on ex-post risk sharing and ex-ante welfare, its focus on pairwise insurance and stationary allocations is restrictive. First, the empirical literature on trust and economic performance from the cross-country analysis has emphasized the effect of trust within larger organizations. The restriction to pairwise matching is too limited when exploring the role of trust across societies with different levels of trust (including larger insurance groups such as countries). Even the evidence from the village insurance in poor countries shows that the size of the

[24]The prediction of the double-edged impact of trust on risk sharing is also consistent the high degree of risk sharing in poor, rural village economies in developing countries (see e.g. Townsend (1994), Ligon, Thomas, and Worrall (2002) or Meghir, Mobarak, Mommaerts and Morten (2022)) versus the relatively lower degree of risk-sharing in richer economies (see e.g. Attanasio and Davis (1996) or Altonji, Hayashi, and Kotlikoff (1997)). Of course, we do not contend that our mechanism limiting risk sharing is the only one consistent with this observation.

[25]There is also substantial evidence that public transfers crowd out private transfers; e.g. Attanasio and Rıos-Rull (2000), Juarez (2009), and Jensen (2004); as Farhi et al. (2009), Krueger and Perri (2011), Park (2014), in our model such public transfers would endogenously change the value of coalitional deviations and impact risk sharing ex-ante and ex-post.

insurance groups is larger than two (e.g., Ligon, Thomas, and Worrall (2002), Bold and Broer (2021)).[26] Second, the assumed stationarity of allocations is not consistent with the empirical evidence suggesting history-dependence of risk-sharing transfers. For example, using the ICRISAT data from rural India, the paper by Morten (2019) discussed above finds that transfers in the village risk sharing depend negatively on the history of past transfers. We now develop, in the next section, a more general model of trust in risk sharing to analyze these richer patterns of history-dependent transfers in large groups.

4 Risk Sharing with a Continuum of Agents

The analysis in Section 2 critically relied on the assumption of risk sharing within a pair of agents whose income is perfectly negatively correlated, since that assumption implied stationarity of the consumption paths in the strong social norm. As is well known, stationarity of optimal risk sharing is not a general property, failing as soon as the income shocks are imperfectly correlated. Moreover, if income shocks are independent, the restriction to risk sharing within pairs is restrictive, since better risk sharing can be achieved by larger groups.

In order to keep the model tractable, we consider risk sharing within continuum groups, so that there is no aggregate uncertainty within the group. We keep the timing of the simple model, so that if an agent decides to defect from the current arrangement, that agent consumes her current income and only then attempts to form a new arrangement with other agents. The fundamental trade-off is unchanged from the simple model: Will high income agents still be willing to sacrifice current consumption for the continued insurance possibilities offered in the current arrangement?

4.1 The Environment: Income, Preferences and Technology

As in Section 2, agents face idiosyncratic income risk in each period of a discrete-time infinite-horizon environment. As there, each agent in each period has low income $y = \ell > 0$ and high income $y = h > \ell$ with equal probability; we write $Y := \{\ell, h\}$ and $\overline{y} := \frac{1}{2}(\ell + h)$. Income realizations are independent across both agents and time. Preferences over *consumption allocations* $\{c(y^t)\}_t$ are represented by the lifetime utility function

$$(1-\beta)E\left\{\sum_{t=1}^{\infty}\beta^{t-1}u(c_t)\right\},$$

[26] The original ICRISAT data does not identify risk sharing groups within the village. The typical villages in the ICRISAT survey data consist of several hundred households, but Bold and Broer (2021) show that the size of the largest renegotiation-proof groups is much smaller.

where u is strictly increasing, strictly concave and satisfies the Inada conditions, and where we multiply period utility by $(1-\beta)$ to express period utility and lifetime utility in the same units. Lifetime utility of autarky is $V^A := \frac{1}{2}(u(\ell) + u(h))$ and of first-best insurance is $V^{FB} := u(\overline{y})$. As usual, we assume that in any positive measure (i.e., *large*) collection of agents, there is no aggregate income risk.

4.2 Coalition Formation and Deviation

At the beginning of the first period, $t = 1$, before each agent's income is realized, a positive measure of agents attempt to form a risk-sharing arrangement. Any arrangement needs to be robust to the possibility of deviations by an agent who could form another risk-sharing coalition after the deviation. Agents decide on deviations after learning their current income. The continual threat of deviations implies that any coalitional arrangement must itself be self-enforcing against the possibility that some members may deviate after that coalition has been formed.

If an agent does deviate from the current agreement, the deviating agent first consumes her current income, and then at the beginning of the next period, attempts to form a new risk-sharing arrangement with other agents. Since every history of income shocks is shared by a positive measure of agents, if an agent finds it profitable to deviate, a positive measure of agents will find it profitable to deviate. Since we have constant returns to scale, any positive measure set of deviating agents can implement any risk-sharing agreement that could have been originally agreed to. It is convenient to sometimes phrase the constraints as those pertaining to just the agents who deviate (while remembering that in principle, deviating agents can form an arrangement that also includes agents who were not in the original arrangement).

As we saw in Section 2, a risk-sharing agreement within a coalition is only reached if its members are confident that future cooperation is sustainable. Rather than explicitly modeling the reasons for this confidence (as we did in Section 2), we assume that any attempt to form a coalition succeeds with an exogenous probability $\pi \in [0, 1]$, the society's *trust*. If the attempt succeeds, then the coalition immediately implements a new risk-sharing agreement. When a coalition fails to form (which happens with probability $1 - \pi$), agents receive their autarky payoff V^A.[27]

[27]The precise specification after a coalition fails to form is not important; see footnote 5.

5 Social Norms

An *allocation* for a coalition is a consumption plan c specifying, for all periods t, an agent's consumption $c(y^t)$ in period t for every possible sequence $y^t \in Y^t$ of individual income shocks. We assume, without loss of generality, that individual consumption depends only on that agent's income history, independent of identity.

When attempting to reach a risk-sharing agreement, since member income levels are not yet known, a positive measure coalition faces the ex ante per capita resource constraint:

Definition 4 *An allocation for a coalition c is* feasible *if*

$$\sum\nolimits_{y^t} c(y^t) \Pr(y^t) \leq \overline{y}, \quad \forall t \geq 1. \tag{13}$$

The lifetime utility from an arbitrary consumption allocation c is given by

$$W^0(c) := (1-\beta) \sum_{\tau=1}^{\infty} \sum_{y^\tau} \beta^{\tau-1} \Pr(y^\tau) u(c(y^\tau)).$$

Initially, all agents are identical, and they will agree to follow any feasible consumption plan c that maximizes $W^0(c)$, as long as they are confident that the consumption plan will be followed in the future. A necessary condition for a consumption plan to be agreed to is that if all the agents do believe in it today, it should not be the case that after some history, an agent finds it optimal to deviate, and after the deviating period join a coalition that follows the *original* consumption plan.[28] Phrased differently, suppose the initial coalition believes that the allocation $\tilde{c}$ is credible, but that after some history y^t with current income y_t, agents receive strictly higher payoff from deviating, and if successfully forming a new coalition, implementing $\tilde{c}$ from the next period. Such a history means that the original coalition should not have believed in the credibility of the original allocation $\tilde{c}$, since it will *not* be implemented in its entirety. Accordingly, we are interested in allocations that are not vulnerable to such a criticism. In the simple model of Section 2, we called a stationary allocation that was not subject to such a criticism a social norm (Definition 1), since such an allocation, when agreed to by all groups, does not induce any deviations.

To define a social norm in the current setting, we need one additional piece of notation. For an arbitrary income history $y^t \in Y^t$, the *continuation* lifetime utility under the allocation

[28] As noted earlier, if an agent finds the deviation optimal, so will a positive measure set of agents, and so the (per capita) feasibility constraint faced by the set of deviating agents is identical to the original (per capita) feasibility constraint.

is

$$W(y^t, c) := (1-\beta)u(c(y^t)) + (1-\beta)\sum_{\tau=1}^{\infty}\sum_{y^\tau}\beta^\tau \Pr(y^\tau)u(c(y^t y^\tau)),$$

where $y^t y^\tau$ denotes the $t+\tau$-history that is the concatenation of t-period history y^t and the τ-period history y^τ.

Definition 5 *A feasible allocation c is a* social norm *if it satisfies* internal-incentive compatibility*, i.e., for all $t \geq 1$ and for all $y^t \in Y^t$,*

$$W(y^t, c) \geq (1-\beta)u(y_t) + \beta[\pi W^0(c) + (1-\pi)V^A]. \tag{14}$$

Let $\mathcal{C}$ denote the set of social norms.

This is a weak notion of credibility to deter deviations. For example, while the autarky allocation is trivially a social norm, that allocation has lower utility than allocations with some insurance. The stability notion is "internal" in the sense that when evaluating the credibility of an allocation, agents only consider the possibility that if accepted, that allocation will also determine the outside option for any deviating set of agents. Agents do not consider the possibility that the payoffs after a deviation may be determined by a different (possibly more attractive) allocation. The stronger requirement of a strong social norm (first introduced in Definition 3 in the simple model and defined for the current setting in Definition 7 below) can lead to nonexistence.

Internal incentive compatibility (14) is the key friction that prevents full consumption insurance within a coalition.

Definition 6 *For given trust π, an allocation c is a* constrained-efficient social norm *if it solves the program*

$$\max_{c\in\mathcal{C}} \; W^0(c).$$

Denote by $\mathbb{W} = \max_{c\in\mathcal{C}} \; W^0(c)$ the resulting optimal lifetime utility and by $\mathbb{F} = \pi\mathbb{W} + (1-\pi)V^A$ the associated ex ante (and so deviation continuation) utility.

The value $\mathbb{W}$ is the maximum per capita value the coalition can achieve, given the credible threat that agents will deviate (and implement the same agreement) if the initial arrangement is not sufficiently generous to that group. If an agent with current income y does deviate, she consumes y in the current period, and then in the next period with probability π, joins a group that is able to coordinate on future risk sharing, with payoff $\mathbb{W}$, and with probability $1-\pi$, does not join a group (and so has no future risk sharing), yielding $(1-\beta)u(y) + \beta\mathbb{F}$ as the expected payoff from deviating.

Since the autarkic allocation is trivially a social norm, the set of social norms is nonempty, and so the supremum of $W^0(c)$ exists and is bounded above by $u(\overline{y})$, the utility of first-best insurance. Moreover, as $\mathcal{C}$ is closed (in the product topology), the supremum is always attained and so constrained-efficient social norms exist. The main focus of our analysis is concerned with the characterization of the constrained-efficient social norm, and how its qualitative properties vary with the degree of trust in society, π.

5.1 Risk Sharing in a Social Norm

As in the simple model, we first consider the possibility of first-best insurance. If first-best insurance is internally incentive compatible (i.e. if it is a social norm), then it is evidently the constrained-efficient social norm. Since first-best insurance is achieved by a stationary allocation, identical calculations that yielded Proposition 1 in the simple model also show that, in the current setting, first-best insurance is the constrained-efficient social norm only when trust is not too large, with the threshold π^{FB} continuing to be given by (8). Furthermore, the cutoff value for the discount factor β^{FB} below which first-best insurance is not a social norm for any level of trust remains as defined in (10) of the simple model. That is, Proposition 1 from the simple model applies completely unchanged to the model with a continuum of agents in this section.

Of more interest is the possibility of partial insurance in a social norm when first-best insurance is not a social norm, which is illustrated by the next example.

Example 1 Consider the allocation in which agents with currently high income transfer ε to all agents that had high income yesterday but have low income today:

$$c_\varepsilon(y^t) := \begin{cases} h-\varepsilon, & y_t = h, \\ \ell+2\varepsilon, & y_{t-1} = h, y_t = \ell, \\ \ell, & \text{otherwise.} \end{cases} \tag{15}$$

Also assume that the discount factor β satisfies

$$\beta > \frac{u'(h)}{u'(\ell)}. \tag{16}$$

Note that since there are only half as many agents with $y_{t-1}y_t = h\ell$ than with $y_t = h$, this allocation satisfies feasibility with equality in every period *except* the initial period, when there are no $y_{t-1}y_t = h\ell$ agents and thus ε resources from every high-income agent are destroyed.

We claim that for $\varepsilon > 0$ small, and as long as condition (16) is satisfied, $c_\varepsilon \in \mathcal{C}$, and thus this partial insurance allocation is internally incentive compatible. A sufficient condition for $c_\varepsilon \in \mathcal{C}$ for $\pi = 1$ is that high-income agents have no incentive to deviate from allocation c_ε (see the internal-incentive compatibility (14) constraint)

$$W(h, c_\varepsilon) \geq (1-\beta)u(h) + \beta W^0(c_\varepsilon). \tag{17}$$

Note that if this constraint is satisfied for $\pi = 1$, it is (strictly) satisfied for all other $\pi < 1$.

By deviating, an h-agent gives up one period of 2ε insurance in the event that she has ℓ income in the next period (which occurs with probability 1/2). So a sufficient condition for (17) to hold for sufficiently small ε is that the marginal benefit of deviating be smaller than the marginal expected delayed cost at $\varepsilon = 0$,

$$(1-\beta)u'(h)\varepsilon < (1-\beta)\frac{\beta}{2}u'(\ell)2\varepsilon,$$

which reduces to the assumed bound on β in equation (16). Note that, since $W^0(c_\varepsilon) > V^A$ for ε small, this allocation indeed provides partial insurance. Condition (16) also turns out to be *necessary* for insurance as well (see Proposition 4 in the next section).[29]

★

Two features of Example 1 deserve further discussion. The first is that the initial period resource destruction plays a critical role in allocation's satisfaction of internal-incentive compatibility. In particular, if the ε resources sacrificed by the initial h-income agents are given to the initial ℓ-income agents (providing additional ex ante insurance), the resulting allocation need not *not* satisfy internal-incentive compatibility.[30]

The second is the time-varying nature of the insurance provided. When first-best insurance is not a social norm, h-income agents optimally secede under the first-best allocation. To reduce this secession incentive, a natural allocation is the stationary allocation of the form studied in the simple example economy of Section 2,

$$c_x(y^t) := \begin{cases} h - x, & y_t = h, \\ \ell + x, & y_t = \ell. \end{cases} \tag{18}$$

[29]This lower bound on the discount factor for partial insurance coincides with that in limited commitment models with exogenous outside option and a continuum of agents, see, e.g., Krueger and Perri (2011).

[30]Using the same calculations for this non-wasteful allocation shows that the conditions for this allocation to satisfy internal incentive are more restrictive. For example, for $\pi = 1$, the bound on the discount factor is $\beta > 2u'(h)/u'(\ell)$, rather than condition (16). The proof of Lemma C.1 uses this property of the modified allocation.

For $x = 0$, c_x is the autarkic allocation, while for $x = h - \overline{y}$, c_x is the first-best allocation. While such an allocation can be a social norm, it is less efficient in its provision of incentives. For example, for $\pi = 0$, c_x satisfies (14) for sufficiently small $x > 0$ only if

$$-(1-\beta)u'(h) + \frac{\beta}{2}[u'(l) - u'(h)] \geq 0 \implies \beta \geq \frac{2u'(h)}{u'(\ell) + u'(h)} > \frac{u'(h)}{u'(\ell)}, \tag{19}$$

which is the bound on β given in (11) of the simple model, and is more restrictive than condition (16) for c_ε to be a social norm.

6 The Strong Social Norm

Characterizing the constrained-efficient social norm allocation is complicated by the nature of the internal-incentive compatibility constraint (14), which implies that the set of social norm allocations is not convex. This non-convexity emerges from the endogeneity of the deviating coalition's payoff in the incentive constraints. As in Section 2, we first characterize, though a fixed-point argument, the *strong* social norm for the subset of trust values π for which it exists (and thus is the constrained-efficient social norm), and then characterize constrained-efficient social norms for the remaining values of π for which the strong social norm does not exist.

Recall that internal-incentive compatibility (14) requires

$$W(y^t, c) \geq (1-\beta)u(y_t) + \beta[\pi W^0(c) + (1-\pi)V^A] \qquad \forall y^t \in \cup_\tau Y^\tau.$$

We begin by considering feasible allocations that satisfy an exogenous version of this constraint, which we call *F-incentive compatibility*,

$$W(y^t, c) \geq (1-\beta)u(y_t) + \beta F \qquad \forall y^t \in \cup_\tau Y^\tau. \tag{20}$$

For an exogenous ex-ante value $F \in \mathbb{R}$ of defection, denote by $\mathcal{C}(F)$ the set of feasible allocations satisfying (20). If F is too large, then $\mathcal{C}(F)$ will be empty. But if c is a social norm for a given level of trust π, then $c \in \mathcal{C}(\pi W^0(c) + (1-\pi)V^A)$, and so the constraint set $\mathcal{C}(F)$ is non-empty for all outside options $F \leq \pi W^0(c) + (1-\pi)V^A$.

In what follows we use $\mathbb{c}$ to denote an allocation c that is the solution to a maximization problem. The notion of a strong social norm from Section 2 extends to the current setting.

Definition 7 *For a given π, an allocation $\mathbb{c}$ is a* strong social norm *if it maximizes $W^0(c)$ over $c \in \mathcal{C}(F)$ when $F = \pi W^0(\mathbb{c}) + (1-\pi)V^A$.*

If a strong social norm exists for a given π, it follows immediately from the definition that it is a constrained-efficient social norm for that π.

When $\mathcal{C}(F) \neq \varnothing$, the expected lifetime utility of a successfully formed coalition that maximizes against the outside option F is

$$\mathbb{V}(F) := \max_{c \in \mathcal{C}(F)} W^0(c). \tag{21}$$

Since $\mathcal{C}(F)$ is a convex set and $W^0(c)$ is a strictly concave function, the above maximization has a unique solution when $\mathcal{C}(F)$ is nonempty. Note that trust π does not appear in the maximization in (21). Instead, the exogenous outside option F determines the optimal allocation and value. But the two are intimately connected. Since agents only successfully coordinate after a deviation with probability π, if F is the implied continuation value of the outside option for a deviating coalition, then, for all $y \in Y$, the value of the outside option is determined by the mapping

$$\mathcal{T}(F; \pi) := \pi \mathbb{V}(F) + (1 - \pi) V^A. \tag{22}$$

The following proposition (proved in Appendix B.1) uses the mapping $\mathcal{T}$ to characterize strong social norms and thus constrained efficient social norms.

Proposition 2 *For a given $\pi \in [0, 1]$, suppose F is a fixed point of $\mathcal{T}(\,\cdot\,; \pi)$. Then there exists a unique allocation $\mathbb{c} \in \mathcal{C}(F)$ satisfying $W^0(\mathbb{c}) = \mathbb{V}(F)$ and $F = \pi W^0(\mathbb{c}) + (1 - \pi) V^A$. The allocation $\mathbb{c}$ is the unique strong social norm, and thus the constrained-efficient social norm for that π, and F is the ex ante value of that social norm. Moreover, F is the only fixed point of $\mathcal{T}(\,\cdot\,; \pi)$.*

Thus, the strong social norm exists for those levels of trust π consistent with outside options that are fixed points of $\mathcal{T}(\,\cdot\,; \pi)$. The fixed point may fail to exist because the constraint set is not a "nice" function of the outside option F, or the constraint set is empty for F in a relevant region. While Proposition 5 below assures us that the former is not an issue (the constraint set is a "nice" function of F), the constraint set *is* empty for large F (which will correspond to large π) and so a fixed point of $\mathcal{T}(\,\cdot\,; \pi)$ does not exist in that case. Define

$$\overline{F} := \sup\{F \mid \mathcal{C}(F) \neq \varnothing\}.$$

as the sup of the outside option for which there are F-incentive-compatible allocations.

6.1 Characterization and Comparative Statics

In Section 2, we established some natural comparative statics of the simple model with respect to the level of trust. In particular,

1. for very low levels of trust, first-best insurance is a strong social norm (when agents are sufficiently patient) and while the chance that it is implemented is very low, its ex ante value is increasing in trust,

2. for increasing levels of trust, the amount of risk sharing falls, but the ex ante value of the risk sharing increases, and

3. for even higher levels of trust, the strong social norm does not exist, the level of insurance in the constrained efficient social norm is decreasing and the ex ante value is constant in trust.

The comparative statics of the current model are not as straightforward as the simple model in large part because the strong social norm (when it is not first-best insurance) is not stationary. However, we can establish that the optimal allocation converges to a stationary *ladder*, defined next, and conduct comparative statics with respect to this limit ladder.

Intuitively, consumption in a ladder in any period is determined by the number of ℓ realizations after the last h realization with consumption falling after each additional ℓ. The critical property is that the history of income realizations before the last h realization is irrelevant. The allocation c_x define in (18) is a stationary ladder with $L = 2$ and c_ε defined in (15) is a ladder with $L = 3$.

Definition 8 *An allocation c is a* ladder sequence *if there is a sequence of finite sequences $\left(\left(c_{t+k}(h\ell^k)\right)_{k=0}^{L}\right)_{t=1}^{\infty}$ such that for all $t \geq 1$ and all $k \geq 0$, $c(y^{t-1}h\ell^k) = c_{t+k}(h\ell^k)$. The finite sequence $\left(c_{t+k}(h\ell^k)\right)_{k=0}^{L}$ is a period-t* ladder*. An allocation c is a* stationary ladder *if $c_{t+k}(h\ell^k)$ is independent of t for all $k \geq 0$; denote this consumption by $c_*(h\ell^k)$.*

When first-best insurance is not a strong social norm, the strong social norm (if it exists) is a ladder sequence characterized by the floor on the consumption of the longtime poor agents c_ℓ and two sequences: the sequence $(c_t(h))$ of consumptions by the rich agents in period t and a growth rate (δ_{t+1}^{-1}) on marginal utilities (which implies consumption decay after each ℓ realization). The proof of the following proposition (most of which is standard, involving variational arguments) is in Appendix B.2.

Proposition 3 *Suppose $\mathbb{c}$ is a strong social norm that is not first-best insurance. There is a sequence $(c_t(h), \delta_{t+1})_{t\geq 0}$ with $\delta_{t+1} < 1$ and $c_\ell > \ell$ such that*

1. $\mathbb{c}(y^{t-1}h) = c_t(h)$ *for all* y^{t-1},

2. *if* $\mathbb{c}(y^t\ell) > c_\ell$, *then for all* $t \geq 1$,

$$u'(\mathbb{c}(y^t)) = \delta_{t+1}u'(\mathbb{c}(y^t\ell)) \quad \textit{for all } y^t, \tag{23}$$

and

3. *there exists* $L > 1$ *such that* $\mathbb{c}(y^t\ell^k) = c_\ell$ *for all* $k \geq L$.

Moreover, $(c_t(h), \delta_{t+1})_{t\geq 0}$ *converges to a limit* $(c_*(h), \delta_*)_{t\geq 0}$ *with* $\delta_* < 1$, *and so* $(c_t(h\ell^k)_t$ *converges for all* k. *Convergence does not occur in finite time. Risk sharing is declining over time, in the sense that for all* $t \geq 1$ *and* $k \geq 2$, $c_t(h) < c_{t+k}(h)$ *and* $\delta_{t+1} > \delta_{t+1+k}$.

The allocation determined by c_ℓ and $(c_*(h), \delta_*)$ is the *limit stationary ladder*.

We can now state the main result of the paper, which captures the comparative statics and summarizes the analysis to follow in the rest of the paper:[31]

Proposition 4 *Constrained-efficient social norms exist for all* $\pi \in [0, 1]$ *and are characterized as follows:*

1. *Suppose* $\beta \leq \underline{\beta} := u'(h)/u'(\ell)$. *There is no risk sharing in any social norm, i.e., autarky is the unique social norm, and therefore trivially the constrained efficient social norm.*

2. *Suppose* $\beta > \underline{\beta}$. *Risk sharing does occur in constrained-efficient social norms. There exist threshold values* $\pi^{FB}(\beta) < 1$ *and* $\overline{\pi}(\beta) \in (0, 1]$ *with* $\pi^{FB}(\beta) < \overline{\pi}(\beta)$ *such that the following hold:*

   (a) *For* $\pi \leq \pi^{FB}(\beta)$, *first-best insurance is the constrained-efficient social norm, with ex-ante value* $F^{FB} = \pi V^{FB} + (1-\pi)V^A$ *which is strictly increasing in* π.

   (b) *For* $\pi \in (\pi^{FB}(\beta), \overline{\pi}(\beta)]$, *the strong social norm* $\mathbb{c}$ *exists and is the constrained-efficient social norm. Its ex ante value* $\pi W^0(\mathbb{c}) + (1-\pi)V^A$ *is strictly increasing in* π, *equaling* $\overline{F} > V^A$ *at* $\overline{\pi}$. *Risk sharing is strictly decreasing in* π *in two senses:*

      i. *The value of the strong social norm* $W^0(\mathbb{c})$ *is strictly decreasing in* π.

      ii. *Risk-sharing in the limit stationary ladder is strictly decreasing in* π, *in the sense that consumption of high-income agents* $c_*(h)$ *is strictly increasing in* π *and consumption falls more rapidly along a spell of low income realizations the larger is* π, *i.e., the decay rate* δ_* *is strictly decreasing in* π.

[31] To simplify notation, we occasionally leave the dependence on β of $\overline{\pi}$, $\overline{F}$, and similar functions implicit.

(c) *For $\pi \in (\overline{\pi}(\beta), 1]$, the strong social norm does not exist and there are multiple constrained-efficient social norms, all with the same ex ante value of $\overline{F}$. The value of every constrained-efficient social norm is given by $[\overline{F} - V^A]/\pi + V^A$, which is strictly decreasing in π.*

3. $\lim_{\beta \searrow \underline{\beta}} \overline{\pi}(\beta) = 0$.[32]

Proof. Existence of constrained-efficient social norms is immediate, as discussed after Definition 6.

1. This is an implication of the machinery we develop to characterize $\overline{F}$, and is Corollary 1 in Section 7.

2. (a) This is an immediate implication of Proposition 2 above and Proposition 5, which is established in the next subsection, and the strict concavity of the problem (21).

   (b) The existence and monotonicity of the strong social norm with respect to π is established in the next two subsections. The comparative statics result concerning the stationary ladder is proved in Appendix B.4.

   (c) Propositions 7 and 8 in Section 8 exhibit two allocations that burn utility in distinct ways with ex ante value $\overline{F}$ and prove these are constrained efficient.

3. This is an implication of the machinery we develop to characterize $\overline{F}$, and is Corollary 2 in Section 7.

□

Several comments on the characterization of the strong social norm are in order. In general, for levels of trust $\pi > \pi^{FB}$ and associated outside options $F > F^{FB}$, the optimal allocation provides maximal risk sharing consistent with the incentive constraints (20). The proof of Proposition 3 reveals that this constraint always holds with equality for h-income agents and sometimes for ℓ-income agents. In order to deter an h-income agent from defecting, the optimal allocation does two things: First, it reduces the transfer to low-income agents below the first-best level. Second, the risk sharing offered is "front-loaded" so that ℓ-income agents who had more recently received a h realization receive more insurance than those who last received a h realization further in the past.[33]

[32] We conjecture that $\overline{\pi} < 1$ for all $\beta \in (0, 1)$ (and not just for β near $\underline{\beta}$. While we have not been able to prove this, all our computed examples have this property; see Appendix S.3.

[33] The property that consumption drifts down with each subsequent ℓ-realization is a common feature of the standard limited commitment model with an exogenous outside option. The front-loading property in this paper does *not* refer to the front-loading of distortions (leading to the back-loaded consumption), which is the typical result of the one-sided limited commitment problem and the self-enforcing equilibrium with capital accumulation (see Albanesi and Armenter (2012) and the references cited therein).

This front-loading, reflected in the declining consumptions on the ladder, implies that consumption eventually, after a sufficiently long string of ℓ-realizations, is determined by F-incentive compatibility for the ℓ-realization. The resulting lower bound on consumption, $c_\ell > \ell$ reflects the following trade-off: Defecting from $\mathbb{c}$ does mean that the agents give up some risk-sharing today, but the benefit is that in a new coalition tomorrow, any agent who receives another ℓ realization receives more generous risk sharing tomorrow (since F-incentive compatibility holds strictly in the first period after ℓ by Lemma B.8, $c_\ell < \mathbb{c}(\ell)$).

In Section S.3 of the appendix we present numerically computed examples of constrained-efficient allocations. These examples demonstrate two points. First, they show that even for $\beta \gg \underline{\beta}$, the threshold for utility burning satisfies $\overline{\pi}(\beta) < 1$. This is guaranteed for low β by part 3 of Proposition 4 but Figure S.3 demonstrates it is a pervasive phenomenon for larger β as well. Second, as Figure S.4 shows, strong social norms converge to the stationary ladder rapidly over time, and thus the stationary ladder and its comparative statics properties are informative about how allocations look like in our model.

6.2 The Fixed Point Problem Characterizing Strong Social Norms

The following result (proved in Appendix B.3) characterizes the range of trust levels π for which the fixed point of the mapping $\mathcal{T}(\,\cdot\,;\pi)$ exists, and with it the strong social norm.

Proposition 5 *Suppose $\beta > u'(h)/u'(\ell)$.*

1. *$\overline{F} > V^A$.*
2. *$\mathcal{C}(\overline{F}) \neq \varnothing$.*
3. *The value of the problem (21), $\mathbb{V}(F)$, is continuous in F for all $F \leq \overline{F}$.*
4. *If $\beta > \beta^{FB}$, then $\overline{F} > F^{FB}$.*[34]
5. *Define*
$$\overline{\pi} := \min\left\{\frac{\overline{F} - V^A}{\mathbb{V}(\overline{F}) - V^A}, 1\right\}. \tag{24}$$
For all $\pi \in (0, \overline{\pi}]$, $\mathcal{T}(\,\cdot\,,\pi)$ has a unique fixed point $F(\pi)$. The function $F(\,\cdot\,)$ is strictly increasing in π. If $\overline{\pi} < 1$, $\overline{F} = F(\overline{\pi})$ and if $F(\overline{\pi}) < \overline{F}$, $\overline{\pi} = 1$.
6. *If $\overline{\pi} < 1$, then for all $\pi \in (\overline{\pi}, 1]$, $\mathcal{T}(\,\cdot\,,\pi)$ does not have a fixed point.*

[34]The largest outside option consistent with first-best insurance F^{FB} was defined in (9) in Proposition 1 which applies to the full model unchanged. If $\beta < \beta^{FB}$, then $F^{FB} \leq V^A$, see Proposition 1.

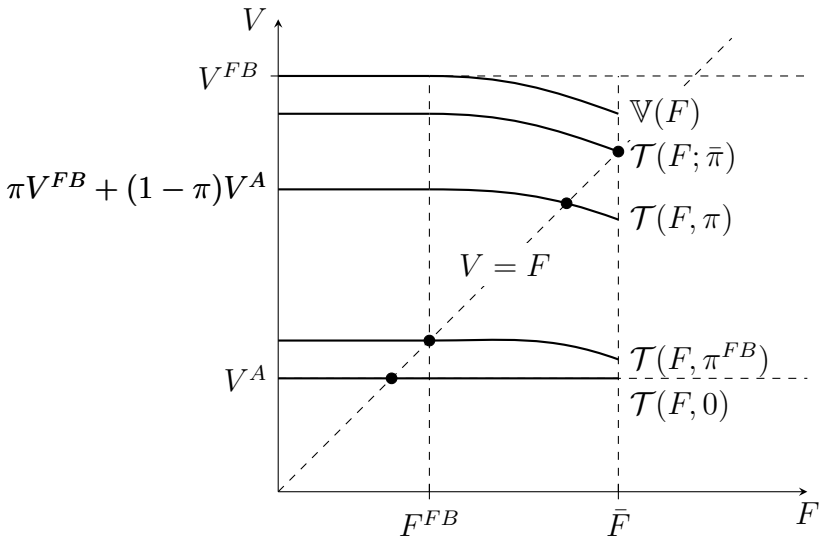

Figure 2: Determination of the fixed point of $\mathcal{T}(F;\pi) = \pi\mathbb{V}(F) + (1-\pi)V^A$ for different values of π. Drawn for $\beta > \beta^{FB}$ and assuming $\mathbb{V}(\overline{F}) > \overline{F}$; if $\beta < \beta^{FB}$, then $F^{FB} < V^A$.

Note that autarky is not a fixed point when $\pi > \overline{\pi}$ (Proposition 5, part 6). Although autarky is a social norm, it is dominated by a better allocation, but this better allocation does not satisfy internal incentive compatibility (14) when a seceding coalition is free to reoptimize.

Figure 2 illustrates this proposition by plotting $\mathbb{V}(F)$ and $\mathcal{T}(F;\pi)$ against the value of the outside option F for various degrees of trust π. First consider the function $\mathbb{V}(F)$, and assume $\beta \geq \beta^{FB}$. Then for small outside options $F \in [V^A, F^{FB}]$ first-best insurance is incentive-compatible and $\mathbb{V}(F) = V^{FB}$ for these outside options. As the outside option rises above F^{FB} the F-incentive compatibility constraint (20) binds at least for agents with currently high income, implying the initial coalition cannot sustain first-best insurance ($\mathbb{V}(F) < V^{FB}$) and that the utility $\mathbb{V}(F)$ it delivers is strictly decreasing in F.

Now consider the mapping $\mathcal{T}(F;\pi)$, which, for a fixed level of trust π, is the convex combination of the function $\mathbb{V}(F)$ (weight π) and the constant V^A. The figure shows how the fixed point $\mathcal{T}(F;\pi)$ varies with π. At one extreme, $\pi = 0$ and we have $\mathcal{T}(F;0) = V^A$ and thus trivially $F = V^A$ is the unique fixed point. In this case $\mathbb{V}(V^A) = V^{FB}$ and the allocation for the initial coalition would feature first-best insurance, but since $\pi = 0$, it never successfully forms. First-best insurance remains the outcome for the successful coalition as long $\pi \leq \pi^{FB} < 1$ and the fixed point $F(\pi) = \pi V^{FB} + (1-\pi)V^A$ is strictly increasing in trust π until it reaches the largest deviation lifetime utility $F^{FB} = \pi^{FB}V^{FB} + (1-\pi^{FB})V^A$ for which first-best insurance can be sustained inside the initial coalition.

For $\pi \in (\pi^{FB}, \overline{\pi}]$, the value of the outside option F continues to be determined as the fixed point of $\mathcal{T}(\,\cdot\,;\pi)$. The fixed point is larger than F^{FB}, and so the incentive constraint (20) binds at least for agents with currently high income, implying that the initial condition cannot sustain first-best insurance ($\mathbb{V}(F) < V^{FB}$), and that the lifetime utility $\mathbb{V}(F)$ the initial coalition delivers is strictly decreasing in F as risk-sharing inside the coalition worsens. The ex-ante (prior to coalition formation) and outside option utility $F(\pi)$, given by the fixed point $F(\pi) = \mathcal{T}(F(\pi);\pi)$ continues to increase since the initial coalition is more likely to form, as Figure 2 shows.

Finally, consider high trust $\pi > \overline{\pi}$ and suppose $\overline{\pi} < 1$. For $\pi > \overline{\pi}$, since $\mathbb{V}(\overline{F}) > V^A$

$$\pi\mathbb{V}(\overline{F}) + (1-\pi)V^A > \overline{F} = \overline{\pi}\mathbb{V}(\overline{F}) + (1-\overline{\pi})V^A.$$

Since $\mathcal{C}(F)$ is empty for $F > \overline{F}$ and thus $\mathbb{V}(F)$ is not defined for these F, $\mathcal{T}(\,\cdot\,;\pi)$ does not have a fixed point ($\mathcal{T}(F;\pi)$ does not intersect the 45-degree line in Figure 2 for $\pi > \overline{\pi}$), and so there is no strong social norm. But there is still a constrained-efficient social norm $\mathbb{c}$ with value $W^0(\mathbb{c})$ (see Proposition 4). This norm must satisfy

$$\mathbb{c} \in \mathcal{C}(\pi W^0(\mathbb{c}) + (1-\pi)V^A),$$

and so

$$\pi W^0(\mathbb{c}) + (1-\pi)V^A \leq \overline{F} = \overline{\pi}\mathbb{V}(\overline{F}) + (1-\overline{\pi})V^A. \tag{25}$$

Since $\pi > \overline{\pi}$, we have $W^0(\mathbb{c}) < \mathbb{V}(\overline{F})$, that is, the initial coalition chooses an allocation that is worse than what it could achieve if maximizing against the exogenous outside option $\overline{F}$. This leads us to the following definition:

Definition 9 *A social norm* $\mathbb{c}$ burns utility *if*

$$W^0(\mathbb{c}) < \mathbb{V}(\overline{F}).$$

A constrained-efficient social norm maximizes ex ante utility (the left side of (25)) over the set of social norms. We show in Section 8 that, for each $\pi \in (\overline{\pi}, 1]$, the constrained-efficient social norm satisfies (25) with equality, i.e., it delivers the ex ante value $\overline{F}$, but at the expense of worsening risk-sharing inside successfully formed coalitions as π increases. It therefore increasingly burns utility, in the sense of Definition 9, as trust π rises.

7 Characterizing $\overline{\pi}$

We now characterize $\overline{\pi}$, or equivalently, $\overline{F}$. It turns that $\overline{F}$ has a simple characterization as the maximum value of the outside option consistent with h-incentive compatibility from a specific stationary ladder. Consider the stationary ladder c_* that maximizes lifetime utility from being on this ladder:

$$\begin{aligned} W(h, c_*) &= (1-\beta)u(c_*(h)) + \tfrac{\beta}{2}\left\{(1-\beta)u(c_*(h\ell)) + W(h, c_*)\right\} \\ &\qquad + \left(\tfrac{\beta}{2}\right)^2\left\{(1-\beta)u(c_*(h\ell^2)) + W(h, c_*)\right\} + \cdots \\ &= (1-\beta)\sum_{k=0}^{\infty}(\tfrac{\beta}{2})^k u(c_*(h\ell^k)) + \tfrac{\beta}{2-\beta}W(h, c_*) \\ &= \left(1-\tfrac{\beta}{2}\right)\sum_{k=0}^{\infty}(\tfrac{\beta}{2})^k u(c_*(h\ell^k)), \end{aligned} \tag{26}$$

subject to F-incentive compatibility for ℓ-income realizations

$$c_*(h\ell^k) \geq c_\ell(F) \text{ for all } k \geq 1 \tag{27}$$

and feasibility

$$\sum_{k=0}^{\infty}(\tfrac{1}{2})^{k+1}c_*(h\ell^k) \leq \overline{y}. \tag{28}$$

Denote the constraint set defined by equations (27) and (28) by $\mathcal{C}_*(F)$ and the maximum value of the program by

$$\mathbb{V}^*(h; F) := \max_{c_* \in \mathcal{C}_*(F)} W(h, c_*). \tag{29}$$

In this problem, h-incentive compatibility does not appear as a constraint because we are maximizing the payoff of the current h agents. Note also that feasibility is being imposed on the ladder, and so there is only one constraint. In contrast, feasibility was not imposed on any ladder in $\mathcal{C}(F)$, being imposed instead in each period.[35]

The next proposition (proved in Appendix C) makes precise the sense in which $\overline{F}$ is the maximum value of the outside option consistent with h-incentive compatibility, and clarifies the role the program in (29) plays in determining this value.

[35]This also means that in general, the stationary ladder solving problem (29) is not the limit stationary ladder characterized in Proposition 3 (which satisfies feasibility and both incentive constraints with equality). However, for $F = \overline{F}$ the two coincide because then F is so large that $\mathcal{C}_*(F)$ becomes a singleton.

Proposition 6 *The set of resource and incentive compatible allocations $\mathcal{C}(F)$ is nonempty if and only if*

$$\mathbb{V}^*(h;F) \geq (1-\beta)u(h) + \beta F =: W^F(h).$$

Moreover,

$$F = \overline{F} \iff \mathbb{V}^*(h;F) = W^F(h).$$

Note that this proposition also shows how to construct $\overline{F}$ as long as $\overline{F} \in (V^A, V^{FB})$ numerically. It is determined by the unique stationary consumption ladder that satisfies h- and ℓ-incentive compatibility (at $F = \overline{F}$) as well as feasibility with equality. Furthermore, from the first-order conditions of the program (29) it follows immediately that the decay rate of marginal utility in this stationary ladder is given by $\bar{\delta} = \beta$. Appendix S.3.3.2 shows how to exploit this observation to compute $\overline{F}$ numerically.

The following corollary gives the condition under which the strong social norm cannot feature any risk sharing.

Corollary 1 *If $\beta u'(\ell) \leq u'(h)$, then*

$$\overline{F} = V^A.$$

Proof. Suppose $\overline{F} > V^A$. By Proposition 6, for all $F \in (V^A, \overline{F}]$,

$$\mathbb{V}^*(h,F) \geq (1-\beta)u(h) + \beta F. \tag{30}$$

But $\beta u'(\ell) \leq u'(h)$ implies that autarky provides an upper bound for (29) and so, using (26)

$$\begin{aligned}\mathbb{V}^*(h,F) &\leq (1-\frac{\beta}{2})u(h) + \frac{\beta}{2}u(\ell)\\ &= (1-\beta)u(h) + \beta V^A\\ &< (1-\beta)u(h) + \beta F,\end{aligned}$$

contradicting (30). Hence, we must have $\overline{F} = V^A$. □

This corollary shows that under the specified condition the highest outside option that can be attained is autarky, and thus the only social norm is one without any insurance. The next corollary (proved in Appendix C) confirms that we have continuity from the right.

Corollary 2

$$\lim_{\beta \searrow u'(h)/u'(\ell)} \overline{\pi}(\beta) = 0.$$

8 The Case of Utility Burning, $\pi > \overline{\pi}$

For high values of trust ($\pi > \overline{\pi}$), constrained efficiency requires utility burning. While constrained-efficient social norms must now impose additional inefficiencies, the precise nature of these inefficiencies is not determined. Rather, these inefficiencies are chosen to exactly offset the increase in trust so that the ex ante value remains at $\overline{F}$. We present two propositions (proved in Appendix D), illustrating possible choices of inefficiencies due to either postponing risk sharing or burning resources. Denote by $\bar{\mathbb{c}}$ the strong social norm for $\pi = \overline{\pi}$. The first proposition describes a constrained-efficient social norm that postpones risk sharing.

Proposition 7 *Suppose $\pi > \overline{\pi}$. Denote by $c^{(T)}$ the allocation specifying T periods of autarkic consumption followed by $\bar{\mathbb{c}}$ in a history independent manner. There exists $T(\pi)$ and $\alpha(\pi) \in [0,1]$ for which the convex combination*

$$c^{(\alpha(\pi))} := \alpha(\pi)c^{(T(\pi)-1)} + (1-\alpha(\pi))c^{(T(\pi))}.$$

is a constrained-efficient social norm, and the value of this allocation is $\overline{F}$.

Thus, the allocation $c^{(\alpha(\pi))}$ postpones risk sharing for $T(\pi) - 1$ periods and then provides intermediate risk sharing in future periods.

The next proposition describes a constrained-efficient social norm that burns resources instead of postponing insurance.

Proposition 8 *Define the consumption allocation $c^{[\alpha]}$ as follows:*

$$c^{[\alpha]}(y^t) = \begin{cases} \bar{\mathbb{c}}(y^t) & \text{if } y^t \neq \ell^t, \\ \alpha\bar{\mathbb{c}}(y^t) + (1-\alpha)c_\ell(\overline{F}), & \text{if } y^t = \ell^t. \end{cases}$$

There exists $\alpha(\pi)$ for which $c^{[\alpha(\pi)]}$ is a constrained efficient social norm whose value is $\overline{F}$.

Note that the consumption allocation $c^{[\alpha]}$ only differs from $\bar{\mathbb{c}}$ at histories ℓ^t. Moreover, since $\bar{\mathbb{c}}(\ell^t) = c_\ell(\overline{F})$ in finite time (Proposition 3), $c^{[\alpha]}(y^t) = \bar{\mathbb{c}}(y^t)$ for $t \geq L$.

Taken together, Propositions 7 and 8 display two options of how sufficient utility can be burned in a constrained-efficient allocation to insure that internal-incentive compatibility is satisfied and risk sharing is optimal, given these constraints.

9 Conclusion

We have proposed a model in which trust facilitates the formation of efficiency-enhancing risk-sharing coalitions as well as coalitional deviations from these original arrangements. The symmetric treatment of initial and deviating coalitions, both with respect to the allocation chosen and the composition of the group, ties together tightly the ex ante payoff and the outside option. This tight link implies that as our notion of trust π increases, these two payoffs rise together. The double-edged aspect of π, making it easier to form both initial and deviating coalitions, leads to the differential impact of a higher π on ex ante utility (which is weakly increasing), and ex post utility conditional on formation as well as the steady state distribution of continuation payoffs (which are weakly decreasing in π). Moreover, at high degrees of trust, constrained-efficient allocations exhibit utility burning as necessary feature.

The comparative statics with respect to π exhibit three regions. With a low probability of forming a coalition, ex ante welfare is linearly increasing in π and conditional on coalition formation, members receive complete insurance. At an intermediate range ex ante welfare is increasing in π but at a decreasing rate and conditional on coalition formation, insurance is incomplete and declining in π. Allocations feature wasteful inequality but are intertemporally efficient. With high levels of trust, ex ante welfare is flat in π, and allocations feature significant inefficiencies, manifested in utility or resource burning within a coalition to prevent defections. In a nutshell, an increase in π enables groups to more readily *trust each other* by agreeing on Pareto improving exchanges but at the same time making this *trust shallower*.

References

Ábrahám, Árpád and Sarolta Laczó (2017), "Efficient risk sharing with limited commitment and storage." *Review of Economic Studies*, 85, 1389–1424.

Aiyagari, S. Rao and Stephen D. Williamson (2000), "Money and dynamic credit arrangements with private information." *Journal of Economic Theory*, 91, 248–279.

Albanesi, Stefania and Roc Armenter (2012), "Intertemporal distortions in the second best." *The Review of Economic Studies*, 79, 1271–1307.

Alesina, Alberto and Paola Giuliano (2014), "Chapter 4 - Family Ties." In *Handbook of Economic Growth* (Philippe Aghion and Steven N Durlauf, eds.), volume 2, 177–215, Elsevier.

Algan, Yann and Pierre Cahuc (2014), "Chapter 2 - Trust, Growth, and Well-Being: New Evidence and Policy Implications." In *Handbook of Economic Growth* (Philippe Aghion and Steven N B T Handbook of Economic Growth Durlauf, eds.), volume 2, 49–120, Elsevier.

Allen, Franklin (1985), "Repeated principal-agent relationships with lending and borrowing." *Economics Letters*, 17, 27–31.

Altonji, Joseph G., Fumio Hayashi, and Laurence J. Kotlikoff (1997), "Parental altruism and inter vivos transfers: Theory and evidence." *Journal of Political Economy*, 105, 1121–1166.

Alvarez, Fernando and Urban J. Jermann (2000), "Efficiency, equilibrium, and asset pricing with risk of default." *Econometrica*, 68, 775–797.

Attanasio, Orazio and Steven J. Davis (1996), "Relative wage movements and the distribution of consumption." *Journal of Political Economy*, 104, 1227–1262.

Attanasio, Orazio and José-Vıctor Rıos-Rull (2000), "Consumption smoothing in island economies: Can public insurance reduce welfare?" *European Economic Review*, 44, 1225–1258.

Bernheim, B. Douglas, Bezalel Peleg, and Michael D. Whinston (1987), "Coalition proof Nash equilibria I: Concepts." *Journal of Economic Theory*, 42, 1–12.

Bernheim, B. Douglas and Debraj Ray (1989), "Collective dynamic consistency in repeated games." *Games and Economic Behavior*, 1, 295–326.

Bold, Tessa and Tobias Broer (2021), "Risk sharing in village economies revisited." *Journal of the European Economic Association*, 19, 3207–3248.

Bramoullé, Yann and Rachel Kranton (2007), "Risk-sharing networks." *Journal of Economic Behavior & Organization*, 64, 275–294.

Bulow, Jeremy and Kenneth Rogoff (1989), "Sovereign debt: Is to forgive to forget?" *American Economic Review*, 79, 43–50.

Carmichael, H. Lorne and W. Bentley MacLeod (1997), "Gift giving and the evolution of cooperation." *International Economic Review*, 38, 485–509.

Cole, Harold L. and Narayana R. Kocherlakota (2001), "Efficient allocations with hidden income and hidden storage." *Review of Economic Studies*, 68, 523–542.

Dixit, Avinash K. (2004), *Lawlessness and Economics: Alternative Modes of Governance*. Princeton University Press.

Easterly, William and Ross Levine (1997), "Africa's growth tragedy: policies and ethnic divisions." *The Quarterly Journal of Economics*, 112, 1203–1250.

Edward, Banfield (1958), "The moral basis of a backward society." *Glencoe*, 111, 85.

Farhi, Emmanuel, Mikhail Golosov, and Aleh Tsyvinski (2009), "A theory of liquidity and regulation of financial intermediation." *Review of Economic Studies*, 76, 973–992.

Farrell, Joseph and Eric Maskin (1989), "Renegotiation in repeated games." *Games and Economic Behavior*, 1, 327–360.

Fukuyama, Francis (1995), *Trust: The social virtues and the creation of prosperity*. Free Press, New York.

Fukuyama, Francis (2001), "Social capital, civil society and development." *Third World Quarterly*, 22, 7–20.

Genicot, Garance and Debraj Ray (2003), "Group formation in risk-sharing arrangements." *Review of Economic Studies*, 70, 87–113.

Ghosh, Parikshit and Debraj Ray (1996), "Cooperation in community interaction without information flows." *Review of Economic Studies*, 63, 491–519.

Golosov, Mikhael and Aleh Tsyvinski (2006), "Designing optimal disability insurance: A case for asset testing." *Journal of Political Economy*, 114, 257–279.

Greenberg, Joseph (1990), *The Theory of Social Situations: An Alternative Game-Theoretic Approach*. Cambridge University Press.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman (1988), "Investment, capacity utilization, and the real business cycle." *American Economic Review*, 402–417.

Greif, Avner (1993), "Contract enforceability and economic institutions in early trade: The maghribi traders' coalition." *The American economic review*, 525–548.

Guiso, Luigi, Luigi Pistaferri, and Fabiano Schivardi (2005), "Insurance within the firm." *Journal of Political Economy*, 113, 1054–1087.

Harris, Milton and Bengt Holmström (1982), "A Theory of Wage Dynamics." *Review of Economic Studies*, 49, 315–333.

Hellwig, Christian and Guido Lorenzoni (2009), "Bubbles and self-enforcing debt." *Econometrica*, 77, 1137–1164.

Jensen, Robert T (2004), "Do private transfers 'displace' the benefits of public transfers? evidence from south africa." *Journal of Public Economics*, 88, 89–112.

Juarez, Laura (2009), "Crowding out of private support to the elderly: Evidence from a demogrant in mexico." *Journal of Public Economics*, 93, 454–463.

Kahn, Charles M and Dilip Mookherjee (1992), "The good, the bad, and the ugly: Coalition proof equilibrium in infinite games." *Games and Economic Behavior*, 4, 101–121.

Kahn, Charles M. and Dilip Mookherjee (1995), "Coalition proof equilibrium in an adverse selection insurance economy." *Journal of Economic Theory*, 66, 113–138.

Kehoe, Timothy J. and David K. Levine (1993), "Debt-constrained asset markets." *Review of Economic Studies*, 60, 865–888.

Kehoe, Timothy J. and David K. Levine (2001), "Liquidity constrained markets versus debt constrained markets." *Econometrica*, 69, 575–598.

Knack, Stephen and Philip Keefer (1997), "Does social capital have an economic payoff? a cross-country investigation." *Quarterly Journal of Economics*, 112, 1251–1288.

Kocherlakota, Narayana R. (1996), "Implications of efficient risk sharing without commitment." *Review of Economic Studies*, 63, 595–609.

Kranton, Rachel E. (1996a), "The formation of cooperative relationships." *Journal of Law, Economics, and Organization*, 12, 214–233.

Kranton, Rachel E. (1996b), "Reciprocal exchange: A self-sustaining system." *American Economic Review*, 86, 830–851.

Krueger, Dirk and Fabrizio Perri (2006), "Does income inequality lead to consumption inequality? Evidence and theory." *Review of Economic Studies*, 73, 163–193.

Krueger, Dirk and Fabrizio Perri (2011), "Public versus private risk sharing." *Journal of Economic Theory*, 146, 920–956.

Krueger, Dirk and Harald Uhlig (2006), "Competitive risk sharing contracts with one-sided commitment." *Journal of Monetary Economics*, 53, 1661–1691.

Leonardi, Robert, Raffaella Y Nanetti, and Robert D Putnam (2001), *Making democracy work: Civic traditions in modern Italy*. Princeton University Press Princeton, NJ, USA.

Ligon, Ethan, Jonathan P. Thomas, and Tim Worrall (2002), "Informal insurance arrangements with limited commitment: Theory and evidence from village economies." *Review of Economic Studies*, 69, 209–244.

MacLeod, W. Bentley and James M. Malcomson (1989), "Implicit contracts, incentive compatibility, and involuntary unemployment." *Econometrica*, 57, 447–480.

Martins-da Rocha, V. Filipe and Mateus Santos (2019), "Self-enforcing debt and rational bubbles." Working paper, FGV.

Meghir, Costas, Mushfiq Mobarak, Corinna Mommaerts, and Melanie Morten (2022), "Migration and informal insurance: evidence from a randomized control trial and a structural model." *Review of Economic Studies*, 89, 452–480.

Morris, Stephen and Hyun Song Shin (2003), "Global games: Theory and applications." In *Advances in Economics and Econometrics (Proceedings of the Eighth World Congress of the Econometric Society)* (M. Dewatripont, L. Hansen, and S. Turnovsky, eds.), Cambridge University Press.

Morten, Melanie (2019), "Temporary migration and endogenous risk sharing in village india." *Journal of Political Economy*, 127, 1–46.

Nunn, Nathan (2008), "The long-term effects of Africa's slave trades." *The Quarterly Journal of Economics*, 123, 139–176.

Nunn, Nathan and Leonard Wantchekon (2011), "The slave trade and the origins of mistrust in africa." *American Economic Review*, 101, 3221–52.

Park, Yena (2014), "Optimal taxation in a limited commitment economy." *Review of Economic Studies*, 81, 884–918.

Phelan, Christopher (1995), "Repeated Moral Hazard and One-Sided Commitment." *Journal of Economic Theory*, 66, 488–506.

Portes, Alejandro and Patricia Landolt (1996), "The downside of social capital." *The American Prospect*, 26, 18–22.

Putnam, Robert D. (2000), *Bowling Alone: The Collapse and Revival of American Community*. Simon and Schuster.

Roth, Felix (2009), "Does too much trust hamper economic growth?" *Kyklos*, 62, 103–128.

Shapiro, Carl and Joseph Stiglitz (1984), "Equilibrium unemployment as a worker discipline device." *American Economic Review*, 74, 433–444.

Tabellini, Guido (2008a), "Presidental Address: Institutions and Culture." *Journal of the European Economic Association*, 6, 255–294.

Tabellini, Guido (2008b), "The Scope of Cooperation: Values and Incentives." *Quarterly Journal of Economics*, 123, 905–950.

Tabellini, Guido (2010), "Culture and Institutions: Economic Development in the Regions of Europe." *Journal of the European Economic Association*, 8, 677–716.

Thomas, Jonathan and Tim Worrall (1988), "Self-Enforcing Wage Contracts." *Review of Economic Studies*, 55, 541–554.

Townsend, Robert M (1994), "Risk and insurance in village india." *Econometrica: journal of the Econometric Society*, 539–591.

Udry, Christopher (1994), "Risk and insurance in a rural credit market: An empirical investigation in northern nigeria." *The Review of Economic Studies*, 61, 495–526.

von Neumann, John and Oskar Morgenstern (1944), *Theory of games and economic behavior*, 1st edition. Princeton University Press. 2nd edn 1947, 3rd edn 1953.

Woolcock, Michael (1998), "Social capital and economic development: Toward a theoretical synthesis and policy framework." *Theory and Society*, 27, 151–208.

Woolcock, Michael (2001), "The place of social capital in understanding social and economic outcomes." *Canadian Journal of Policy Research*, 2, 11–17.

Woolcock, Michael and Deepa Narayan (2000), "Social capital: Implications for development theory, research, and policy." *The World Bank Research Observer*, 15, 225–249.

Zak, Paul J and Stephen Knack (2001), "Trust and Growth." *The Economic Journal*, 111, 295–321.

Appendix

A Optimality of stationary risk sharing in Section 2

Since both agents in a compatible pair are ex ante identical, and utility is strictly concave, we can restrict attention to anonymous allocations (that is, consumption will not depend upon the name of the partner in a pair). An anonymous allocation specifies for all periods t, an agent's consumption $c(y^t)$ in period t for every possible sequence of $y^t \in Y^t$ of that agent's income shocks. Given a history y^t of income shocks for an agent, denote by $\overrightarrow{y}^t$ the history of income shocks of its partner (recall that income shocks are perfectly negatively correlated within a compatible pair). An anonymous allocation is *feasible* if for all t and all $y^t \in Y^t$,

$$c(y^t) + c(\overrightarrow{y}^t) = 2\overline{y}.$$

The lifetime utility from the allocation c is

$$W^0(c) := (1-\beta)\sum_{\tau=1}^{\infty}\sum_{y^\tau}\beta^{\tau-1}\Pr(y^\tau)u(c(y^\tau)),$$

and the continuation lifetime utility at an income history y^t is

$$W(y^t, c) := (1-\beta)u(c(y^t)) + (1-\beta)\sum_{\tau=1}^{\infty}\sum_{y^\tau}\beta^{\tau}\Pr(y^\tau)u(c(y^t y^\tau)),$$

where $y^t y^\tau$ denotes the $t+\tau$-history that is the concatenation of t-period history y^t and the τ-period history y^τ.

Lemma A.1 *Suppose c^* maximizes $W^0(c)$ over all feasible c satisfying*

$$W(y^t, c) \geq (1-\beta)u(y_t) + \beta F \qquad \forall t, y^t \in Y^t.$$

Then there exists a transfer $x^* \geq 0$ such that for all y^{t-1}, $c^*(y^{t-1}h) = h - x^*$.

Proof. We first argue that c^* is history independent. Suppose not. There are then two histories $\tilde{y}^\tau$ and $\hat{y}^\tau$ such that $c^*(\tilde{y}^\tau h) \neq c^*(\hat{y}^\tau h)$. Define a new consumption plan $c^\dagger$ as follows: for histories y^t shorter than τ or whose τ initial periods differ from $\tilde{y}^\tau$ or $\hat{y}^\tau$, set $c^\dagger(y^t h) = c^*(y^t h)$. For histories $y^t = (y^\tau, y^{t-\tau})$, $y^\tau \in \{\tilde{y}^\tau, \hat{y}^\tau\}$, set

$$c^\dagger(y^\tau, y^{t-\tau}h) = \frac{1}{2}[c^*(\tilde{y}^\tau, y^{t-\tau}h) + c^*(\hat{y}^\tau, y^{t-\tau}h)].$$

Since $W(y^t, c)$ is a concave function of c, $W(y^\tau, y^{t-\tau}h, c^\dagger) \geq \frac{1}{2}[W(\tilde{y}^\tau, y^{t-\tau}h, c^*)+W(\hat{y}^\tau, y^{t-\tau}h, c^*)]$, and so $c^\dagger$ satisfies all the constraints. Moreover, $W^0(c^\dagger) > W^0(c^*)$, and so c^* cannot have been optimal.

Since income is perfectly negative correlated, feasibility implies for all histories $\tilde{y}^\tau$ and $\hat{y}^\tau$, $c(\tilde{y}^\tau \ell) = c^*(\hat{y}^\tau \ell)$. In other words, the optimal consumption is given by a history independent sequence of transfers $(x^t)_t$ from the rich agent to the poor agent.

Suppose $x^t < x^\tau$ for some $t \neq \tau$. Since the outside option is binding in every period, the lifetime utility from period $t+1$ is larger than from period $\tau+1$. Define a sequence of transfers $(\hat{x}^s)_s$ by $\hat{x}^s = x^s$ for $s \leq \tau$ and $x^{t+s-\tau}$ for $s \geq \tau+1$. This raises the level of lifetime utility from period $\tau+1$, raising $W^0(c)$. Hence, the optimal transfers must be stationary. □

B Proofs for Section 6

B.1 Proof of Proposition 2

Since $\mathcal{C}(F)$ is a convex set and $W^0(c)$ is a strictly concave function, there is a unique allocation $\mathbb{c} \in \mathcal{C}(F)$ satisfying $W^0(\mathbb{c}) = \mathbb{V}(F)$ and so F is the ex ante value of the social norm if $\mathbb{c}$ is the strong social norm. It remains to argue that $\mathbb{c}$ is the strong social norm.

Since $\mathbb{c} \in \mathcal{C}(F)$, $\mathbb{c}$ satisfies (14). If $\mathbb{c}$ is not the strong social norm, there exists another social norm c' with

$$W^0(c') > W^0(\mathbb{c}).$$

Then, since c' is internally-incentive compatible, for all $t \geq 1$ and $y^t \in Y^t$,

$$\begin{aligned} W(y^t, c') &\geq (1-\beta)u(y_t) + \beta[\pi W^0(c') + (1-\pi)V^A] \\ &> (1-\beta)u(y_t) + \beta[\pi W^0(\mathbb{c}) + (1-\pi)V^A] \\ &= (1-\beta)u(y_t) + \beta F, \end{aligned}$$

and so $c' \in \mathcal{C}(F)$, implying $W^0(\mathbb{c})$ could not be a fixed point of $\mathcal{T}(\,\cdot\,;\pi)$.

Finally, the fixed point is unique because $\mathbb{V}(\,\cdot\,)$, and thus $\mathcal{T}(\,\cdot\,;\pi)$, is weakly decreasing in F.

B.2 Proof of Proposition 3

Suppose $\mathbb{c}$ is a strong social norm that is not first-best insurance, and let $F = (1-\pi)W^0(\mathbb{c}) + \pi V^A$. From Proposition 2, $\mathbb{c}$ is F-incentive compatible and so $\mathbb{V}(F) = W^0(\mathbb{c}) \geq F$. Since $\mathbb{c}$ is not first-best insurance, $F > F^{FB}$. We first prove a collection of intermediate results.

Lemma B.1 *There exists $\delta_{t+1} < 1$ such that if F-incentive compatibility holds as a strict inequality at y^{t+1}, then*

$$\frac{u'(\mathbb{c}(y^t))}{u'(\mathbb{c}(y^{t+1}))} = \delta_{t+1}$$

and so

$$\mathbb{c}(y^t) > \mathbb{c}(y^{t+1}).$$

Proof. We first argue that if F-incentive compatibility holds strictly at $\tilde{y}^{t+1}$, then for all $\hat{y}^{t+1}$

$$\frac{u'(\mathbb{c}(\tilde{y}^t))}{u'(\mathbb{c}(\tilde{y}^{t+1}))} \leq \frac{u'(\mathbb{c}(\hat{y}^t))}{u'(\mathbb{c}(\hat{y}^{t+1}))}. \tag{B.1}$$

Suppose not, so that (B.1) holds with a strict inequality in the reverse direction.

Define a new allocation $c^\dagger$ by setting

$$c^\dagger(y^\tau) = \begin{cases} \mathbb{c}(\tilde{y}^t) + \varepsilon, & y^\tau = \tilde{y}^t \\ \mathbb{c}(\hat{y}^t) - \varepsilon, & y^\tau = \hat{y}^t \\ \mathbb{c}(\tilde{y}^{t+1}) - \eta, & y^\tau = \tilde{y}^{t+1} \\ \mathbb{c}(\hat{y}^{t+1}) + \eta, & y^\tau = \hat{y}^{t+1} \\ \mathbb{c}(y^\tau), & \text{otherwise.} \end{cases}$$

Since $\Pr(\hat{y}^t) = \Pr(\tilde{y}^t)$ and $\Pr(\hat{y}^{t+1}) = \Pr(\tilde{y}^{t+1})$, the allocation $c^\dagger$ is feasible.

Choosing $\eta = \eta(\varepsilon)$ so that

$$u(\mathbb{c}(\tilde{y}^t) + \varepsilon) + \frac{\beta}{2}u(\mathbb{c}(\tilde{y}^{t+1}) - \eta(\varepsilon)) = u(\mathbb{c}(\tilde{y}^t)) + \frac{\beta}{2}u(\mathbb{c}(\tilde{y}^{t+1}))$$

ensures that F-incentive compatibility is satisfied along the sequence $\tilde{y}^t$. For small η, it is also satisfied at $\tilde{y}^{t+1}$.

Differentiating with respect to ε and evaluating at $\varepsilon = 0$, we get

$$\eta'(0) = \frac{2u'(\mathbb{c}(\tilde{y}^t))}{\beta u'(\mathbb{c}(\tilde{y}^{t+1}))}.$$

At $\varepsilon = 0$, the derivative of

$$u(\mathbb{c}(\hat{y}^t) - \varepsilon) + \frac{\beta}{2} u(\mathbb{c}(\hat{y}^{t+1}) + \eta(\varepsilon))$$

is

$$\begin{aligned} -u'(\mathbb{c}(\hat{y}^t)) + \frac{\beta}{2} u'(\mathbb{c}(\hat{y}^{t+1}))\eta'(0) &= -u'(\mathbb{c}(\hat{y}^t)) + \frac{\beta}{2} u'(\mathbb{c}(\hat{y}^{t+1})) \frac{2u'(\mathbb{c}(\tilde{y}^t))}{\beta u'(\mathbb{c}(\tilde{y}^{t+1}))} \\ &= u'(\mathbb{c}(\hat{y}^{t+1})) \left\{ -\frac{u'(\mathbb{c}(\hat{y}^t))}{u'(\mathbb{c}(\hat{y}^{t+1}))} + \frac{u'(\mathbb{c}(\tilde{y}^t))}{u'(\mathbb{c}(\tilde{y}^{t+1}))} \right\} \\ &> 0. \end{aligned}$$

This implies that the values of the agents with histories $\hat{y}^t$ and $\hat{y}^{t+1}$ have increased, and so the ex ante value of $c^\dagger$ must exceed $\mathbb{c}$, contradicting the optimality of $\mathbb{c}$.

Hence, (B.1) must hold as written. If F-incentive compatibility also holds strictly at $\hat{y}^{t+1}$, then the weak inequality holds as an equality.

We now argue that if F-incentive compatibility holds strictly at $\tilde{y}^{t+1}$, then

$$u'(\mathbb{c}(\tilde{y}^t)) < u'(\mathbb{c}(\tilde{y}^{t+1})).$$

If not, then for all histories y^{t+1} at which F-incentive compatibility holds strictly,

$$u'(\mathbb{c}(y^t)) \geq u'(\mathbb{c}(y^{t+1})).$$

But this implies that for all such y^{t+1},

$$\mathbb{c}(y^t) \leq \mathbb{c}(y^{t+1}).$$

Feasibility then implies that for at least one history $\hat{y}^{t+1}$ (at which F-incentive compatibility holds with equality), $\mathbb{c}(\hat{y}^t) \geq \mathbb{c}(\hat{y}^{t+1})$. If there is no history at which the inequality is strict, then $\mathbb{c}(y^t) = \mathbb{c}(y^{t+1})$ for all y^{t+1}, which implies that $\mathbb{c}$ is the first best allocation. But $F > F^{FB}$ precludes the first best allocation as a solution.

To complete the argument, suppose there is a history $\hat{y}^{t+1}$ for which $\mathbb{c}(\hat{y}^t) > \mathbb{c}(\hat{y}^{t+1})$. Then,

$$u'(\mathbb{c}(\hat{y}^t)) < u'(\mathbb{c}(\hat{y}^{t+1})),$$

and so the reverse direction of (B.1) holds as a strict inequality for any history $\tilde{y}^{t+1}$ at which F-incentive compatibility holds strictly and $\hat{y}^{t+1}$. The rest of argument applies without change to yield a contradiction.

□

Lemma B.2 *At the optimal allocation* $\mathbb{c}$*, if the incentive constraint holds with equality at* $\tilde{y}^t$ *and* $\hat{y}^t$ *with* $\tilde{y}_t = \hat{y}_t$*, then*

$$\mathbb{c}(\tilde{y}^t) = \mathbb{c}(\hat{y}^t).$$

Proof. Suppose not. then the incentive constraint holds with equality at two histories $\tilde{y}^t$ and $\hat{y}^t$ with $\tilde{y}_t = \hat{y}_t$, and

$$\mathbb{c}(\tilde{y}^t) \neq \mathbb{c}(\hat{y}^t).$$

Define a new consumption allocation $c^\dagger$ as follows:

$$c^\dagger(y^\tau) = \begin{cases} \frac{1}{2}\mathbb{c}(\hat{y}^t{}_t y^\tau) + \frac{1}{2}\mathbb{c}(\tilde{y}^t{}_t y^\tau), & \tau \geq t, {}^t y^\tau = \tilde{y}^t, \hat{y}^t, \\ \mathbb{c}(y^\tau), & \text{otherwise,} \end{cases}$$

where ${}_t y^\tau$ is the last $\tau - t$ periods of the income history y^t (so that $y^\tau = {}^t y^\tau {}_t y^\tau$)). Since $\Pr(\tilde{y}^t) = \Pr(\hat{y}^t)$, $c^\dagger$ satisfies (13).

Moreover, the incentive constraints are satisfied at all histories:

1. For $\tau < t$, since the incentive constraints bind at two histories $\tilde{y}^t$ and $\hat{y}^t$ with $\tilde{y}_t = \hat{y}_t$, $W(\tilde{y}^t, \mathbb{c}) = W(\hat{y}^t, \mathbb{c})$, and so $W(y^t, c^\dagger) \geq W(y^t, \mathbb{c})$ for all y^t (with equality holding for $y^t \notin \{\tilde{y}^t, \hat{y}^t\}$). Hence,

$$\begin{aligned} W(y^\tau, c^\dagger) &= (1-\beta)u(\mathbb{c}(y^\tau)) + (1-\beta)\sum_{r=1}^{t-\tau-1} \beta^r \sum_{y^r} \Pr(y^r) u(\mathbb{c}(y^\tau y^r)) \\ &\qquad + \beta^{t-\tau} \sum_{y^t} \Pr(y^t) W(y^t, c^\dagger) \\ &\geq W(y^\tau, \mathbb{c}). \end{aligned}$$

2. For $\tau \geq t$, the concavity of u implies

$$W(y^t, c^\dagger) \geq \min\{W(\hat{y}^t{}_t y^\tau, \mathbb{c}), W(\tilde{y}^t{}_t y^\tau, \mathbb{c})\} \geq W^F(y_\tau).$$

Finally, concavity implies $W^0(c^\dagger) > W^0(\mathbb{c})$, which is impossible, since $\mathbb{c}$ is by assumption optimal. □

Lemma B.3 *In the optimal allocation, F-incentive compatibility holds with equality at all y^t for which $y_t = h$, and so for all y^{t-1},*

$$W(y^{t-1}h, \mathbb{c}) = W^F(h) := (1-\beta)u(h) + \beta F.$$

Proof. Since $F > F^{FB}$,

$$(1-\beta)u(\bar{y}) + \beta V^{FB} < W^F(h),$$

and so

$$(1-\beta)u(\bar{y}) + \beta \mathbb{V}(F) < W^F(h).$$

Thus, F-incentive compatibility at $\hat{y}^{t-1}h$ requires $\mathbb{c}(\hat{y}^{t-1}h) > \bar{y}$. Suppose

$$W(\hat{y}^{t-1}h, \mathbb{c}) > W^F(h).$$

Define

$$c^\varepsilon(y^\tau) = \begin{cases} \mathbb{c}(\hat{y}^{t-1}h) - \varepsilon, & y^\tau = \hat{y}^{t-1}h, \\ \mathbb{c}(\hat{y}^{t-1}\ell) + \varepsilon, & y^\tau = \hat{y}^{t-1}\ell, \\ \mathbb{c}(y^t), & \text{otherwise.} \end{cases}$$

Since h and ℓ are equally likely, c^ε satisfies feasibility. For sufficiently small $\varepsilon > 0$, c^ε satisfies internal-incentive compatibility, and so we have a contradiction (since c^ε has higher ex ante utility than $\mathbb{c}$). Thus, the incentive constraint holds with equality at all $\hat{y}^t$ for which $\hat{y}_t = h$. □

Lemma B.4 *For all $\tilde{y}^{t-1}, \hat{y}^{t-1}$,*

$$\mathbb{c}(\tilde{y}^{t-1}) \geq \mathbb{c}(\hat{y}^{t-1}) \implies \mathbb{c}(\tilde{y}^{t-1}y) \geq \mathbb{c}(\hat{y}^{t-1}y) \text{ and } W(\tilde{y}^{t-1}\ell, \mathbb{c}) \geq W(\hat{y}^{t-1}\ell, \mathbb{c}).$$

Proof. Lemmas B.2 and B.3, imply

$$\mathbb{c}(\tilde{y}^{t-1}h) = \mathbb{c}(\hat{y}^{t-1}h) \qquad \forall \tilde{y}^{t-1}, \hat{y}^{t-1}.$$

Suppose now, en route to a contradiction that there are two histories $\tilde{y}^{t-1}$ and $\hat{y}^{t-1}$ such that

$$\mathbb{c}(\tilde{y}^{t-1}) \geq \mathbb{c}(\hat{y}^{t-1}) \text{ and } \mathbb{c}(\tilde{y}^{t-1}\ell) < \mathbb{c}(\hat{y}^{t-1}\ell).$$

The idea is to construct a dominating consumption allocation by moving consumption from the relatively high-consumption histories to the low-consumption histories. For any small $\varepsilon > 0$, define $\eta(\varepsilon)$ as the value η solving

$$u(\mathbb{c}(\tilde{y}^{t-1}) - \eta) + \tfrac{\beta}{2}u(\mathbb{c}(\tilde{y}^{t-1}\ell) + \varepsilon) = u(\mathbb{c}(\tilde{y}^{t-1})) + \tfrac{\beta}{2}u(\mathbb{c}(\tilde{y}^{t-1}\ell)),$$

and define a new consumption allocation as

$$c^{\varepsilon}(y^{\tau}) = \begin{cases} \mathbb{c}(y^{\tau}) - \eta(\varepsilon), & y^{\tau} = \tilde{y}^{t-1}, \\ \mathbb{c}(y^{\tau}) + \eta(\varepsilon), & y^{\tau} = \hat{y}^{t-1}, \\ \mathbb{c}(y^{\tau}) + \varepsilon, & y^{\tau} = \tilde{y}^{t-1}\ell, \\ \mathbb{c}(y^{\tau}) - \varepsilon, & y^{\tau} = \hat{y}^{t-1}\ell, \\ \mathbb{c}(y^{\tau}), & \text{otherwise.} \end{cases}$$

From the concavity of u, $u'(\mathbb{c}(\tilde{y}^{t-1})) \leq u'(\mathbb{c}(\hat{y}^{t-1}))$ and

$$\xi := u'(\mathbb{c}(\tilde{y}^{t-1}\ell)) - u'(\mathbb{c}(\hat{y}^{t-1}\ell)) > 0.$$

Moreover, the function η is $\mathcal{C}^1$ with $\eta'(0) > 0$. Then we have (where each function o_j, for $j = 1, \ldots, 4$ satisfies $o_j(\varepsilon)/\varepsilon \to 0$ as $\varepsilon \to 0$),

$$\begin{aligned} \tfrac{\beta}{2}\{u(\mathbb{c}(\hat{y}^{t-1}\ell)) - u(\mathbb{c}(\hat{y}^{t-1}\ell) - \varepsilon)\} &= \tfrac{\beta}{2}\{u'(\mathbb{c}(\hat{y}^{t-1}\ell))\varepsilon + o_1(\varepsilon)\} \\ &= \tfrac{\beta}{2}\{u'(\mathbb{c}(\tilde{y}^{t-1}\ell))\varepsilon - \xi\varepsilon + o_1(\varepsilon)\} \\ &= \tfrac{\beta}{2}\{u(\mathbb{c}(\tilde{y}^{t-1}\ell) + \varepsilon) - u(\mathbb{c}(\tilde{y}^{t-1}\ell)) - \xi\varepsilon\} + o_2(\varepsilon) \\ &= u(\mathbb{c}(\tilde{y}^{t-1})) - u(\mathbb{c}(\tilde{y}^{t-1}) - \eta(\varepsilon)) - \tfrac{\beta}{2}\xi\varepsilon + o_2(\varepsilon) \\ &= u'(\mathbb{c}(\tilde{y}^{t-1}))\eta(\varepsilon) - \tfrac{\beta}{2}\xi\varepsilon + o_3(\varepsilon) \\ &\leq u'(\mathbb{c}(\hat{y}^{t-1}))\eta(\varepsilon) - \tfrac{\beta}{2}\xi\varepsilon + o_3(\varepsilon) \\ &= u(\mathbb{c}(\hat{y}^{t-1}) + \eta(\varepsilon)) - u(\mathbb{c}(\hat{y}^{t-1})) - \tfrac{\beta}{2}\xi\varepsilon + o_4(\varepsilon). \end{aligned}$$

Rearranging,

$$u(\mathbb{c}(\hat{y}^{t-1})) + \frac{\beta}{2}u(\mathbb{c}(\hat{y}^{t-1}\ell)) + \tfrac{\beta}{2}\xi\varepsilon \leq u(\mathbb{c}(\hat{y}^{t-1}) + \eta(\varepsilon)) + \frac{\beta}{2}u(\mathbb{c}(\hat{y}^{t-1}\ell) - \varepsilon) + o_4(\varepsilon),$$

and so, if $\varepsilon > 0$ is sufficiently small that

$$|o_4(\varepsilon)| < \tfrac{\beta}{2}\xi\varepsilon,$$

we have

$$u(\mathbb{c}(\hat{y}^{t-1})) + \frac{\beta}{2}u(\mathbb{c}(\hat{y}^{t-1}\ell)) < u(\mathbb{c}(\hat{y}^{t-1}) + \eta(\varepsilon)) + \frac{\beta}{2}u(\mathbb{c}(\hat{y}^{t-1}\ell) - \varepsilon).$$

Since $\mathbb{c}(y^\tau) \leq c^\varepsilon(y^\tau)$ for all y^τ, with a strict inequality on one positive-measure history, $\mathbb{c}$ cannot have been optimal.

The inequality on continuation values then immediately follows from the following calculation: For any y^t, denote by $y^t\ell^k$ the history formed by adding k periods of ℓ after y^t (so that $y^t\ell^0 = y^t$). Then,

$$\begin{aligned} W(y^t, \mathbb{c}) &= (1-\beta)u(\mathbb{c}(y^t)) + \tfrac{\beta}{2}\{W^F(h) + W(y^t\ell, \mathbb{c})\} \\ &= (1-\beta)\sum_{k=0}^{\infty}(\tfrac{\beta}{2})^k u(\mathbb{c}(y^t\ell^k)) + \tfrac{\beta}{2-\beta}W^F(h). \end{aligned} \tag{B.2}$$

□

Lemma B.5 *If F-incentive compatibility holds with equality at $y^t\ell$, then for all $\hat{y}^t$,*

$$\mathbb{c}(y^t\ell) \leq \mathbb{c}(\hat{y}^t\ell).$$

Proof. Suppose

$$\mathbb{c}(y^t\ell) > \mathbb{c}(\hat{y}^t\ell).$$

Then, from Lemma B.4,

$$\begin{aligned} u(\mathbb{c}(y^t\ell)) + \tfrac{\beta}{2}\{W^F(h) + W(y^t\ell\ell, \mathbb{c})\} &> \mathbb{c}(\hat{y}^t\ell) + \tfrac{\beta}{2}\{W^F(h) + W(\hat{y}^t\ell\ell, \mathbb{c})\} \\ &\geq W^F(\ell), \end{aligned}$$

which is impossible if F-incentive compatibility holds with equality at $y^t\ell$. □

Lemma B.6 *Suppose F-incentive compatibility holds with equality at some $y^{t-1}\ell$ in an optimal allocation. Then F-incentive compatibility holds with equality at $y^{t-1}\ell\ell$.*

Proof. Suppose F-incentive compatibility binds at $y^{t-1}\ell$ but not at $y^{t-1}\ell^2$. Then

$$\begin{aligned} u(\mathbb{c}(y^{t-1}\ell)) + \tfrac{\beta}{2}\{W^F(h) + W(y^{t-1}\ell^2, \mathbb{c})\} &= W^F(\ell), \\ W(y^{t-1}\ell^2, \mathbb{c}) = u(\mathbb{c}(y^{t-1}\ell^2)) + \tfrac{\beta}{2}\{W^F(h) + W(y^{t-1}\ell^3, \mathbb{c})\} &> W^F(\ell), \end{aligned}$$

and (because the last F-incentive compatibility constraint is strict)

$$\mathbb{c}(y^{t-1}\ell) > \mathbb{c}(y^{t-1}\ell^2).$$

Since

$$u(\mathbb{c}(y^{t-1}\ell)) > u(\mathbb{c}(y^{t-1}\ell^2)),$$

we therefore have (because F-incentive compatibility holds with equality at $y^{t-1}\ell$)

$$W(y^{t-1}\ell^3, \mathbb{c}) > W(y^{t-1}\ell^2, \mathbb{c}) > W^F(\ell), \tag{B.3}$$

and F-incentive compatibility is also strict at $y^{t-1}\ell^3$. This implies

$$\mathbb{c}(y^{t-1}\ell) > \mathbb{c}(y^{t-1}\ell^3),$$

and so

$$W(y^{t-1}\ell^4, \mathbb{c}) > W(y^{t-1}\ell^2, \mathbb{c}) > W^F(\ell).$$

Repeated applications of this argument shows that F-incentive compatibility is strict for any history $y^{t-1}\ell^r$, $r \geq 2$, and so $(\mathbb{c}(y^{t-1}\ell^r))_{r\geq 1}$ is a monotonically declining sequence. Hence, from (B.2), so is $(W(y^{t-1}\ell^r, \mathbb{c}))_{r\geq 1}$. But this contradicts (B.3). □

Lemma B.7 *If F-incentive compatibility holds with equality at $y^t\ell$, then*

$$\mathbb{c}(y^t\ell) = c_\ell,$$

where $c_\ell > \ell$ is the unique consumption satisfying

$$u(c_\ell) = u(\ell) + \beta(F - V^A) > u(\ell).$$

Note that c_ℓ is an increasing function of F, so that for $F > F^{FB}$ (i.e., for $\pi > \pi^{FB}$) but arbitrarily close, c_ℓ is bounded away from $\overline{y}$.

Proof. Since F-incentive compatibility holds with equality at $y^t\ell$ (and so at $y^t\ell^2$), we have

$$(1-\beta)u(\mathbb{c}(y^t\ell)) + \tfrac{\beta}{2}\{W^F(h) + W^F(\ell)\} = W^F(\ell).$$

Rearranging and dividing by $(1-\beta)$ yields

$$u(\mathbb{c}(y^t\ell)) = (1 - \tfrac{\beta}{2})u(\ell) - \tfrac{\beta}{2}u(h) + \beta F,$$

which is the displayed equation (recall that $V^A = Eu(y)$). $\square$

Lemma B.8 *F-incentive compatibility holds strictly in the initial period at ℓ and after any history of the form $y^t h\ell$.*

Proof. If F-incentive compatibility holds with equality in the initial period, then

$$\begin{aligned}\mathbb{V}(F) &= \tfrac{1}{2}(1-\beta)\{u(h)+u(\ell)\}+\beta F\\ &= (1-\beta)V^A+\beta F\\ &\implies \mathbb{V}(F) < F,\end{aligned}$$

which is impossible, since $\mathbb{V}(F) \geq F$.

Suppose F-incentive compatibility holds with equality after a history of the form $y^t h\ell$. Since F-incentive compatibility always holds with equality after any realization of h, we have

$$\begin{aligned}&(1-\beta)u(h)+\beta F = (1-\beta)u(\mathbb{c}(y^t h))+\beta\{(1-\beta)V^A+\beta F\}\\ &\implies (1-\beta)u(h) = (1-\beta)u(\mathbb{c}(y^t h))-\beta(1-\beta)(F-V^A)\\ &\implies u(\mathbb{c}(y^t h)) = u(h)+\beta(F-V^A)\\ &\implies \mathbb{c}(y^t h) > h,\end{aligned}$$

which is ruled out by feasibility and $\mathbb{c}(y^{t+1}) \geq c_\ell > \ell$. $\square$

Lemma B.9 *For $t > L$, $\mathbb{c}(y^t h) < \mathbb{c}(y^{t+k}h)$ for all y^{t+k} and $k \geq 2$.*

Proof. We prove by contradiction, so suppose there is a $k \geq 2$ and history $\tilde{y}^{t+k}$, $t > L$, such that $\mathbb{c}(\tilde{y}^t h) \geq \mathbb{c}(\tilde{y}^{t+k}h)$.

Define a new allocation $c^\dagger$ by setting

$$c^\dagger(y^\tau) = \begin{cases} \mathbb{c}(\tilde{y}^t h)-\varepsilon, & y^\tau = \tilde{y}^t h,\\ \mathbb{c}(\tilde{y}^{t+1-L}\ell^L)+\varepsilon, & y^\tau = \tilde{y}^{t+1-L}\ell^L,\\ \mathbb{c}(\tilde{y}^t h^{k+1})+\eta, & y^\tau = \tilde{y}^t h^{k+1},\\ \mathbb{c}(\tilde{y}^{t+1-L}\ell^L h^{k-1}\ell)-\eta, & y^\tau = \tilde{y}^{t+1-L}\ell^L h^{k-1}\ell,\\ \mathbb{c}(y^\tau), & \text{otherwise.}\end{cases}$$

Since all histories of the same length have the same probability, the allocation $c^\dagger$ is feasible.

Choosing $\eta = \eta(\varepsilon)$ so that

$$u(\mathbb{c}(\tilde{y}^t h) - \varepsilon) + \frac{\beta^k}{2^k} u(\mathbb{c}(\tilde{y}^t h^{k+1}) + \eta(\varepsilon)) = u(\mathbb{c}(\tilde{y}^t h)) + \frac{\beta^k}{2^k} u(\mathbb{c}(\tilde{y}^t h^{k+1}))$$

ensures that F-incentive compatibility is satisfied on the sequence $\tilde{y}^t h$. For small η, it is also satisfied at $\tilde{y}^{t+1-L}\ell^L h^{k-1}\ell$ (from Lemma B.8, $\mathbb{c}$ satisfies F-incentive compatibility strictly there).

Differentiating with respect to ε and evaluating at $\varepsilon = 0$, we get

$$\eta'(0) = \frac{2^k u'(\mathbb{c}(\tilde{y}^t h))}{\beta^k u'(\mathbb{c}(\tilde{y}^t h^{k+1}))} \leq \frac{2^k}{\beta^k}$$

(where the inequality follows from our beginning supposition). At $\varepsilon = 0$, the derivative of

$$u(\mathbb{c}(\tilde{y}^{t+1-L}\ell^L) + \varepsilon) + \frac{\beta^k}{2^k} u(\mathbb{c}(\tilde{y}^{t+1-L}\ell^L h^{k-1}\ell) - \eta(\varepsilon))$$

is

$$\begin{aligned} u'(\mathbb{c}(\tilde{y}^{t+1-L}\ell^L) - \frac{\beta^k}{2^k} u'(\mathbb{c}(\tilde{y}^{t+1-L}\ell^L h^{k-1}\ell))\eta'(0) \\ \geq u'(\mathbb{c}(\tilde{y}^{t+1-L}\ell^L) - u'(\mathbb{c}(\tilde{y}^{t+1-L}\ell^L h^{k-1}\ell)) \\ > 0, \end{aligned}$$

where the strict inequality is an implication of

$$\mathbb{c}(\tilde{y}^{t+1-L}\ell^L) = c_\ell(F) < \mathbb{c}(\tilde{y}^{t+1-L}\ell^L h^{k-1}\ell)).$$

This implies both that F-incentive compatibility is strictly satisfied at $\tilde{y}^{t+1-L}\ell^L$ and that the ex ante value of $c^\dagger$ must exceed $\mathbb{c}$, contradicting the optimality of $\mathbb{c}$.

□

Lemma B.10 *There exists $\delta_* < 1$ such that $(\delta_t) \to \delta_*$ and consumptions $(\mathbb{c}_*(h\ell^k))$ such that for all $k \geq 0$,*

$$\mathbb{c}(y^{t-1-k}h\ell^k) = c_t(h\ell^k) \to \mathbb{c}_*(h\ell^k).$$

Proof. We first argue that $(c_t(h))$ converges. Define the sequence (x_t) by setting $x_t := c_t(h)$. Since the sequence takes values in a compact set, it has a convergent subsequence (x_{t_k}) with limit $x_* =: \mathbb{c}_*(h)$. If (x_t) does not converge to x_*, Lemma B.9 implies there is an $\varepsilon > 0$ and a different subsequence (x_{t_n}) with $x_{t_n} < x_* - \varepsilon$. But this is impossible, since for

t_k large and $t_n > t_k + 2$, we have

$$x_{t_n} > x_{t_k} > x_* - \varepsilon.$$

We now argue that (δ_t) converges to some limit $\delta_* < 1$. Given an arbitrary $c(h)$ and $\delta \in (0,1)$, define inductively the sequence $(c^\dagger(h\ell^k))$ for $k \geq 1$ by setting $c^\dagger(h) = c(h)$ and solving

$$u'(c^\dagger(h\ell^{k-1})) = \delta u'(c^\dagger(h\ell^k)) \tag{B.4}$$

for $c^\dagger(h\ell^k)$. Define $c(h\ell^k) = \max\{c^\dagger(h\ell^k), c_\ell\}$ and let $\delta(c(h))$ be the unique value of δ for which $(c(h\ell^k))_{k\geq 0}$ is exactly feasible (when it exists). Finally, define $\delta_* := \delta(\mathbb{c}_*(h))$. Observe that $\delta_* < 1$ (since $\delta^* = 1$ implies constant consumption, violating feasibility). Since $\delta(\cdot)$ is continuous when it is well defined, we have

$$\delta_t = \delta(c_t(h)) \to \delta(\mathbb{c}_*(h)) = \delta_*.$$

Since (B.4) is continuous in δ, the consumptions also converge. □

Lemma B.11 *Suppose $\pi > \pi^{FB}$. In the optimal allocation, there exists L such that the incentive constraint holds with equality at any history of the form $y^t\ell^L$.*

Proof. Lemma B.4 implies that optimal consumption in any period is determined by the number of ℓ realizations after the last h realization. From Lemma B.6, once the ℓ incentive constraint holds with equality, it continues to bind after each subsequent ℓ realization.

We need to prove that the number of ℓ realizations before the ℓ incentive constrain binds is bounded as we vary the period in which h is realized.

We prove by contradiction: Suppose there is a subsequence of periods with the property that the number of ℓ realizations after an h realization before the ℓ-incentive constraint holds with equality goes to ∞. Without loss of generality, assume there is a subsequence $(t_n)_n$ with the property that the ℓ-incentive constraint holds strictly in period t_n after a history $y^{t_n-n-1}h\ell^n$. This implies that the ℓ-incentive constraint holds strictly in period t_n after all histories $y^{t_n-k-1}h\ell^k$ for all $1 \leq k \leq n$. But this implies $\delta_{t_n} \to 1$, which is impossible by Lemma B.10. □

Proof of Proposition 3.

Lemmas B.2 and B.3 imply that for all y^{t-1} and $\hat{y}^{t-1}$,

$$\mathbb{c}(y^{t-1}h) = \mathbb{c}(\hat{y}^{t-1}h) =: c_t(h).$$

From Lemma B.8, F-incentive compatibility holds strictly at the history $y^t h\ell$, and so from Lemma B.1, $\mathbb{c}(y^t h\ell)$ is determined by (23), and so can be denoted $c_{t+1}(h\ell)$. Equation (23) continues to determine $c_{t+k}(h\ell^k)$ as a decreasing sequence as long as F-incentive compatibility holds strictly. From Lemma B.6, once F-incentive compatibility holds with equality, it continues to hold with equality after additional ℓ realizations. By Lemma B.11, there is an L such that incentive compatibility binds at $y^t \ell^L$, and so from Lemma B.7, $\mathbb{c}(y^t \ell^L) = c_\ell$. Finally, the convergence of consumptions is Lemma B.10. Lemma B.9 implies convergence does not occur in finite time and declining risk sharing over time. □

B.3 Proof of Proposition 5

We first establish two preliminary results.

Lemma B.12

1. *$\mathcal{C}(F') \supset \mathcal{C}(F'')$ for $F' < F''$, and so $\mathcal{C}(F) \neq \varnothing$ for all $F \leq \overline{F}$.*
2. *$\mathcal{C}(F)$ is closed and convex for all $F \leq \overline{F}$.*
3. *$\mathcal{C}$ is a continuous correspondence at all $F \leq \overline{F}$ (at $\overline{F}$, the continuity is from the left).*

Proof.

1. This is immediate.
2. This is also immediate.
3. Since $\mathcal{C}$ is a decreasing correspondence in F, we need only show upper hemicontinuity from the left and lower hemicontinuity from the right. Upper hemicontinuity is immediate, since all the constraints are closed. Turning to lower hemicontinuity, we need to show that if $c \in \mathcal{C}(F)$ and $(F_k)_k$ is a sequence with $F_k \searrow F$, then there exists $c_k \in \mathcal{C}(F_k)$ with $c_k \to c$. Fix $c^\dagger \in \mathcal{C}(\overline{F})$. We now verify that for all k, there exists $\alpha_k \in [0,1]$ such that $\alpha_k c^\dagger + (1-\alpha_k)c \in \mathcal{C}(F_k)$ and $\alpha_k \to 0$.

   Fix k, and let $\alpha_k = (F_k - F)/(\overline{F} - F_k) > 0$. Then,

$$\begin{aligned} W(y^t, \alpha_k c^\dagger + (1-\alpha_k)c) &\geq \alpha_k W(y^t, c^\dagger) + (1-\alpha_k) W(y^t, c) \\ &\geq (1-\beta)u(y_t) + \alpha_k \beta \overline{F} + (1-\alpha_k)\beta F \\ &= (1-\beta)u(y_t) + \beta F_k, \end{aligned}$$

and so F-incentive compatibility (20) is satisfied. Since (13) is trivially satisfied, we are done.

□

Lemma B.13 *If $\beta > \beta^{FB}$, then $\overline{F} > F^{FB}$.*

Proof. Recall the allocation c_ζ defined in (18):

$$c_\zeta(y^t) = \begin{cases} h - \zeta, & y_t = h, \\ \ell + \zeta, & y_t = \ell. \end{cases}$$

We now argue that there exists $\xi > 0$ such that for all $F \in (F^{FB}, F^{FB} + \xi]$, for

$$\zeta = \zeta^{FB} - 2\beta(F - F^{FB})/[(1-\beta)u'(\bar{y})], \tag{B.5}$$

where $\zeta^{FB} = h - \bar{y}$, we have $c_\zeta \in \mathcal{C}(F)$, and so $\overline{F} > F^{FB}$.

By the definition of F^{FB},

$$W(h, c^{FB}) = (1-\beta)u(h) + \ \beta F^{FB},$$

and so

$$W(\ell, c^{FB}) = W(h, c^{FB}) > (1-\beta)u(\ell) + \beta F^{FB}. \tag{B.6}$$

Because marginal changes in ζ from ζ^{FB} result in only second losses to ex ante payoffs ($W^0(c_\zeta)$), we have

$$\frac{\partial W(h, c^{FB})}{\partial \zeta} = -(1-\beta)u'(\bar{y}),$$

and so

$$\begin{aligned} W(h, c_\zeta) =& W(h, c^{FB}) - (1-\beta)u'(\bar{y})(\zeta - \zeta^{FB}) + o((\zeta - \zeta^{FB})^2) \\ =& W(h, c^{FB}) + (\zeta^{FB} - \zeta)[(1-\beta)u'(\bar{y}) + o((\zeta - \zeta^{FB})^2)/(\zeta - \zeta^{FB})]. \end{aligned}$$

For $\zeta^{FB} - \zeta < \xi'$, where $\xi' > 0$ is a sufficiently small constant, the magnitude of the last term is less than $(1-\beta)u'(\bar{y})/2$, and so

$$W(h, c_\zeta) > W(h, c^{FB}) + (\zeta^{FB} - \zeta)(1-\beta)u'(\bar{y})/2.$$

For $F = F^{FB} + (\zeta^{FB} - \zeta)(1-\beta)u'(\bar{y})/(2\beta)$ (this is just a rewriting of (B.5)), we then have

$$W(h, c_\zeta) > (1-\beta)u(h) + \beta F.$$

Moreover, there is $\xi'' > 0$, such that for $\zeta^{FB} - \zeta < \xi''$, the strict inequality on the ℓ-incentive constraint (B.6) is preserved:

$$W(\ell, c_\zeta) > (1-\beta)u(\ell) + \beta F.$$

Setting

$$\xi := \min\{\xi', \xi''\}u'(\bar{y})/(2\beta)$$

completes the proof. □

Proof of Proposition 5.

1. Since, for ε small, the allocation in Example 1 is internally-Incentive compatible for $\pi = 1$ and provides partial insurance, $\mathcal{C}(F) \neq \varnothing$ for some $F > V^A$, and so $\overline{F} > V^A$. This also shows that $\mathbb{V}(V^A) > V^A$.

2. Suppose $(F_k) \nearrow \overline{F}$ is a sequence satisfying $\mathcal{C}(F_k) \neq \varnothing$. Since the space of consumption allocations is sequentially compact (being the countable product of sequentially-compact spaces), we can assume there is a convergent sequence $(c_k)_k$, with $c_k \in \mathcal{C}(F_k)$ and limit c_∞. Since all the constraints defining $\mathcal{C}$ are closed (and continuous in F), the limit also satisfies these constraints (including (20) at $F = \overline{F}$), and so $c_\infty \in \mathcal{C}(\overline{F})$, and $\mathcal{C}(\overline{F}) \neq \varnothing$.

3. The continuity of $\mathbb{V}$ follows from the continuity of $\mathcal{C}$ (Lemma B.12) and the maximum theorem.

4. This is Lemma B.13.

5. The function $p : [V^A, \overline{F}] \to [0, \overline{\pi}]$ defined by

$$p(F) := \frac{F - V^A}{\mathbb{V}(F) - V^A}$$

is strictly increasing, continuous, and onto (since $\mathbb{V}(V^A) > V^A$). It is straightforward to verify that for $\pi \in (0, \overline{\pi}]$, the fixed point is given by $F(\pi) := \pi\mathbb{V}(p^{-1}(\pi)) + (1-\pi)V^A$. The remaining claims are immediate.

6. Finally, for $\pi > \overline{\pi}$, the required F is strictly greater than $\overline{F}$, implying that the constraint set is empty, and so there is no fixed point.

□

B.4 Comparative Statics of the Limit Stationary Ladder

Lemma B.14 *Suppose a strong social norm exists but is not first-best insurance. The limit stationary ladder features decreasing risk sharing in π: $c_*(h)$ is an increasing function of π, while δ_* is a decreasing function of π.*

Proof. From Proposition 5.5, it is enough to show that $c_*(h)$ is an increasing function of F, while δ_* is a decreasing function of F.

From Lemma B.7, the floor on consumption c_ℓ is determined by

$$u(c_\ell) = u(\ell) + \beta(F - V^A).$$

Any specification of $c(h)$ and $\delta \in (0,1)$ (and c_ℓ) implies a unique allocation c (see the proof of Lemma B.10). In order to be feasible, c must satisfy

$$\sum_{k=0}^{\infty} \frac{1}{2^{k+1}} c(h\ell^k) = \bar{y}. \tag{B.7}$$

Since, for each $k \geq 1$, $c(h\ell^k)$ is a decreasing function of δ (when $c(h\ell^k) > c_\ell$), (B.7) describes a downward sloping continuous function in $c(h)$-δ space.

The F-incentive compatibility constraint at h,

$$W(h,c) = (1-\beta)u(h) + \beta F, \tag{B.8}$$

describes another downward sloping continuous function in $c(h)$-δ space.

The limit stationary ladder must satisfy (B.7) and (B.8), and so $(c_*(h), \delta_*)$ is a fixed point of (B.7) and (B.8). However, since the functions described by (B.7) and (B.8) are nonlinear, there may be multiple fixed points.

Let $(\hat{c}(h), \hat{\delta})$ denote the fixed point maximizing ex ante flow utility

$$\tfrac{1}{2}u(c(h)) + \sum_{k=1}^{\infty} \left(\tfrac{1}{2}\right)^{k+1} u(c(h\ell^k).$$

Since the value of ex ante flow utility is decreasing in $c(h)$ along (B.7), this is the fixed point with the smallest value of $c(h)$.

We claim that $\mathbb{c}(y^{t-1}h) = c_t(h) < \hat{c}(h)$ for all t. We prove the weak inequality, with the strict inequality following from the strict monotonicity of Lemma B.9.

Suppose, en route to a contradiction, that $c_t(h) > \hat{c}(h)$ for some t. Lemma B.9 implies that $c_{t+j}(h) > \hat{c}(h)$ and so $c_{t+j}(h\ell^k) \geq \hat{c}(h\ell^k)$ for all $j \geq 2$. Observe that c_{t+j} satisfies (B.7) for all $j \geq 2$, and so, since (23) also holds, the ex ante value of period $t+j$ utility of c_{t+j} is less than that of $(\hat{c}(h), \hat{\delta})$. Define a new allocation $c^\dagger$ as

$$c^\dagger(y^\tau) = \begin{cases} \mathbb{c}(y^\tau), & \tau \leq t+1, \\ \hat{c}(y^{[\tau-t-1]}), & \tau \geq t+2, \end{cases}$$

where $y^{[\tau-t-1]}$ is the last $\tau - t - 1$ periods of y^τ. That is, $c^\dagger$ agrees with $\mathbb{c}$ until period $t+2$, at which point the allocation is constant at $\hat{c}$. Observe that the new allocation satisfies F-incentive compatibility because $\hat{c}$ does, and consumptions after $h\ell^k$ are never decreased by the switch from period $t+2$. But $c^\dagger$ has higher ex ante welfare, a contradiction.

Since $\mathbb{c}(y^{t-1}y)$ is bounded away from every fixed point of (B.7) and (B.8) other than $(\hat{c}(h)\hat{\delta})$, the limit stationary ladder satisfies $(c_*(h), \delta_*) = (\hat{c}(h), \hat{\delta})$.

Since $F > F^{FB}$, $c_\ell < \bar{y}$, and so $(\bar{y}, 1)$ satisfies (B.7). Moreover, for the stationary ladder implied by $(\bar{y}, 1)$, F-incentive compatibility at h fails. This implies that at $(\hat{c}(h), \hat{\delta})$, the graph of (B.8) is steeper than then graph of (B.7).

If F increases, c_ℓ increases, implying the graph of the feasibility constraint (B.7) rotates around $(\bar{y}, 1)$ clockwise. At the same time, an increase in F moves the graph of (B.8) to the right. This implies that the new intersection with the lowest value of $c(h)$ has moved to the south east, that is, a higher $\hat{c}(h)$ and a lower $\hat{\delta}$. □

C Proofs for Section 7

Proof of Proposition 6. The outside option F only affects $\mathbb{V}^*(h; F)$ through c_ℓ (which is a strictly increasing function of F, and so makes the constraints strictly more demanding). Hence, $\mathbb{V}^*(h; F)$ is strictly decreasing function of F. It remains to prove that $\mathbb{V}^*(h; F) = W^F(h)$ at $\overline{F}$.

If $\mathbb{c}_*$ is the stationary ladder yielding $\mathbb{V}^*(h;F)$, define an allocation as follows

$$c^F(y^t) := \begin{cases} \mathbb{c}_*(h), & \text{if } y_t = h, \\ \mathbb{c}_*(h\ell^\tau), & \text{if } y^t = y^{t-\tau-1}h\ell^\tau, \\ \hat{c}(\ell^t), & \text{if } y^t = \ell^t, \end{cases} \tag{C.1}$$

where $\hat{c}(\ell^t)$ satisfies

$$\Pr(\ell^t)\hat{c}(\ell^t) = \overline{y} - \sum\nolimits_{y^t \neq \ell^t} \Pr(y^t)c^F(y^t).$$

By construction, c^F satisfies feasibility, and F-incentive compatibility for any history ending in a realization of ℓ (since $\mathbb{c}_*$ satisfies (28), $\hat{c}(\ell^t) \geq c_\ell$).

If $\mathbb{V}^*(h;F) \geq W^F(h)$, then the incentive constraint on $y^t h$ is satisfied under c^F for all y^t. Hence, $c^F \in \mathcal{C}(F)$, and so $F \leq \overline{F}$.

Suppose $\mathbb{V}^*(h;F) > W^F(h)$. A marginal increase in F preserves the inequality and so $F < \overline{F}$.

Finally, we prove that if $F \leq \overline{F}$, then $\mathbb{V}^*(h;F) \geq W^F(h)$. We do this by proving that if $\mathcal{C}(F)$ is nonempty, then there is a feasible *stationary* ladder of the form (C.1). We construct the stationary ladder by time averaging over histories that have the same number of y realizations after an h realization.

Suppose $\mathbb{c} \in \mathcal{C}(F)$ is optimal. From Lemma B.11, there exists $L \geq 2$ such that for all $\tau \geq L$, the ℓ incentive constraint holds with equality at any history of the form $y^t\ell^\tau$ and so, from Lemma B.7,

$$\mathbb{c}(y^t\ell^\tau) = c_\ell, \qquad \forall \tau \geq L. \tag{C.2}$$

For $M \geq 1$, define the ladder $(c_k^M)_{k=0}^L$ (recall that $\Pr(y^t) = 2^{-t}$):

$$\begin{aligned} c_k^M &= \begin{cases} \frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-k-1}}(\frac{1}{2})^{t-k-1}\mathbb{c}(y^{t-k-1}h\ell^k), & 0 \leq k < L \\ \frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-L}}(\frac{1}{2})^{t-L}\mathbb{c}(y^{t-L}\ell^L), & k = L \end{cases} \\ &= \begin{cases} \frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-k-1}}(\frac{1}{2})^{t-k-1}\mathbb{c}(y^{t-k-1}h\ell^k), & 0 \leq k < L \\ c_\ell, & k = L. \end{cases} \end{aligned}$$

We claim that $(c_k^M)_k$ satisfies (28) (where we set $c_k^M = c_\ell$ for $k > L$):

$$\begin{aligned}
\sum_{k=0}^{\infty}(\tfrac{1}{2})^{k+1}c_k^M &= \sum_{k=0}^{L-1}(\tfrac{1}{2})^{k+1}\frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-k-1}}(\tfrac{1}{2})^{t-k-1}\mathbb{c}(y^{t-k-1}h\ell^k) + (\tfrac{1}{2})^L c_\ell \\
&= \frac{1}{M+1}\sum_{t=L}^{L+M}\left\{\sum_{k=0}^{L-1}(\tfrac{1}{2})^{k+1}\sum_{y^{t-k-1}}(\tfrac{1}{2})^{t-k-1}\mathbb{c}(y^{t-k-1}h\ell^k) + (\tfrac{1}{2})^L c_\ell\right\} \\
&= \frac{1}{M+1}\sum_{t=L}^{L+M}\left\{\sum_{k=0}^{L-1}\sum_{y^{t-k-1}}(\tfrac{1}{2})^{t}\mathbb{c}(y^{t-k-1}h\ell^k) + (\tfrac{1}{2})^L c_\ell\right\} \\
&= \frac{1}{M+1}\sum_{t=L}^{L+M}\left\{\sum_{y^t \neq y^{t-L}\ell^L}\Pr(y^t)\mathbb{c}(y^t) + (\tfrac{1}{2})^L c_\ell\right\}.
\end{aligned}$$

But (C.2) implies

$$(\tfrac{1}{2})^L c_\ell = \sum_{y^t = y^{t-L}\ell^L}\Pr(y^t)\mathbb{c}(y^t)$$

and so

$$\sum_{k=0}^{\infty}(\tfrac{1}{2})^{k+1}c_k^M = \frac{1}{M+1}\sum_{t=L}^{L+M}E\mathbb{c}(y^t) \leq \overline{y}.$$

Since $\mathbb{c}(y^{t-1}\ell) \geq c_\ell$, it is immediate that c_k^M satisfies (27). Thus, for each M, $c^M \in \mathcal{C}_*(F)$.

Since $(c_k^M)_k \in [0,h]^L$, a closed and bounded set, the sequence $\left((c_k^M)_k\right)_M$ has a convergent subsequence with limit $(c_k^*)_k$. We now argue that $W(h,c^*) \geq W^F(h)$, completing the proof of the Proposition.

Since $\mathbb{c}$ is incentive compatible, for all y^t,

$$W^F(h) \leq W(y^t h, \mathbb{c}).$$

Consequently, taking time averages

$$\begin{aligned}
W^F(h) &\leq \frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-1}}(\tfrac{1}{2})^{t-1}W(y^{t-1}h,\mathbb{c}) \\
&= \frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-1}}(\tfrac{1}{2})^{t-1}(1-\beta)\sum_{k=0}^{\infty}(\tfrac{\beta}{2})^k u(\mathbb{c}(y^{t-1}h\ell^k)) + \tfrac{\beta}{2-\beta}W^F(h) \\
&= (1-\beta)\sum_{k=0}^{\infty}(\tfrac{\beta}{2})^k\frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-1}}(\tfrac{1}{2})^{t-1}u(\mathbb{c}(y^{t-1}h\ell^k)) + \tfrac{\beta}{2-\beta}W^F(h).
\end{aligned}$$

Since u is strictly concave,

$$\frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-1}}(\tfrac{1}{2})^{t-1}u(\mathbb{c}(y^{t-1}h\ell^k)) \leq u\left(\tfrac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-1}}(\tfrac{1}{2})^{t-1}\mathbb{c}(y^{t-1}h\ell^k)\right),$$

and so

$$W^F(h) \leq (1-\beta)\sum_{k=0}^{\infty}(\tfrac{\beta}{2})^k u\left(\frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-1}}(\tfrac{1}{2})^{t-1}\mathbb{c}(y^{t-1}h\ell^k)\right) + \tfrac{\beta}{2-\beta}W^F(h). \tag{C.3}$$

If the arguments of the utility function were c_k^M (which they are not), the proof would be done without the need to pass to the limit, since then the expression on the right hand side is simply $W(h, c^M)$.

However, we are almost done, since the discrepancy can be made arbitrarily small. For $k < L < M$, we have

$$\begin{aligned}
c_k^M &- \frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-1}}(\tfrac{1}{2})^{t-1}\mathbb{c}(y^{t-1}h\ell^k) \\
&= \frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-k-1}}(\tfrac{1}{2})^{t-k-1}\mathbb{c}(y^{t-k-1}h\ell^k) - \frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-1}}(\tfrac{1}{2})^{t-1}\mathbb{c}(y^{t-1}h\ell^k) \\
&= \frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-k-1}}(\tfrac{1}{2})^{t-k-1}\mathbb{c}(y^{t-k-1}h\ell^k) - \frac{1}{M+1}\sum_{t=L+k}^{L+M+k}\sum_{y^{t-k-1}}(\tfrac{1}{2})^{t-k-1}\mathbb{c}(y^{t-k-1}h\ell^k) \\
&= \frac{1}{M+1}\sum_{t=L}^{L+k-1}\sum_{y^{t-k-1}}(\tfrac{1}{2})^{t-k-1}\mathbb{c}(y^{t-k-1}h\ell^k) - \frac{1}{M+1}\sum_{t=L+M+1}^{L+M+k}\sum_{y^{t-k-1}}(\tfrac{1}{2})^{t-k-1}\mathbb{c}(y^{t-k-1}h\ell^k).
\end{aligned}$$

The magnitude of this expression is bounded above by $kh/(M+1)$. An identical argument shows that we have the bound of $Lh/(M+1)$ for the divergence of c_L^M.

Using (C.2), we can rewrite (C.3) as

$$\begin{aligned}
W^F(h) \leq (1-\beta)\sum_{k=0}^{L}(\tfrac{\beta}{2})^k u\left(\frac{1}{M+1}\sum_{t=L}^{L+M}\sum_{y^{t-1}}(\tfrac{1}{2})^{t-1}\mathbb{c}(y^{t-1}h\ell^k)\right) \\
+ \frac{2(1-\beta)}{(2-\beta)}\left(\frac{\beta}{2}\right)^{L+1}u(c_\ell) + \frac{\beta}{2-\beta}W^F(h).
\end{aligned} \tag{C.4}$$

For all $\varepsilon > 0$, there exists M_1^ε such that if $M > M_1^\varepsilon$, for all $k = 0, \ldots, L$ the upper bound of $Lh/(M+1)$ on consumption divergences is sufficiently small that the right side of (C.4)

is within ε of $W(h, c^M)$, implying $W^F(h) < W(h, c^M) + \varepsilon$. Moreover, there exists M_2^ε such that for all $M > M_2^\varepsilon$, $|W(h, c^*) - W(h, c^M)| < \varepsilon$. So, for $M > \max\{M_1^\varepsilon, M_2^\varepsilon\}$, we have

$$\begin{aligned} W^F(h) &< W(h, c^*) - W(h, c^*) + W(h, c^M) + \varepsilon \\ &< W(h, c^*) + 2\varepsilon. \end{aligned}$$

Since this holds for all $\varepsilon > 0$, we have

$$W^F(h) \leq W(h, c^*),$$

completing the proof. □

In the next lemma, an allocation is π-internally-incentive compatible if it satisfies the internal-incentive compatibility constraint (14) at the value π.

Lemma C.1 *Define the allocation $c_{\varepsilon,\alpha}$ by*

$$c_{\varepsilon,\alpha}(y^t) := \begin{cases} h - \varepsilon, & y_t = h, \\ \ell + \alpha\varepsilon, & y_{t-1} = h, y_t = \ell, \\ \ell + (2-\alpha)\varepsilon, & y_{t-1} = y_t = \ell, \\ \ell + \varepsilon, & t = 1, y_1 = \ell. \end{cases}$$

Define $\underline{\beta} := u'(h)/u'(\ell)$. For all $\pi > 0$, there exists $\eta > 0$, such that for all $\beta \in [\underline{\beta}, \underline{\beta} + \eta]$, all $\varepsilon' \in (0, \eta)$, and all $\alpha \in [1, 2]$, the allocation $c_{\varepsilon,\alpha}$ is not π-internally-incentive compatible.

Proof. We first calculate the values of the allocation $c_{\varepsilon,\alpha}$ after different histories, where we simplify notation by writing $c_h = h - \varepsilon$, $c'_\ell = \ell + \alpha\varepsilon$, and $c''_\ell = \ell + (2-\alpha)\varepsilon$:

$$\begin{aligned} W(h, c_{\varepsilon,\alpha}) &= (1-\beta)u(c_h) + \tfrac{\beta}{2}(W(h, c_{\varepsilon,\alpha}) + W(h\ell, c_{\varepsilon,\alpha})), \\ W(h\ell, c_{\varepsilon,\alpha}) &= (1-\beta)u(c'_\ell) + \tfrac{\beta}{2}(W(h, c_{\varepsilon,\alpha}) + W(\ell\ell, c_{\varepsilon,\alpha})), \\ W(\ell\ell, c_{\varepsilon,\alpha}) &= (1-\beta)u(c''_\ell) + \tfrac{\beta}{2}(W(h, c_{\varepsilon,\alpha}) + W(\ell\ell, c_{\varepsilon,\alpha})), \\ \text{and} \qquad W(\ell, c_{\varepsilon,\alpha}) &= (1-\beta)u(2\bar{y} - c_h) + \tfrac{\beta}{2}(W(h, c_{\varepsilon,\alpha}) + W(\ell\ell, c_{\varepsilon,\alpha})). \end{aligned}$$

Hence,

$$W(\ell\ell, c_{\varepsilon,\alpha}) = \frac{1}{2-\beta}\{2(1-\beta)u(c''_\ell) + \beta W(h, c_{\varepsilon,\alpha})\}$$

and so

$$W(h\ell, c_{\varepsilon,\alpha}) = (1-\beta)u(c'_\ell) + \frac{\beta}{2}\left\{W(h, c_{\varepsilon,\alpha}) + \frac{1}{2-\beta}\{2(1-\beta)u(c''_\ell) + \beta W(h, c_{\varepsilon,\alpha})\}\right\}$$
$$= (1-\beta)\left\{u(c'_\ell) + \frac{\beta}{2-\beta}u(c''_\ell)\right\} + \frac{\beta}{(2-\beta)}W(h, c_{\varepsilon,\alpha}).$$

Thus,

$$W(h, c_{\varepsilon,\alpha}) = (1-\beta)u(c_h) + \frac{\beta}{2}\left\{W(h, c_{\varepsilon,\alpha}) + (1-\beta)\left\{u(c'_\ell) + \frac{\beta}{2-\beta}u(c''_\ell)\right\} + \frac{\beta}{(2-\beta)}W(h, c_{\varepsilon,\alpha})\right\}$$
$$= (1-\beta)\left\{u(c_h) + \frac{\beta}{2}u(c'_\ell) + \frac{\beta^2}{2(2-\beta)}u(c''_\ell)\right\} + \frac{\beta}{(2-\beta)}W(h, c_{\varepsilon,\alpha}),$$

which implies

$$2(1-\beta)W(h, c_{\varepsilon,\alpha}) = (1-\beta)(2-\beta)\left\{u(c_h) + \frac{\beta}{2}u(c'_\ell) + \frac{\beta^2}{2(2-\beta)}u(c''_\ell)\right\},$$

that is,

$$W(h, c_{\varepsilon,\alpha}) = \frac{(2-\beta)}{2}u(c_h) + \frac{\beta(2-\beta)}{4}u(c'_\ell) + \frac{\beta^2}{4}u(c''_\ell). \tag{C.5}$$

A necessary condition for $c_{\varepsilon,\alpha}$ to be π-internally-Incentive compatible is

$$f(\varepsilon;\beta) := W(h, c_{\varepsilon,\alpha}) - (1-\beta)u(h) - \frac{\beta}{2}\overline{\pi}\left[W(h, c_{\varepsilon,\alpha}) + W(\ell, c_{\varepsilon,\alpha})\right] - \beta(1-\overline{\pi})V^A \geq 0.$$

Note that for all β, $f(0;\beta) = 0$. We now argue that there exists $\eta > 0$, such that for all $\beta \in [\underline{\beta}, \underline{\beta} + \eta]$ and all $\varepsilon' \in (0, \eta)$, $\partial f(\varepsilon';\beta)/\partial\varepsilon < 0$, implying

$$f(\varepsilon';\beta) < 0 \qquad \forall \beta \in [\underline{\beta}, \underline{\beta} + \eta],\ \varepsilon' \in (0, \eta).$$

Recalling our definition of $c_{\varepsilon,\alpha}$ and differentiating (C.5) with respect to ε,

$$\frac{\partial}{\partial\varepsilon}W(h, c_{\varepsilon,\alpha}) = \frac{1}{4}\left\{-2(2-\beta)u'(c_h) + \beta(2-\beta)\alpha u'(c'_\ell) + \beta^2(2-\alpha)u'(c''_\ell)\right\}.$$

Evaluating this expression at $\beta = \underline{\beta} = u'(h)/u'(\ell)$ and $\varepsilon = 0$ yields (for any $\alpha \in [1, 2]$)

$$\tfrac{1}{4}u'(\ell)\{-2(2-\underline{\beta})\underline{\beta} + \underline{\beta}(2-\underline{\beta})\alpha + \underline{\beta}^2(2-\alpha)\} \leq 0,$$

and so there exists η' such that for all $\beta \in [\underline{\beta}, \underline{\beta} + \eta']$ and all $\varepsilon' \in (0, \eta')$,

$$\frac{\partial}{\partial \varepsilon} W(h, c_{\varepsilon,\alpha}) < \frac{\pi \underline{\beta}}{(2 - \pi \underline{\beta})} \frac{1}{3} [u'(\ell) - u'(h)].$$

Turning to $W(\ell, c_{\varepsilon,\alpha})$, we have

$$\frac{\partial}{\partial \varepsilon} W(\ell, c_{\varepsilon,\alpha}) = (1 - \beta) u'(\ell + \varepsilon) + \frac{\beta}{2} \left\{ \frac{\partial}{\partial \varepsilon} W(h, c_{\varepsilon,\alpha}) + \frac{\partial}{\partial \varepsilon} W(\ell\ell, c_{\varepsilon,\alpha}) \right\}.$$

Note that $(1 - \beta) u'(\ell + \varepsilon) = u'(\ell) - u'(h)$ for $\beta = \underline{\beta}$ and $\varepsilon = 0$. Moreover, the term in $\{ \cdot \}$ is nonnegative at $\beta = \underline{\beta}$ and $\varepsilon = 0$. Thus, there exists η'' such that for all $\beta \in [\underline{\beta}, \underline{\beta} + \eta'']$ and all $\varepsilon' \in (0, \eta'')$,

$$\frac{\partial}{\partial \varepsilon} W(\ell, c_{\varepsilon,\alpha}) > \frac{2}{3} [u'(\ell) - u'(h)]. \tag{C.6}$$

Taking $\eta' = \min\{\eta', \eta''\}$, we thus have for all $\beta \in [\underline{\beta}, \underline{\beta} + \eta]$ and all $\varepsilon' \in (0, \eta)$,

$$\partial f(\varepsilon'; \beta) / \partial \varepsilon < 0.$$

This implies that for the specified bounds on β and ε, the allocation $c_{\varepsilon,\alpha}$ for *any* value of $\alpha \in [1, 2]$ is not π-internally Incentive compatible.

□

Proof of Corollary 2. We first argue that, for β larger than but near $\underline{\beta}$, the stationary ladder solving (29) for $F = \overline{F}$ has length 2 (which will allow us to use Lemma C.1): If the ladder is 3 or longer, then the consumption lower bound after realizations ℓ and $\ell\ell$ is not binding, and so

$$u'(\bar{c}_*(h)) = \beta u'(\bar{c}_*(h\ell)) = \beta^2 u'(\bar{c}_*(h\ell\ell)).$$

But for β close to $\underline{\beta}$, $c_*(h\ell)$ and $c_*(h\ell\ell)$ are both close to ℓ, and so $u'(\bar{c}_*(h\ell)) \approx u'(\bar{c}_*(h\ell\ell))$, implying β is close to 1, a contradiction.

Let $c^{\overline{F}}$ denote the allocation defined in (C.1) using the stationary ladder for $\overline{F}$. While we do not explicitly indicate the dependence of $\overline{F}$ and so $c^{\overline{F}}$ on β, both objects will vary with β: For β close to $\underline{\beta}$, the allocation $c^{\overline{F}}$ is given by $c_{\varepsilon,\alpha}$, the allocation defined in Lemma C.1, for an appropriate choice of ε and $\alpha \in [1, 2]$. Moreover, ε converges to 0 as β tends to $\underline{\beta}$.

For each $\pi > 0$, denote by $\eta(\pi) > 0$ the η-bound identified in Lemma C.1 (note that $\eta(\pi)$ is a nondecreasing function of π). There then exists $\eta'''(\pi) > 0$ such that for $\beta \in [\underline{\beta}, \underline{\beta} + \eta'''(\pi)]$, the ε associated with $c^{\overline{F}}$ is smaller than $\eta(\pi)$. This implies that for $\beta \in [\underline{\beta}, \underline{\beta} + \min\{\eta(\pi), \eta'''(\pi)\}]$, $c^{\overline{F}}$ cannot be π-internally-Incentive compatible.

We first prove that $\overline{\pi}(\beta) < 1$ for β close to $\underline{\beta}$. For if not, then for β close to $\underline{\beta}$,

$$\mathbb{V}(\overline{F}) \leq \overline{F}.$$

But this implies $c^{\overline{F}}$ is π-internally Incentive compatible for $\pi = 1$ (and so for any smaller π):

$$\begin{aligned} W(y, c^{\overline{F}}) &\geq (1-\beta)u(y) + \beta\overline{F} \\ &\geq (1-\beta)u(y) + \beta\mathbb{V}(\overline{F}) \\ &\geq (1-\beta)u(y) + \beta W^0(c^{\overline{F}}), \end{aligned}$$

which we have just seen is impossible for all $\beta \in [\underline{\beta}, \underline{\beta} + \min\{\eta(1), \eta'''(1)\}]$.

If

$$\mathbb{V}(\overline{F}) > \overline{F},$$

then $\overline{\pi}(\beta) < 1$, and we again have that the allocation $c^{\overline{F}}$ is $\overline{\pi}(\beta)$-internally-incentive efficient:

$$\begin{aligned} W(y, c^{\overline{F}}) &\geq (1-\beta)u(y) + \beta\overline{F} \\ &= (1-\beta)u(y) + \beta\{\overline{\pi}\mathbb{V}(\overline{F}) + (1-\overline{\pi})V^A\} \\ &\geq (1-\beta)u(y) + \beta\{\overline{\pi}W^0(c^{\overline{F}}) + (1-\overline{\pi})V^A\}. \end{aligned}$$

This completes the argument, since for any fixed $\pi > 0$, for β sufficiently close to $\underline{\beta}$, $c^{\overline{F}}$ is not π-internally-Incentive compatible. □

D Proofs for Section 8

Proof of Proposition 7. Note first that $c^{(T)} \in \mathcal{C}(\overline{F})$ for all $T \geq 0$. Denote by $T(\pi)$ the unique value of T satisfying

$$\pi[(1-\beta^T)V^A + \beta^T\mathbb{V}(\overline{F})] + (1-\pi)V^A < \overline{F} \leq \pi[(1-\beta^{T-1})V^A + \beta^{T-1}\mathbb{V}(\overline{F})] + (1-\pi)V^A.$$

Since $W^0(c^{(\alpha)})$ is continuous function of α, with

$$\pi W^0(c^{(0)}) + (1-\pi)V^A < \overline{F} \leq \pi W^0(c^{(1)}) + (1-\pi)V^A,$$

there exists $\alpha(\pi)$ such that

$$\pi W^0(c^{(\alpha(\pi))}) + (1-\pi)V^A = \overline{F}.$$

The convexity of $\mathcal{C}(\overline{F})$ implies $c^{\alpha(\pi)} \in \mathcal{C}(\overline{F})$.

Thus, $c^{\alpha(\pi)}$ is a social norm:

$$\begin{aligned} W(y^t, c^{\alpha(\pi)}) &\geq \alpha(\pi)W(y^t, c^{T(\pi)-1}) + (1-\alpha(\pi)W(y^t, c^{T(\pi)}) \\ &\geq (1-\beta)u(y_t) + \beta\overline{F} \\ &= (1-\beta)u(y_t) + \beta[\pi W^0(c^{(\alpha(\pi))}) + (1-\pi)V^A]. \end{aligned}$$

Finally, $c^{\alpha(\pi)}$ is constrained efficient, because no social norm can have higher value: Suppose c is a social norm with value $W^0(c)$. Then, $c \in \mathcal{C}((\pi W^0(c) + (1-\pi)V^A)$, and so $\mathcal{C}((\pi W^0(c) + (1-\pi)V^A) \neq \varnothing$, implying $\pi W^0(c) + (1-\pi)V^A \leq \overline{F}$. □

Proof of Proposition 8. Since $\bar{\mathbb{c}}(\ell^t) \geq c_\ell(\overline{F})$ for all t, $c^{[\alpha]} \in \mathcal{C}(\overline{F})$.

Since the payoff to any agent receiving the income h in the initial period is the same as under $\bar{\mathbb{c}}$ and the h $\overline{F}$-incentive compatibility constraint is always binding, the h payoff is given by

$$(1-\beta)u(h) + \beta\overline{F}.$$

The consumption $c_\ell(\overline{F})$ is determined by the requirement that the ℓ $\overline{F}$-incentive compatibility constraint is binding, and so the payoff to any agent receiving the income ℓ in the initial period under $c^{[0]}$ is

$$(1-\beta)u(\ell) + \beta\overline{F}.$$

This implies

$$W^0(c^{[0]}) < \overline{F},$$

so that

$$\pi W^0(c^{[0]}) + (1-\pi)V^A < \overline{F} < \pi W^0(c^{[1]}) + (1-\pi)V^A.$$

Thus, there exists $\alpha(\pi)$ such that

$$\pi W^0(c^{[\alpha(\pi)]}) + (1-\pi)V^A = \overline{F},$$

and so (applying the same argument as in the last paragraph of the proof of Proposition 7) $c^{[\alpha(\pi)]}$ is a constrained-efficient social norm for $\pi > \overline{\pi}$.

□

Supplementary Appendix: For Editorial Review Only

S.1 The Simple Model: Details and Derivations

In this appendix we provide the detailed derivations underlying Figure 1 in the main text. For concreteness we give a concrete parametric example (which we use to draw the figure), but its qualitative properties hold for any strictly concave utility function.

Let the period utility function is logarithmic, $u(c) = \log(c)$ and the endowment process by given by $h = 1+\epsilon$ and $\ell = 1-\epsilon$ so that expected (average) income in society is $E(y) = 1$, and ϵ measures the degree of income risk (the standard deviation of income). This is also the parameterization analyzed in Krueger and Perri (2006), which allows us to readily compare our results to theirs.

Direct calculations reveal that expected lifetime utility from the autarkic allocation, the first-best insurance allocation (consuming 1 in every period of life), and a stationary allocation with constant transfer $x \in [0, \epsilon]$ are given by

$$\begin{aligned} V^A &= V(x=0) = \log\left[\sqrt{1-\epsilon^2}\right] < 0, \\ V^{FB} &= V(x=\epsilon) = 0, \text{ and} \\ V(x) &= \frac{1}{2}\left[\log(1-\epsilon+x) + \log(1+\epsilon-x)\right] = \log\left[\sqrt{1-(\epsilon-x)^2}\right] \leq 0. \end{aligned} \tag{S.1}$$

For the parametric example the parameter thresholds for autarky and first-best insurance are available in closed form. Autarky is the only social norm for all discount factors $\beta \leq \underline{\beta} = 2u'(h)/[u'(\ell) + u'(h)] = 1 - \epsilon$, independent of trust π.

At the other extreme, the trust threshold π^{FB} for first-best insurance is given by

$$\pi^{FB} = 1 - \left(\frac{2(1-\beta)}{\beta}\right)\frac{[\log(1+\epsilon)]}{[-\log(1-\epsilon^2)]} < 1. \tag{S.2}$$

If, in addition, $\beta < \underline{\beta}^{FB} := 2\log(1+\epsilon)/[2\log(1+\epsilon) - \log(1-\epsilon^2)] < 1$, then first-best insurance is not a social norm for any trust $\pi \in [0,1]$. Note that $\underline{\beta} < \underline{\beta}^{FB}$, that is, there exist $\beta \in (\underline{\beta}, \underline{\beta}^{FB})$ such that the constrained efficient social norm is non-autarkic but also does not exhibit first-best insurance for any $\pi \in [0,1]$.

We can also analytically characterize the maximally attainable ex-ante lifetime utility (i.e., lifetime utility before being in a coalition) $\overline{F}$, the trust threshold at which this lifetime utility is attained $\overline{\pi}$ (and above which no fixed point to the operator $\mathcal{T}$ exists) and the

transfer $\bar{x}$ that implements this value. These are given by

$$\bar{x} = \epsilon + \beta - 1, \tag{S.3}$$

$$\overline{F} = \frac{1-\beta}{\beta}\log\left[\frac{2-\beta}{1+\epsilon}\right] + \frac{1}{2}\log\left[\beta(2-\beta)\right] \tag{S.4}$$

$$\overline{\pi} = 1 - \left(\frac{2(1-\beta)}{\beta}\right)\left(\frac{\log(1+\epsilon) - \log(2-\beta)}{\log(\beta(2-\beta)) - \log(1-\epsilon^2)}\right). \tag{S.5}$$

As long as $\beta > \underline{\beta} = 1 - \epsilon$ (otherwise autarky is the only social norm for all π) we have that $\bar{x} > 0$ and $\pi^{FB} < \overline{\pi} < 1$, since then both the numerator and denominator in the last fraction are positive. It is straightforward to see that $\overline{\pi}$ is strictly decreasing in income risk ϵ. The larger is income risk, the smaller is the set of trusts π for which a fixed point exists, and the larger is the set of trusts for which utility needs to be burned to dissuade the high-income agent from leaving the arrangement. Direct calculations also reveal that $\overline{F}$ is strictly decreasing in ϵ and strictly increasing in β.

Figure 1 in the main text plots the value $\Gamma(x)$ of a high-income agent from being in a coalition, and the value $\Psi(x;\pi)$ of a high-income agent from deviating from the coalition, against the stationary transfer x. For the parametric example these are given explicitly by

$$\begin{aligned}\Gamma(x) &= (1-\beta)\log(1+\epsilon-x) + \beta V(x) \text{ and} \\ \Psi(x;\pi) &= (1-\beta)\log(1+\epsilon) + \beta\left[\pi V(x) + (1-\pi)V^A\right] = (1-\beta)\log(1+\epsilon) + \beta F(x),\end{aligned}$$

where $V(x)$ was given in equation $(S.1)$ and

$$F(x) = \pi V(x) + (1-\pi)V^A.$$

Figure 1 in the main text was drawn for parameter values for which first-best insurance is never attainable ($\pi^{FB} < 0$), yet partial insurance is feasible $\beta > \underline{\beta}$. In this appendix, for concreteness we depict the two complementary cases. First, assume $\beta \leq \underline{\beta} = 1 - \epsilon$. Then the value of being a currently high-income agent in a coalition with no risk sharing, $\Gamma(x = 0) = (1-\beta)\log(1+\epsilon) + \beta V^A$, exceeds that of first-best insurance, V^{FB}, and is declining in x. Since the value of deviating at $x = 0$ is $\Psi(x = 0;\pi) = \Gamma(x = 0)$ is increasing in x, Figure S.1 shows that the only allocation satisfying the incentive constraint is autarky, and this is the only social norm (and a fixed point).

Now assume that $\beta > \underline{\beta} = 1 - \epsilon$, and thus some social norms with positive risk sharing exist. The main text displayed this scenario under the assumption that $\beta \in (\underline{\beta}, \underline{\beta}^{FB})$ and thus $\pi^{FB} < 0$, that is, constrained-efficient social norms always exhibit only partial insurance.

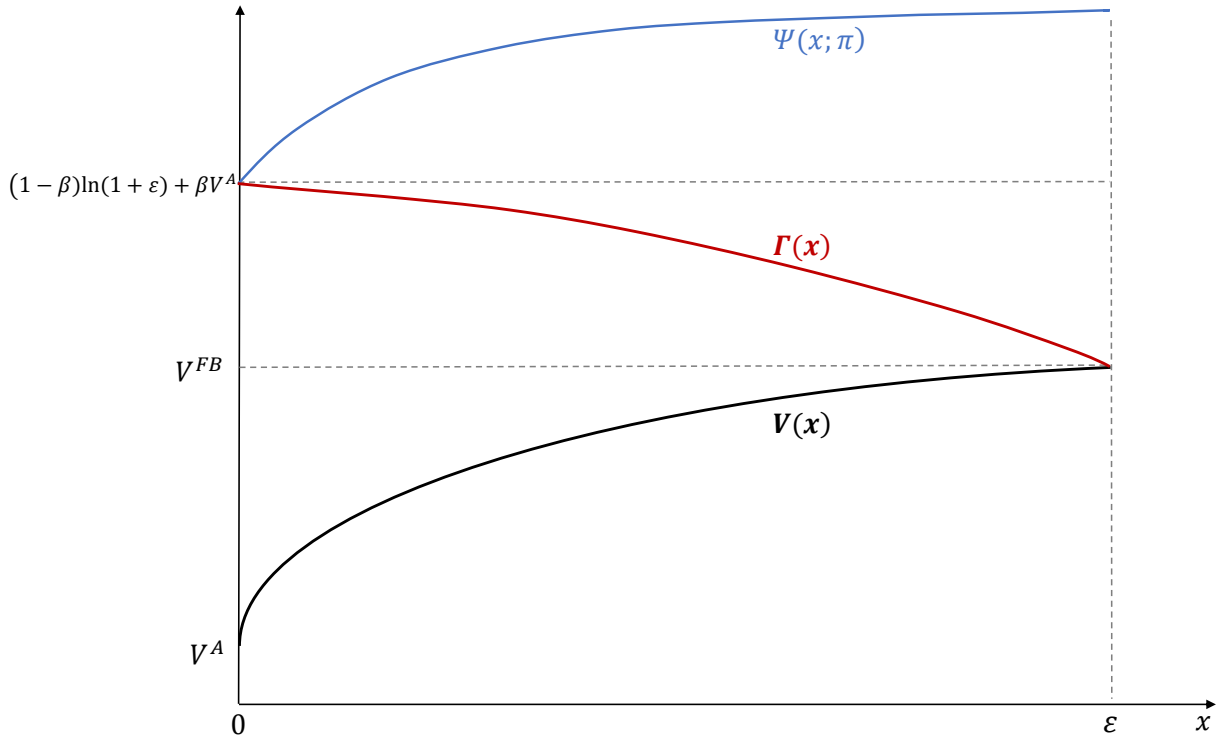

Figure S.1: The functions V, Γ, and Ψ when $\beta < \underline{\beta}$: a graphical representation of autarky.

Here we complement the analysis in the main text by displaying, in Figure S.2 the constrained-efficient stationary allocation when $\beta > \underline{\beta}^{FB}$, and thus $\pi^{FB} \geq 0$ in equation $(S.2)$. In this case, for sufficiently low levels of trust π the constrained-efficient social norm displays full consumption insurance.

Figure S.2 traces out how the unique constrained-efficient stationary allocation changes with trust π, as only the outside option $\Psi(x;\pi)$ changes with π. Specifically, this outside option tilts upward as π increases from 0 to 1, around the point $(x = 0, \Gamma(x = 0))$ for all $\pi \in [0, 1]$.

1. For all $\pi \in [0, \pi^{FB}]$ the full-insurance allocation $x(\pi) = \epsilon$ satisfies the incentive constraint since the $\Gamma(x)$-curve lies above the $\Psi(x;\pi)$-curve at $x = \epsilon$. The value of being in a coalition is maximal, at $V(\epsilon)$, and the ex-ante value of being in the unmatched pool, $F(\pi) = \pi V^{FB} + (1-\pi)V^A$ is strictly increasing in π, starting from V^A for $\pi = 0$.[1] Graphically, π^{FB} obtains when the $\Psi(x;\pi)$-curve has just tilted upward enough so that $\Gamma(\epsilon) = \Psi(\epsilon;\pi^{FB})$.

2. As π increases further to a $\pi \in (\pi^{FB}, \overline{\pi})$ and the $\Psi(x;\pi)$-curve tilts further upward, the incentive constraint becomes binding, and the constrained-efficient transfer $x(\pi)$ is

[1] For $\pi = 0$, there is first-best insurance inside a coalition, but since coalitions never form, $F(\pi = 0) = V^A$.

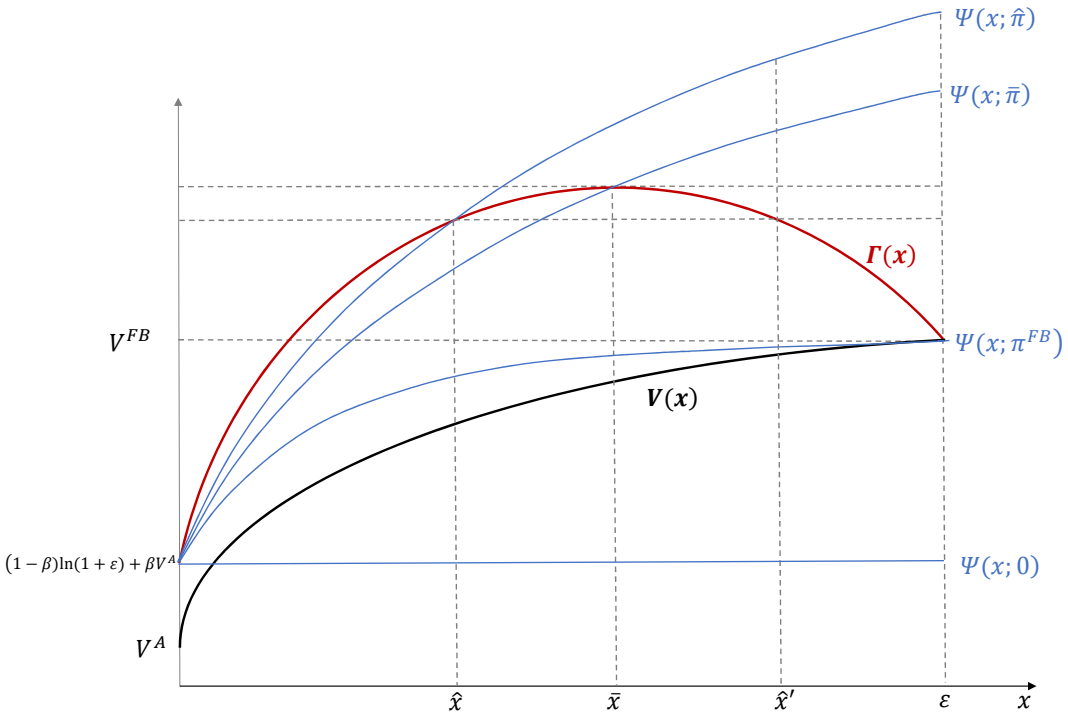

Figure S.2: The functions V, Γ, and Ψ when $\beta^{FB} < \beta$.

determined by the intersection between the $\Psi(x;\pi)$-curve and the $\Gamma(x)$-curve:

$$\left[1 - \frac{\beta(1+\pi)}{2}\right] \log\left(\frac{1+\epsilon}{1+[\epsilon - x(\pi)]}\right) = \frac{\beta}{2}(1-\pi)\log\left(\frac{1-[\epsilon - x(\pi)]}{1-\epsilon}\right) \tag{S.6}$$

The risk-sharing transfer $x(\pi)$ is strictly decreasing in π and the ex ante value $F(\pi) \equiv F(x(\pi))$ can be read off from the y-axis and is strictly increasing in π, even though expected utility of being matched in a coalition, $V(x(\pi))$ is strictly decreasing in π.

3. As π reaches $\overline{\pi}$ given in equation $(S.5)$ the intersection of the $\Psi(x;\overline{\pi})$-curve and the $\Gamma(x)$-curve occurs at $\bar{x} = \epsilon + \beta - 1$ and the maximally attainable ex ante utility of being unmatched is given by $\overline{F}$ given in equation (S.4).

4. Finally, as π increases further beyond $\overline{\pi}$, say to $\hat{\pi} \in (\overline{\pi}, 1]$, the $\Psi(x;\hat{\pi})$-curve and the $\Gamma(x)$-curve intersect, but at an $\hat{x} < \bar{x}$, with associated ex-ante value $\hat{F}$, and thus the allocation $\hat{x}$ satisfies the incentive constraint if the outside option is given by $\hat{F}$. However, a coalition faced with this outside option $\hat{F}$ will choose allocation $\hat{x}'$ with implied value $\hat{F}' > \hat{F}$, and thus $(\hat{x}, \hat{F})$ is not a fixed point for $\hat{\pi}$. For such high trust $\hat{\pi} > \overline{\pi}$, the full model analysis will show that the ex-ante value will remain at $\overline{F} < \pi V(\bar{x}) + (1-\pi)V^A$, and will be implemented by a non-stationary allocation

(which cannot be depicted in the figure) that "burns" utility relative to the better social norm $\bar{x}$ which, however, if implemented, would result in an outside option so high that the coalition itself is left with an empty set of resource- and incentive-compatible allocations.

S.2 Model Extensions

In this appendix we discuss two extensions of our model. In the first we consider a more general model of temporary delays to agreement after an initial failure to successfully form a coalition. In the second we extend out model to allow for production.

S.2.1 Temporary Delay

We have assumed that a deviating coalition succeeds with probability π and is in permanent autarky with complementary probability. We now assume that a failure to form a coalition is followed by $T \geq 1$ periods of autarky before another attempt can be made (so that if $T = 1$, a new attempt can be made in the next period after a failure). Under this assumption, after a deviation, coalition formation always eventually occurs. For fixed trust π a reduction of T increases the outside option. We now argue that the extension with a delay of T is equivalent to our original model with trust

$$\pi^{\dagger} := \frac{\pi}{1-(1-\pi)\beta^T}.$$

Suppose $\mathbb{c}^{\dagger}$ is an efficient allocation in the model with T-period delay. Then, the value of the outside option after deviating satisfies

$$W^d = \pi W^0(\mathbb{c}^{\dagger}) + (1-\pi)[(1-\beta^T)V^A + \beta^T W^d],$$

that is,

$$W^d = \pi^{\dagger} W^0(\mathbb{c}^{\dagger}) + (1-\pi^{\dagger})V^A.$$

It is easy to verify that since $\mathbb{c}^{\dagger}$ is an efficient allocation in the model with T-period delay, it must also be an efficient allocation in our original model for trust $\pi^{\dagger}$.

With finite exclusion, all agents are eventually in a risk-sharing arrangement, irrespective of the level of trust. However, the level of risk-sharing is declining in trust.

S.2.2 Risk Sharing and Production

We now briefly discuss how to extend our model to a production economy where output is produced and consumption is allocated within coalitions we will call production clubs, or firms for short. Output y_t produced by agent at time t depends upon idiosyncratic productivity $e_t \in E = \{e_\ell, e_h\}$ and labor effort l_t.

$$y_t = e_t l_t$$

Individual preferences are given by

$$(1-\beta)\mathbb{E}\Big\{\sum_{t=1}^{\infty}\beta^t U(c_t, l_t)\Big\},$$

and labor effort is bounded by the unit interval, so $l_t \in [0,1]$. All other aspects of the environment are the same as in the endowment economy studied thus far.

As before, risk-sharing incentives lead to continuum-sized firms being efficient, just as in our endowment economy. Since this implies that there is no aggregate output risk within a firm, an allocation within a continuum-sized firm are sequences of consumption and labor effort, both functions of the individual productivity history, $\{c_t(e^t), l_t(e^t)\}$.

In the special case in which labor is inelastically supplied at 1, and preferences are separable in consumption and labor, the efficient allocation of our model becomes essentially the same as in the endowment case, with endowment income $y \in \{\ell, h\}$ replaced by production income $y \in \{e_\ell \times 1, e_h \times 1\}$. This is the content of the next proposition.

Proposition S.1 *Suppose flow utility $U(c_t, l_t) = u(c_t) - v(l_t)$ is separable between consumption and labor, $(e_\ell, e_h) = (\ell, h)$, and $u'(y_h)y_\ell \geq v'(1)$.*

1. *There exists an efficient allocation with a consumption allocation that is identical to that in the endowment economy with $c(e^t) = c(y^t)$ and labor equal to $l_t(e^t) = 1$.*

2. *The payoff to forming a firm is the same as in the coalition payoff in the endowment economy, net of the cost of labor effort:*

$$(1-\beta)\mathbb{E}\Big\{\sum_{t=1}^{\infty}\beta^t[u(c_t(e^t)) - v(l_t(e^t))\Big\} = W^0(c) - v(1).$$

3. *The largest probability of successfully forming a firm for which there is a fixed point is still $\overline{\pi}$ from the endowment economy, however the associated highest feasible outside option is $\overline{F} - v(1)$.*

This proposition follows from the fact that the rankings of consumption sequences is unaffected by subtracting a constant labor cost in each period. For $\pi > \overline{\pi}$ utility-burning needs to occur in an efficient allocation, and while this can be done just as in the endowment case, richer possibilities involving the labor allocation emerge in the production economy.

The key to the previous proposition is that the within-firm consumption-labor allocation can be solved sequentially. In a first step the optimal labor allocation is determined, and in a second step the consumption risk-sharing allocation is chosen, taking as given the stochastic income process from the first stage. For a general utility function where labor is interior both consumption and labor are determined jointly.

An exception are utility functions without income effects on labor supply. For example, suppose households have Greenwood, Hercowitz, and Huffman (1988) preferences of the form

$$U(c,l) = \frac{1}{1-\gamma}\left\{c - \Psi\frac{l^{1+\theta}}{1+\theta}\right\}^{1-\gamma}$$

then the optimal labor allocation is determined by $l_t(e^t) = (e_t/\Psi)^{1/\theta}$ if Ψ is sufficiently large relative to e_t so that $l_t(e^t) < 1$. Now idiosyncratic income is given as $y(e^t) = \frac{e_t^{1+1/\theta}}{\Psi^{1/\theta}}$ and is efficiently shared within the firm as before, leading to a consumption allocation similar to the endowment economy. However, now we need to adjust the payoffs to take account of the differential labor utility costs. For example, the decay condition (23) in Proposition 3 becomes

$$\frac{\left(c(e^t) - (e_t/\Psi)^{(1+\theta)/\theta}\right)^{-\gamma}}{\left(c(e^{t+1}) - (e_{t+1}/\Psi)^{(1+\theta)/\theta)}\right)^{-\gamma}} = \delta_{t+1}.$$

Finally, it is easy to accommodate the notion that firms can realize increasing returns to scale, up to a point, in the size of its workforce, and that the production coalitions we model partially form not only for risk sharing purposes, but also for production efficiency purposes. Suppose that individual output within a firm is now given by

$$y_t = z e_t l_t$$

where $z = z(x)$ is a positive and weakly increasing function of the size x of the workers of the firm, with $z(x) = 1$ for $x \geq X$. That is, for firms larger than size $X < \infty$, which include those with an infinite number or a continuum of members, $z(x) = 1$. When $z(0) < 1$, then producing in autarky involves not only a loss in consumption smoothing but also a reduction in productivity. This again leads to a consumption allocation that has the same characteristics as in the endowment economy, but with a reduction in the value of autarky.

With period utility that is separable and CRRA in consumption the utility from autarky is scaled to $u(z(0))V^A(y)$.[2] Scaling down the utility from autarky raises π^{FB} and $\overline{\pi}$, the trust at which first-best insurance can be sustained and the threshold trust for which the fixed-point exists and utility burning is unnecessary. Thus, while the qualitative features of the analysis are unaffected by productivity benefits of large coalitions, quantitatively such production coalitions can provide better insurance when formed.

Our model of production clubs can qualitatively account for a number of well known features of the data. In the context of the literature on trust, Fukuyama (1995, p. 309, 312) asserts that while "there continues to be a steady proliferation of interest groups of all sorts in American life ... communities of shared values whose members are willing to subordinate their private interests for the sake of larger goals of the community ... have become rarer." This is consistent with the prediction of our model that more coalitions forming goes hand in hand with shallower cooperation within coalitions. On the issue of risk sharing within a firm, Guiso, Pistaferri, and Schivardi (2005) find that while temporary shocks are well-insured, permanent ones are not. This is consistent with our model, since a permanent shock to a worker's income would rescale their outside option and hence lead to a permanently different consumption ladder.[3]

[2]If the disutility of labor such that it is always efficient to supply a unit of labor in autarky for all levels of idiosyncratic productivity, then this simply shifts down the autarky payoff in the production economy relative to the endowment economy and is given by

$$(1-\beta)[u(z(0)y)-v(1)]+\beta\mathbb{E}_{y'}[u(z(0)y')-v(1)]=u(z(0))V^A(y)-v(1).$$

[3]With homothetic preferences, a permanent multiplicative shock to productivity for a (positive measure) subset of agents would simply scale these agents' consumption allocation by the permanent shock, since these agents with the positive shock can always secede and guarantee themselves the scaled consumption process.

S.3 Numerical Examples and Comparative Statics

In this section we provide numerical examples illustrating, in Subsection S.3.1, how the trust threshold for utility burning $\bar{\pi}$ changes with patience β away from $\underline{\beta}$, and, in Subsection S.3.2, how the dynamics of strong social norms evolves towards the limit stationary ladder. Subsection S.3.3 provides the details of the computational algorithm employed to derive these numerical results.

S.3.1 Insurance Possibilities and Trust Thresholds π^{FB} and $\bar{\pi}$

We first numerically characterize the threshold trust values for full insurance, π^{FB} and for the existence of strong social norms (and thus for the absence of utility burning), $\bar{\pi}$. These calculations complement the third part of Proposition 4 in the main text where we proved that near $\underline{\beta}$, the threshold $\bar{\pi}$ is less than one and thus utility burning must occur for sufficiently high values of trust.

Figure S.3 plots the threshold values for π^{FB} and $\bar{\pi}$ against the discount factor β, for an income process with $h = 1.25$ and $\ell = 0.75$, and logarithmic period utility, $u(c) = \log(c)$. This parameterization implies that $\underline{\beta} = u'(h)/u'(\ell) = 0.75/1.25 = 0.6$.

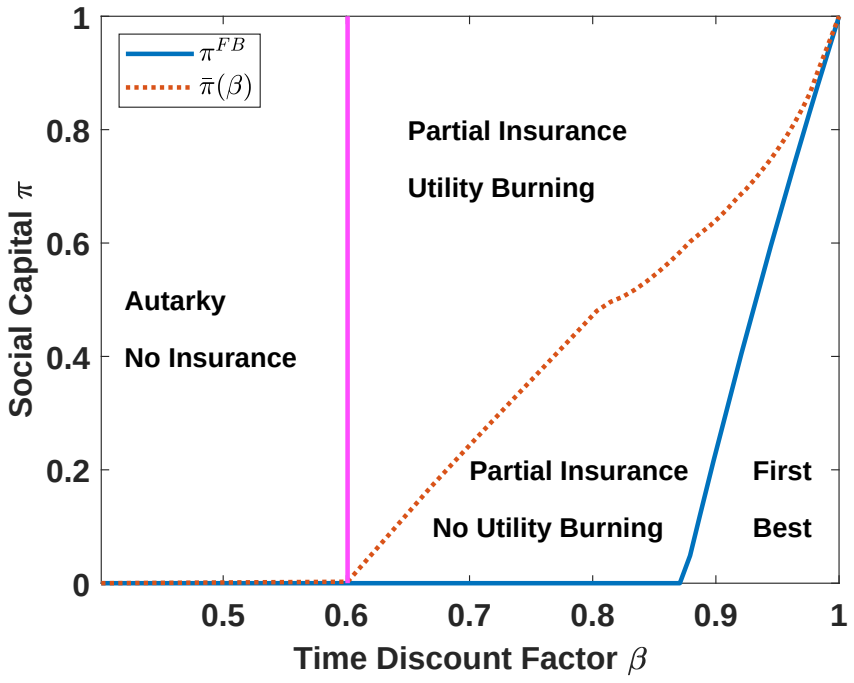

Figure S.3: Insurance possibilities as a function of (β, π), for $h = 1.25$, $\ell = 0.75$, $u = \log$.

The figure demonstrates that for discount factors $\beta \leq \underline{\beta}$ the constrained-efficient allocation is autarkic independent of trust π. For values of $\beta > \underline{\beta}$, in contrast, the constrained-efficient allocation changes qualitatively as trust π increases. Take $\beta = 0.9$ for concreteness: for low values $\pi \leq \pi^{FB}$ first-best insurance can be sustained, for intermediate values $\pi \in (\pi^{FB}, \overline{\pi}]$ there is partial risk sharing but no utility burning, and for $\pi > \overline{\pi}$ the constrained-efficient social norm requires utility burning. Importantly, this numerical example shows that for all $\beta < 1$, the threshold trust level $\overline{\pi}(\beta)$ above which utility burning needs to occur as part of the constrained-efficient allocation is always less than one, a feature that we have robustly found through many parameterizations we have explored.

S.3.2 The Dynamics of Strong Social Norms

In this subsection we present results for an illustrative set of examples to convey the qualitative properties of strong social norms. Throughout this section we assume a CRRA period utility function. This functional form implies that equation (23) characterizing strong social norms can be written as

$$\forall y^t, \mathbb{c}(y^t \ell) > c_\ell(F) \implies \frac{\mathbb{c}(y^t)^{-\gamma}}{\mathbb{c}(y^t \ell)^{-\gamma}} = \delta_{t+1},$$

for some $\delta_{t+1} < 1$. Since $\delta_{t+1} < 1$, and defining $g_{t+1} := (\delta_{t+1})^{1/\gamma} < 1$,

$$\forall y^t, \mathbb{c}(y^t \ell) > c_\ell(F) \implies \mathbb{c}(y^t \ell) = g_{t+1} \mathbb{c}(y^t).$$

By Proposition 3 strong social norms have the form of a sequence of consumption ladders (as defined in Definition 8), where the period t-ladder is determined by an initial consumption after the high income $y = h$ realization, $\mathbb{c}_t(h)$, and then a decreasing sequence of lower consumptions $g_{t+1}\mathbb{c}_t(h), g_{t+1}g_{t+2}\mathbb{c}_t(h), \ldots$, until the lower bound $c_\ell(F)$ is reached (after $L-1$ realizations of ℓ). The strong social norm converges to a stationary ladder and associated constant decay rate $g_t = g_{t+1} = g$.

With these observations from our theoretical results in hand, the computation of strong social norms with associated outside option $F \in (V^A, \overline{F}]$ (and thus for trust π associated with that outside option) proceeds as follows.[4] The algorithm first computes a stationary consumption ladder and associated consumption decay rate g that satisfies the h-incentive-feasibility constraint associated with F with equality (as well as the resource constraint and the ℓ-incentive-feasibility constraint with equality for those at the very bottom of the

[4]The details of the computational procedure are described in Subsection S.3.3

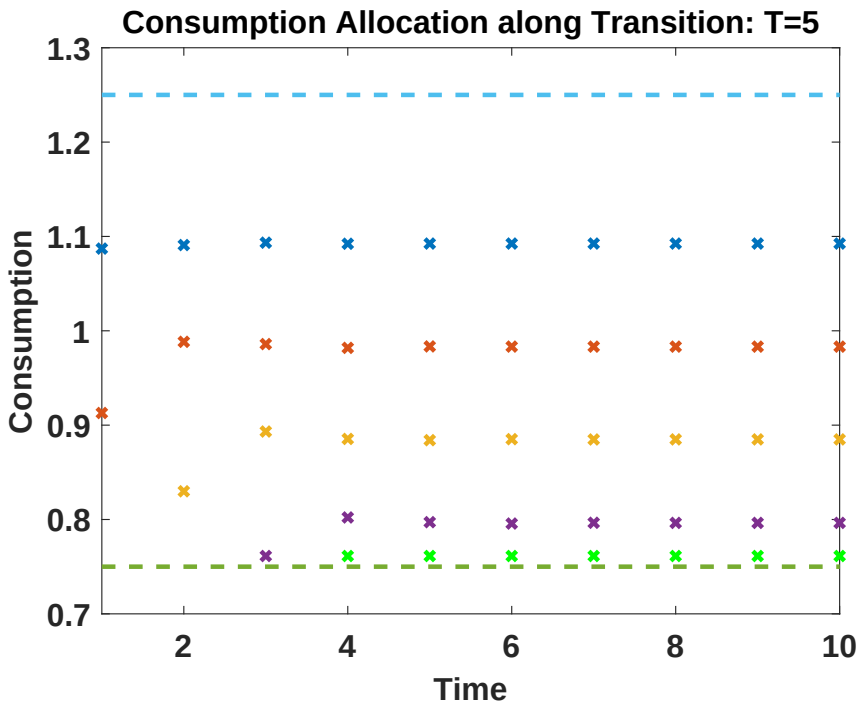

Figure S.4: Consumption allocation along transition, with $\ell = 0.75$ (indicated by lower dashed horizontal line), $h = 1.25$ (upper dashed horizontal line), $\beta = 0.9$, $\pi = 0.41$, and $\gamma = 1$.

ladder). It then determines the full dynamic strong social norm by imposing convergence to the stationary ladder in finite (but potentially long) time.

Figure S.4 plots the dynamics of the efficient consumption allocation with $u(c) = \log(c)$, incomes are $(\ell, h) = (0.75, 1.25)$, and the discount factor is chosen as $\beta = 0.9$. The level of trust is set to $\pi = \overline{\pi} = 0.41$ so that the value of the outside option is given by $F = \overline{F}$. Table 1 provides additional summary statistics for the allocation in this parameterization, as well as for alternative values of (β, γ) to display the comparative statics of the model with respect to its preference parameters (the values of $\overline{F}$ and $\overline{\pi}$ changes with (β, γ)).

From Figure S.4 we observe that as the transition unfolds, consumption spreads out over time, and eventually converges to the stationary ladder, which for this parameterization has five consumption steps. Consumption insurance worsens over time but remains positive: for high income agents the outside option is binding, but they consume substantially less than their income h (indicated by the upper dashed line) and thus provide insurance to low-income agents. Initially low income agents consume significantly more than their income (lower dashed line), and also more than implied by a binding outside option, $c_\ell(\overline{F})$. Over time those with continuously low income see their consumption drift down until the outside option binds and $c = c_\ell(\overline{F})$. This occurs in period four of the transition.

The strong social norm can generate high initial consumption insurance because the allocation does not inherit any implicit promises to past high income types. As time evolves, the consumption level of $c(\ell^t)$ declines as the burden of efficient smoothing of consumption to past high income types makes consumption scarcer. The allocation also becomes statically inefficient since agents with the same current income receive different consumption levels. Finally, the figure shows that although we do not force convergence to the stationary ladder until period 10 (the last period of the transition phase prior to imposing a stationary ladder) in this example, effectively allocations have converged to the stationary ladder by period four of the transition. Expanding the length of the transition yields utility gains that are indistinguishable from zero. Thus, although theoretically convergence to the stationary ladder is only asymptotic, our example suggest that numerically convergence occurs very rapidly; an observation shared by all examples we have computed.

	$\gamma = 1$		$\gamma = 2$	
Statistic	$\beta = 0.9$	$\beta = 0.95$	$\beta = 0.9$	$\beta = 0.95$
$V^{FB}/V(\overline{F})$ in %	0.63%	0.22%	0.45%	0.12%
$V(\overline{F})/\overline{F}$ in %	0.94%	0.71%	1.24%	0.80%
$\overline{\pi}$	0.41	0.66	0.69	0.85
$c_\ell(\overline{F})$	0.761	0.767	0.776	0.782
c_h	1.092	1.049	1.050	1.025
Steps	5	8	7	12
$\frac{EU(c_1)}{EU(c_\infty)}$ in %	0.28%	0.11%	0.25%	0.07%
$\frac{EU(c_2)}{EU(c_\infty)}$ in %	0.05%	0.05%	0.11%	0.03%
$Var(c_\infty)$	0.01	0.004	0.004	0.001
$\frac{Var(c_1)}{Var(c_\infty)}$	0.62	0.55	0.55	0.52
$\frac{Var(c_2)}{Var(c_\infty)}$	0.94	0.80	0.81	0.77

Table 1: Summary Statistics of the Transition

Notes: Ratios of (lifetime) utilities are converted into consumption equivalent variation and give the percentage increase in consumption (uniform across all states or histories) required to equalize period (or lifetime) utility across the two alternatives. The first two lines measure the welfare loss from imperfect consumption insurance relative to first-best insurance, and the welfare gain of coalition allocations relative to the outside option. The second panel provides summary statistics of the stationary ladder, and the third and forth panels show how expected utility and consumption insurance declines over time.

Table 1 contains summary statistics of strong social norms along the transition for alternative parameterizations of the model. Focus first on the benchmark case in the first column: we observe that the consumption allocation a coalition can implement improves significantly (worth 0.94% of consumption) on the outside option, by providing insurance to initially poor agents, but also needs to leave significant insurance opportunities unexploited (worth 0.63%

of consumption relative to first-best insurance). Insurance gets worse over time as expected period utility falls and consumption dispersion rises over time.[5] As households become more patient (higher β) and more risk-averse (higher γ), the strong social norm gets closer to first-best insurance, but the gains from coalition risk sharing relative to the outside option become smaller. The stationary ladder has more steps and the support of the consumption distribution tightens. We also observe that increased patience (higher β), elevates the gains of coalition risk sharing (compared to the outside option) mostly through an improvement of the stationary ladder. An increase in risk aversion (larger γ), in contrast, leads to better risk sharing both because of an improved stationary ladder and longer initial insurance and thus slower convergence to the ladder.

S.3.3 Computational Details for Section S.3

In this subsection we provide the details of how we compute strong social norms. Section S.3.3.1 describes how to compute a stationary ladder that delivers an outside option $F \in (V^A, \overline{F})$. Section S.3.3.2 describes how to determine the value of $\overline{F}$ (and the associated threshold trust value $\bar{\pi}$) together with the stationary ladder attaining it. Finally, Section S.3.3.3 describes the calculation of an entire dynamic efficient consumption allocation converging to a stationary ladder.

S.3.3.1 Stationary Ladder

For a *fixed* F, a stationary ladder $c_* = (c_*(h), gc_*(h), g^2c_*(h), \ldots, c_\ell)$ that satisfies feasibility and F-incentive compatibility for high income individuals (henceforth h-incentive compatibility) with equality as well as F-incentive compatibility for individuals at the bottom of the stationary ladder (henceforth ℓ-incentive compatibility) with equality is fully characterized by the upper and lower bound of consumption $(c_*(h), c_\ell)$, the decay rate g and the length of the ladder L. These values, all functions of a given $F \in (V^A, \overline{F})$, are calculated as follows:

1. Determine the unique consumption floor $c_\ell = c_\ell(F)$ from Lemma B.7, i.e.,

$$u(c_\ell(F)) = u(\ell) + \beta\left(F - V^A\right)$$

and recall the value of the outside option for the high income agents is

$$W^F(h) := (1-\beta)u(h) + \beta F.$$

[5] We only display the first two periods, relative to the stationary ladder.

2. A stationary ladder attaining F is then determined by three equations in the three unknowns $c_*(h), g, L$ from

$$L = \max\left\{k : g^{k-1}c_*(h) > c_\ell(F)\right\}, \tag{S.7}$$

$$\frac{1}{2}\sum_{t=0}^{L-1}\left(\frac{1}{2}\right)^t c_*(h)g^t + \left(\frac{1}{2}\right)^L c_\ell(F) = \bar{y}, \tag{S.8}$$

and, using $W(h, c_*) = W^F(h)$ in equation (26) characterizing lifetime utility from a stationary consumption ladder,

$$W^F(h) = \left(1 - \frac{\beta}{2}\right)\left[\sum_{k=0}^{L-1}\left(\frac{\beta}{2}\right)^k u(c_*(h)g^k)\right] + \left(\frac{\beta}{2}\right)^L u(c_\ell(F)). \tag{S.9}$$

This system of equations can be reduced to one non-linear equation in the unknown decay rate $g \in [\ell/h, 1]$. Use equation (S.7) to solve for the unique length of the ladder $L(g, c_*(h))$ given g, and then equation (S.8) to solve for the unique entry level consumption $c_*(h)$ (exploiting the fact that the period utility function is of CRRA variety), and insert both entities into equation (S.9) to obtain one equation in the unknown consumption decay rate g. The result of solving this one-dimensional nonlinear equation in g is a stationary ladder summarized by $(c_*(h)(F), g(F), L(F))$ as a function of the outside option F.

In general the stationary ladder associated with an outside option F need not be unique, although it is for $F = \overline{F}$, as we have seen in Section 7 of the main text. Computationally, since g must be bounded between $g = 1$ (no consumption decay, as in the full-insurance allocation) and $g = \ell/h$ (the consumption decay in the autarkic allocation), it is straightforward to determine all solutions to this one-dimensional nonlinear equation.

However, to better understand conceptually the potential multiplicity of stationary ladders and to determine which of the potentially several ladders is the relevant limit stationary ladder of the strong social norm (for the given F), it is instructive to proceed as follows. Instead of calculating the consumption decay rate g (and the associated $(c_*(h), L)$) as a function of F, in step 2 above we reverse the order and calculate, for a given stationary consumption ladder decay rate $g \in (\ell/h, 1)$, the outside option $F(g)$ attained by this g (and the associated consumption ladder) and plot F against $g \in [\ell/g, 1]$.

Numerically, we find that the mapping $F(\,\cdot\,)$ is hump-shaped with a maximum at $\bar{g} := \beta^{1/\gamma} < 1$ that delivers the maximum value $\overline{F}$.[6] That the decay rate at $F = \overline{F}$ is given by

[6]This result accords well with the results in the simple model where the outside option F from a stationary allocation was also hump-shaped in the extent of consumption insurance (which in the simple model was measured by the size of the transfer x), with the maximum $\overline{F}$ attained at $\bar{x}$.

$\bar{g} = \beta^{1/\gamma}$ can easily be shown theoretically and follows directly from the first-order condition of the program defining in $\overline{F}$ in (29) of the main text. The reason for the hump-shape of $F(\,\cdot\,)$ is as follows. Start at $g = 1$, and thus a constant consumption allocation with first-best insurance, and now lower g infinitesimally. Individuals with current income $y = h$ strictly prefer a more front loaded consumption allocation even though it entails more consumption risk in the future. As g initially falls from $g = 1$, both $W(h, c_*)$ and $c_*(h)$ increase, which in turn leads the outside option $F(g)$ to increase as g falls. At $g = \beta^{1/\gamma}$ the optimal front loading is attained from the perspective of the current h types; by reducing g further the associated increased future consumption risk more than offsets the higher current consumption $c_*(h)$ chosen to satisfy the resource constraint. Thus $W(h, c_*)$ and $F(g)$ decline as g falls beyond $g = \beta^{1/\gamma}$.

We cannot prove that $F(g)$ is hump-shaped in g but always found this to be the case in our numerical examples. This implies, in particular, that for any $F < \overline{F}$ there are two associated stationary ladders that deliver the same outside option F, one with little risk sharing ($g < \bar{g}$) and one with more risk sharing ($g > \bar{g}$). Since the algorithm for computing a dynamic strong social norm is based on the convergence of the allocation to a stationary ladder, it is important to know which ladder to pick, for a given $F < \overline{F}$.

In Lemma B.14 we have shown that although there might be multiple candidate stationary ladders, the one the strong social norm converges to asymptotically is the one with the smallest entry level of consumption $c_*(h)$ and thus the slowest consumption decay and the largest extent of risk sharing. Thus, for the purpose of the computation of dynamic strong social norms we restrict attention to stationary ladders with decay rates $g \in [\bar{g}, 1]$.

S.3.3.2 Determination of the Outside Option $\overline{F}$

To determine $\overline{F}$ we proceed as follows: At $F = \overline{F}$, Proposition 6 implies that there is a unique stationary ladder satisfying h-incentive compatibility and this ladder solves (29), so we know that the consumption decay rate is given by

$$g(\overline{F}) = \beta^{1/\gamma}.$$

In effect, $\overline{F}$ is the peak of the $F(\,\cdot\,)$ map discussed above, and is reached at $g = \bar{g}$. Since the value of $\overline{F}$ itself is unknown, we have to determine the lower consumption floor $c_\ell = c_\ell(\overline{F})$

jointly with $\overline{F}$, $c_*(h)$, and L. The relevant equations, with $g = g(\overline{F}) = \beta^{1/\gamma}$ are

$$u(c_\ell) = u(\ell) + \beta\left(\overline{F} - V^A\right), \tag{S.10}$$

$$\bar{y} = \frac{1}{2}\sum_{t=0}^{L-1}\left(\frac{1}{2}\right)^t c_*(h)g^t + \left(\frac{1}{2}\right)^L c_\ell, \tag{S.11}$$

$$L = \max\{k : g^{k-1}c_*(h) > c_\ell\}, \quad \text{and} \tag{S.12}$$

$$(1-\beta)u(h) + \beta\overline{F} = \left(1 - \frac{\beta}{2}\right)\left[\sum_{k=0}^{L-1}\left(\frac{\beta}{2}\right)^k u(c_*(h)g^k)\right] + \left(\frac{\beta}{2}\right)^L u(c_\ell). \tag{S.13}$$

The algorithm to determine $\overline{F}$ is then a slightly modified version of the procedure from the previous subsection, with $\overline{F}$ replacing g as the unknown to be computed, and are identical to the computations we carry out when solving for $F(g)$ for a given $g \neq \bar{g}$.

1. Guess $\overline{F} \in (V^A, V^{FB})$.
2. For a given $\overline{F}$:
   (a) Solve for c_ℓ from (S.10).
   (b) Jointly solve for $(c_*(h), L)$ from (S.11) and (S.12).
   (c) Calculate the right side of (S.13).
3. Solve $\overline{F}$ such that (S.13) holds.

Finally, once $\overline{F}$ is computed, we can determine $\bar{\pi}$ from equation (24).

S.3.3.3 Computation of the Transition

As discussed at the beginning of this section, the computational procedure solves for the strong social norm for a given F, imposing the stationary ladder from an exogenously specified period T. We now describe the computation of the allocations for fixed T and fixed outside option $F \in [V^A, \overline{F}]$. We take as given the stationary ladder associated with F, summarized by $(c_*(h)(F), g(F), L(F))$, including the lifetime continuation utilities $V_{i,*}(F)$ from being in step i of the stationary ladder, as described in the previous two subsections.[7] As described at the beginning of this section, the algorithm calculates consumption in three phases.

In the first $t \leq T$ periods the algorithm picks time-varying consumption of agents with currently high income (and so have binding incentive constraints), $(\mathbb{c}_t(h))_{t=1}^T$ and uses the

[7] The only part that distinguishes the calculations for $F < \overline{F}$ and $F = \overline{F}$ is the calculation of the stationary ladder(s), and in case of $F < \overline{F}$, the selection of the "right" ladder.

resource constraints and the fact that agents without binding constraints have common consumption decay rates (or consume the lower bound consume $c_\ell(F)$) to pin down the remainder of the consumption allocation. In a second phase, from $t = T+1, ..., T+L(F)$ the allocation blends into the stationary ladder: all agents with high income consume according to the stationary ladder, and all households with low income drift down from consumption in the previous period at a common (across individuals, but time-varying) decay rate g_t. Finally, for all $t > T + L(F)$, the allocation coincides with the stationary ladder. More precisely, the algorithm works as follows:

1. Guess $(\mathbb{c}_t(h))_{t=1}^T \in (\bar{y}, h)^T$.

2. Calculate the consumption allocation implied by this guess, imposing the characterization of a strong social norm from Proposition 3: the h-incentive-feasibility constraint holds with equality in every period, and all agents with low income either have non-binding constraints and their consumption decays at a common rate or they consume c_ℓ. The implied consumption allocations $(c_{i,t})_{i=0}^t$ for all $t = 1, \ldots, T, T+1, \ldots, T+L(F)$, are calculated as follows, where i again indicates the position on the consumption ladder:

   (a) Set

   $$\begin{aligned} & c_{0,t} = \mathbb{c}_t(h) \text{ for } t = 1, \ldots, T, \\ \text{and} \quad & c_{0,t} = c_*(h)(F) \text{ for } t = T+1, \ldots, T+L(F). \end{aligned}$$

   (b) For $t = 1$, determine $c_{1,1}$ from

   $$\frac{1}{2}\left[c_{0,1} + c_{1,1}\right] = \bar{y}.$$

   (c) For $t = 2, \ldots, T$, determine the consumption decay rates $(g_t)_{t=2}^T$ recursively (beginning with $t = 2$) as follows:

   The consumption decay g_t solves

   $$\frac{1}{2}\sum_{i=0}^{t-1}\left(\frac{1}{2}\right)^i c_{i,t} + \left(\frac{1}{2}\right)^t c_{t,t} = \bar{y},$$

   where for all $i = 1, \ldots, t$,

   $$c_{i,t} = \max\{g_t c_{i-1,t-1}, c_\ell(F)\}.$$

For each t, g_t is determined by one equation. The equations are solved forward in time since the allocations $\{c_{i,t}\}$ require knowledge of allocations $\{c_{i-1,t-1}\}$.

(d) For $t = T+1, \ldots, T+L(F)$, part of the consumption allocations are on the stationary ladder. For each $t = T+1, \ldots, T+L(F)$, the consumption decay g_t solves

$$\frac{1}{2}\sum_{i=0}^{t-1}\left(\frac{1}{2}\right)^i c_{i,t} + \left(\frac{1}{2}\right)^t c_{t,t} = \bar{y},$$

where

$$c_{i,t} = \begin{cases} g^i c_h(F), & \text{for } i = 1, \ldots, t-T-1, \\ \max\{g_t c_{i-1,t-1}, c_\ell(F)\}, & \text{for } i = t-T, \ldots, t. \end{cases}$$

3. For a given guess $(\mathbb{c}_t(h))_{t=1}^T$, the previous step delivers the entire allocation $(c_{i,t})_{i=0}^t$ for periods $t = 1, \ldots, T, T+1, \ldots, T+L(F)$. From date $t = T+L(F)+1$ on the consumption allocation coincides, by assumption, with the stationary ladder. Now we need to determine $(\mathbb{c}_t(h))_{t=1}^T$. These values must yield a consumption allocation that delivers the outside option $W^F(h)$ for all $t = 1, \ldots, T$. Construct the lifetime utility in period t after the history $y^{t-1-i}h\ell^i$, $V_{i,t}$, from the consumption allocation computed in the previous step. This can be done recursively, going backward in time. Lifetime utilities are given by, for each $t = T+L, \ldots, 1$ (working backwards in time) and all $i = 0, \ldots, t$,

$$V_{i,t} = (1-\beta)u(c_{i,t}) + \frac{\beta}{2}\left[V_{0,t+1} + V_{i+1,t+1}.\right]$$

Note that these calculations are the same before and in the blended phase, because $V_{0,t}$ is a function of $V_{i,t+i}$ for $i = 1, \ldots, L$, with $V_{L,T+L} = (1-\beta)u(\ell) + \beta F$ and $t \leq T+L$. The only role the consumption levels from the stationary ladder play is in step 2 above in determining $c_{i,t}$ via feasibility.

Finally we need to check whether the entry consumption levels $(\mathbb{c}_t(h))_{t=1}^T$ are such that the resulting consumption allocation hits the outside option for each $t = 1, \ldots, T$

$$V_{0,t} = (1-\beta)u(h) + \beta F.$$

If yes, we are done. If not, go back to step 1 and adjust the guess for $(\mathbb{c}_t(h))_{t=1}^T$.