Research Division
FEDERAL RESERVE BANK of MINNEAPOLIS

The Risk of Becoming Risk Averse:
A Model of Asset Pricing and Trade Volumes

Fernando Alvarez
University of Chicago and NBER

Andy Atkeson
University of California, Los Angeles,
NBER, and Federal Reserve Bank of Minneapolis

Staff Report 577
December 2018

DOI: https://doi.org/10.21034/sr.577
Keywords: Liquidity; Trade volume; Asset pricing; Tobin taxes
JEL classification: G12

The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Federal Reserve Bank of Minneapolis • 90 Hennepin Avenue • Minneapolis, MN 55480-0291
https://www.minneapolisfed.org/research/

The Risk of Becoming Risk Averse: A Model of Asset Pricing and Trade Volumes*

Fernando Alvarez† and Andy Atkeson‡

Abstract

We develop a new general equilibrium model of asset pricing and asset trading volume in which agents' motivations to trade arise due to uninsurable idiosyncratic shocks to agents' risk tolerance. In response to these shocks, agents trade to rebalance their portfolios between risky and riskless assets. We study a positive question — When does trade volume become a pricing factor? — and a normative question — What is the impact of Tobin taxes on asset trading on welfare? In our model, economies in which marketwide risk tolerance is negatively correlated with trade volume have a higher risk premium for aggregate risk. Likewise, for a given economy, we find that assets whose cash flows are concentrated on states with high trading volume have higher prices and lower risk premia. We then show that Tobin taxes on asset trade have a first-order negative impact on ex-ante welfare, i.e., a small subsidy to trade leads to an improvement in ex-ante welfare. Finally, we develop an alternative version of our model in which asset trade arises from uninsurable idiosyncratic shocks to agents' hedging needs rather than shocks to their risk tolerance. We show that our positive results regarding the relationship between trade volume and asset prices carry through. In contrast, the normative implications of this specification of our model for Tobin taxes or subsidies depend on the specification of agents' preferences and non-traded endowments.

*We are grateful for comments from Martin Eichenbaum, Sergio Rebelo, Ivan Werning, Harald Uhlig, Lars Hansen, Marco Bassetto, Monika Piazzesi, and participants at workshops in the University of Chicago, UCLA, Federal Reserve Bank of Chicago, EIEF, ECB, LACEA, and Bank of Portugal conference. We thank the Goldman Sachs Global Markets Institute for financial support. The previous of this paper circulated under the title "Random Risk Aversion and Liquidity: A Model of Asset Pricing and Trade Volumes".

†Department of Economics, University of Chicago; and NBER.

‡Department of Economics, University of California, Los Angeles; NBER; and Federal Reserve Bank of Minneapolis.

1 Introduction

In this paper, we develop a new general equilibrium model of asset pricing and asset trading volume in which investors' motivations to trade arise due to uninsurable idiosyncratic and aggregate shocks to investors' risk tolerance. In response to these shocks, investors trade to rebalance their portfolios between risky and riskless assets. The volume of asset trade in our model is driven by the dispersion of idiosyncratic shocks to risk tolerance. Our model delivers simple analytical expressions for asset prices and trading volume as functions of aggregate variables and the distribution of idiosyncratic shocks to agents' risk tolerance. We use these formulas to study a positive and a normative question: To what extent is trading volume a factor that helps price risky assets? And what are the welfare implications of Tobin taxes and subsidies to asset trading?

We show three main positive results regarding the relationship between trading volume and asset prices in our model. First, with positive trading volume, interest rates are lower than in an otherwise identical representative agent economy. Second, if aggregate shocks to risk tolerance are negatively correlated with trade volume, then the risk premium for aggregate risk is higher than in an otherwise identical representative agent economy with no rebalancing trade. Third, risky assets whose cash flows are concentrated on states in which trading volume is high sell at a higher price, i.e., they have lower expected excess returns. To help develop intuition for the asset pricing results in our model, we show that there is a mathematical correspondence between our asset pricing formulas and those in Mankiw (1986) and Constantinides and Duffie (1996) regarding the role of uninsurable idiosyncratic income shocks in asset pricing. The primary difference between our model and theirs, however, is that the shocks to risk tolerance in our model lead to a positive volume of rebalancing trade, while there is no trade in the equilibrium in these other papers.

We also use our model to evaluate the impact on ex-ante welfare of a Tobin tax on asset transactions. Contrary to the standard public finance result that in an undistorted equilibrium, a tax (or subsidy) has a zero first-order effect on welfare, in our case a Tobin tax on asset transactions has a first-order negative welfare effect. This is because it turns out that a transaction tax levied in the equilibrium with uninsurable risk tolerance shocks, through its incidence, exacerbates the equilibrium failure to share risk efficiently. In particular, in equilibrium, agents cannot insure against an idiosyncratic shock to their risk tolerance, which turns out to act the same way as an idiosyncratic uninsured wealth shock. Thus, a small subsidy to trade leads to first-order ex-ante welfare improvement because it improves upon the incomplete risk sharing achieved in equilibrium.

In the final section of the paper, we consider two alternative specifications of our model in which agents' desire to trade assets is driven by uninsurable idiosyncratic shocks to agents' non-tradable risky endowments of consumption goods rather than by shocks to agents' risk tolerance. We compare these alternative specifications of our model with the baseline specification with shocks to risk tolerance to highlight the economics of our positive and normative results.

The specification of preferences in our model is key for its positive and normative implications. We consider a three-period endowment economy, where, to simplify, consumption takes place only in the first and last periods. In period $t = 0$, all agents are identical. In period $t = 1$, all investors receive common signals about period $t = 2$ output, and each investor's preferences for consumption at $t = 2$ are realized. Specifically, we assume that in period $t = 1$, each investor has a utility function of the equicautious HARA family, which we index as $U_\tau(\cdot)$.[1] Formally, this is the class of utilities where risk tolerance is linear in consumption.[2] The intercept of the linear risk tolerance function, which we denote by τ, is allowed to be investor specific, and this is the preference shock that we consider.

What is central to our results is the way that investors view at time $t = 0$ the prospect of a time $t = 1$ random shock to their risk tolerance. Here we use a recursive representation of agents' preferences. For each realization of the risk tolerance parameter τ at $t = 1$, we define for each agent a time $t = 1$ level of certainty equivalent consumption based on that agent's realized risk tolerance τ and the (stochastic) consumption allocated to that agent at $t = 2$. Then each investor's time $t = 0$ preferences are given by an additively separable utility V over time $t = 0$ consumption plus the discounted value of the expected utility over the time $t = 1$ *certainty equivalent* of continuation consumption, also computed with the utility function V. This gives a non-expected utility as of time $t = 0$, as in Kreps and Porteus (1978) or Selden (1978). For the particular case where the distribution of risk tolerance τ is degenerate, ex-ante preferences are exactly as in Selden (1978). In the general case, time $t = 0$ investors evaluate the prospects of preference shocks *only* by considering their effect on their implied certainty equivalent consumption. In particular, we assume that investors are risk averse with respect to randomness in certainty equivalent consumption regardless of whether the variation

[1] As is well known, this family includes utility functions with constant relative risk aversion, constant absolute risk aversion, and quadratic utility, where the origin can be displaced from zero.

[2] Recall that risk tolerance is defined as the reciprocal of risk aversion.

in this certainty equivalent consumption comes from randomness in the time $t = 2$ allocation of consumption or from time $t = 1$ preference shocks. This formulation, as opposed to simply adding a shifter to standard additively separable preferences, isolates the effect of randomness of risk tolerance without having extra effects due to the particular cardinal representation of utility. Thus, our specification captures *the risk of becoming risk averse*. In addition, this specification has been used in social choice theory when considering foundations for ex-ante Rawlsian preferences "behind the veil of ignorance" to take into account the effect of different realized risk tolerances; see Grant et al. (2010) or Mongin and Pivato (2015).

The correspondence between an idiosyncratic shock to risk tolerance in our model and an idiosyncratic shock to income in Mankiw (1986) and related models can be understood as follows. The Arrow-Pratt theorem states that an investor with a risk tolerance lower than another, given the same budget set for tradable assets, also has lower certainty equivalent consumption. In our model, all investors are ex-ante identical at $t = 0$ and thus have the same asset position right before their time $t = 1$ preference shock is realized. Hence, they face the same budget set for tradable assets at $t = 1$. Therefore, the idiosyncratic risk tolerance shock is akin to a negative income shock in the sense that such a shock makes it more costly for that investor to attain any given level of certainty equivalent consumption through trade in assets at $t = 1$. Since the assumed time $t = 0$ preferences are in terms of expected utility over *certainty equivalent* consumption, the equivalence of a risk tolerance shock with an income shock is exact. Thus, as in Mankiw (1986) and Constantinides and Duffie (1996), the effect on time $t = 0$ marginal valuation of the idiosyncratic variation on *certainty* equivalent consumption depends on whether preferences feature precautionary savings. This claim is formally stated in Proposition 6. When $V''' > 0$, so investors are prudent, then states that correspond to high trade volume (i.e., high dispersion of certainty equivalence) are states with high marginal valuations. This effect explains the three positive asset pricing implications described above.

Our model can be formulated with a general specification of preferences $U_\tau(\cdot)$ realized at $t = 1$. Our assumption that the realized $U_\tau(\cdot)$ time $t = 1$ utility functions are of the equicautious HARA family has several important implications that make the model particularly tractable and which help us to understand the logic of our asset pricing results. First, the two-mutual fund separation theorem holds once agents' risk tolerances are realized at $t = 1$, and hence, for pricing assets at $t = 1$, it suffices for investors to trade in a Lucas tree and an uncontingent bond. No further claims on output in period $t = 2$ are needed. We refer to this type of trade as *portfolio rebalancing*. Second, these preferences admit Gorman aggregation. That is, for the purpose of pricing securities at $t = 1$ that pay consumption at $t = 2$ contingent on output realized at that date once agents' preference shocks have been realized, there is a representative investor whose preferences depend exclusively on the *average* of the risk tolerance parameter τ across investors. These results imply that asset prices realized at time $t = 1$, given agents' preference shocks and signals about aggregate output, are independent of time $t = 1$ trade volume. The third key implication that follows from our assumption of HARA preferences is that trade volume in the Lucas tree and the uncontingent bond at $t = 1$ maps directly into and depends exclusively on the realized *dispersion* of the risk tolerance parameter τ across investors.[3] A fourth key implication of HARA preferences is that the equilibrium allocation of certainty equivalent consumption to an agent with realized risk tolerance τ at $t = 1$ is linear and increasing in the size of that agent's purchases of the risky security. Hence, our model implies that data on the volume of rebalancing trade at $t = 1$ map directly into the dispersion of certainty equivalent consumption across agents at that date. This is summarized in Proposition 5.

These four properties of equicautious HARA preferences together imply that the only effect of trade volume on asset pricing comes from the marginal valuation that investors attach, from the point of view of time $t = 0$, to time $t = 1$ prices that occur given different realizations of the dispersion of idiosyncratic shocks to risk tolerance. In other words, the connection between trade volume and ex-ante asset prices comes from investors' valuation in the presence of risk to the *dispersion* of preference shocks that drive the desire for portfolio rebalancing. In our model, this risk is manifest in variation in the volume of rebalancing trade at $t = 1$ across the various states that may be realized at that date.

The main asset pricing positive implications of our model follow directly from the insights derived from these four properties of the equicautious HARA family of preferences. In particular, we obtain a decomposition of risk premia in Proposition 7, a comparative statics result on trade volume and interest rates in Proposition 8, and a comparative statics result on trade volume and risk premia across economies in Proposition 9 and across securities in a given economy in Proposition 10.

To gain intuition for our normative results regarding Tobin taxes, first observe that as discussed above, the initial undistorted equilibrium allocation has imperfect sharing of the

[3] Recall that investors are identical at time $t = 0$ and that τ shocks are uninsurable, so investors start their time $t = 1$ with identical portfolios.

idiosyncratic risk of shocks to risk tolerance τ. In particular, an agent who receives a low idiosyncratic realization of risk tolerance (high risk aversion) has low certainty equivalent consumption at $t = 1$ relative to an agent who receives a high risk tolerance shock. Thus, the welfare implications of a Tobin tax on asset trade whose proceeds are rebated lump sum to investors depend on the incidence of the tax: does the tax fall on risk-tolerant or risk-averse investors? We find fairly general sufficient conditions where the tax is borne by the risk-averse investors and hence the certainty equivalent consumption for these agents is pushed even lower from the imposition of the tax. These results are presented in Propositions 13 and 14. Our result that the tax falls primarily on the risk-averse investors follows from the classic result in finance that the elasticity of an investor's demand for risky assets is increasing in his or her risk tolerance and the classic result in public finance that tax incidence is determined by demand elasticities.

For most of the paper, we focus on our baseline model in which asset trade is driven by idiosyncratic shocks to risk tolerance. In the final section of the paper, we consider two alternative specifications of our model in which agents' desire to trade assets at $t = 1$ is driven by uninsurable idiosyncratic shocks to agents' tradable and non-tradable endowments of consumption at $t = 2$ rather than being driven exclusively by shocks to agents' risk tolerance. We compare these alternative specifications of our model with the baseline specification with shocks to risk tolerance to highlight the economics of our positive and normative results.

In the first alternative specification of our model, agents receive at time $t = 1$ a random amount of a non-tradable endowment of consumption at $t = 2$ where the risk in this endowment is diversifiable. The equilibrium of such a model is essentially the same as the original model, except that now there is more trade at time $t = 1$, since agents want to (and can) eliminate their exposure to this idiosyncratic shock. Thus, in this alternative specification of our model, we reach the same conclusions for the relationship between the volume of portfolio rebalancing trade and asset prices and, at the same time, allow for additional trade volume that is not portfolio rebalancing trade.

In the the second alternative specification of our model, agents receive at time $t = 1$ a random amount of non-tradable of consumption at time $t = 2$ where the risk in this endowment is exposed to the aggregate endowment of consumption at $t = 2$. Thus, relative to the first specification, the risk in this endowment at time $t = 1$ is non-diversifiable. In this case, for simplicity, we suppress the idiosyncratic random shocks to risk tolerance.[4] The idiosyncratic shock in this alternative specification is completely analogous to an idiosyncratic income shock in terms of certainty equivalent consumption at $t = 1$ because this shock affects the set of certainty equivalent consumption that the agent can afford at $t = 1$. This shock also motivates the agent to rebalance his portfolio at $t = 1$ to hedge the risk in his or her non-traded endowment to be realized at $t = 2$. The direction of this trade depends on the correlation of the agent's non-traded endowment at $t = 2$ and the payoffs of traded securities. Thus, the tight link in our baseline model between the observed rebalancing trade of an individual investor at $t = 1$ and that investor's certainty equivalent consumption is broken because this correlation could be positive or negative or zero. Nevertheless, we show that our three positive results regarding trade volume (in rebalancing trade) and asset pricing carry through directly to these alternative specifications. In contrast, our normative results regarding Tobin taxes hold for some specifications of preferences and non-traded endowment shocks, but not for others, because in this specification of our model, we no longer have a tight link for each agent between the level of certainty equivalent consumption for that agent in the initial undistorted equilibrium and that agent's realized elasticity of demand for aggregate risk.

The remainder of our paper is organized as follows. In subsection 1.1, we discuss the related literature. In section 2, we present the model with a general specification of preferences. We consider socially optimal allocations under the assumption that it is possible to condition agents' consumption at $t = 2$ on the realization of their preference shock at $t = 1$. We then define equilibrium with asset markets that are incomplete in the sense that traded claims are contingent only on aggregate shocks and not on the realizations of individuals' preference shocks. It is this form of equilibrium that we study. In section 3, we consider our model with preferences at $t = 1$, $U_\tau(\cdot)$ specialized to the HARA class. Here we develop the properties of these preferences that are key to making the model tractable. We then fully characterize equilibrium allocations, asset prices, and trade volumes. We also develop the mathematical correspondence between asset prices with shocks to risk tolerance and shocks to income. In section 4, we present our main results regarding the equilibrium relationship between trading volume and asset prices. In section 5, we present our normative results regarding Tobin taxes on asset trade. In section 6, we discuss the alternative specifications of our model in which asset trade is driven by hedging needs, and we compare the positive and normative implications of these alternative

[4] Thus, in this case one can consider exactly the same preferences as in Selden (1978), or even expected utility, if we so desire.

specifications with those of our baseline model. In the appendix, we complement our analysis of a simple linear tax of trade rebated lump sum, with the analysis of the optimal non-linear tax-subsidy.

1.1 Related Literature

There is a large theoretical and empirical literature on the relationship between trading volume and asset prices.

The idea that idiosyncratic preference shocks affect investors' precautionary demand for an asset (in this case, money) is central to Lucas (1980). Also, the idea that shocks to the demand side for risky assets are important is emphasized by Albuquerque et al. (2016). The model in that paper, as well as several other related models, incorporates risky preference shocks so that the model can account for the weak correlation of asset prices with traditional supply side factors emphasized in the literature. We concentrate on the relationship between aggregate and idiosyncratic preference shocks so we can examine the implied relationships between trade volume and asset pricing. Guiso, Sapienza, and Zingales (2018) provide evidence of changes in the risk aversion of individual Italian investors after the 2008 crisis.

Random Risk Tolerance. There is a small theoretical asset pricing literature that uses random changes in risk tolerance. An early example, particularly related because it addresses properties of the volume of transactions, is Campbell, Grossman, and Wang (1993). The aim of that paper is to investigate the temporal patterns in asset returns and trade volume. This paper considers shocks to risk tolerance in the context of a model with expected utility, so these shocks also correspond to shocks to agents' intertemporal elasticity of substitution. On the pure portfolio side Steffensen (2011), analyzes the implications randomness of risk tolerance, also using expected utility. Gordon and St-Amour (2004) use a time-separable utility with a state-dependent CRRA parameter to jointly fit consumption and asset pricing moments. In contrast to these earlier papers, here we consider an environment in which agents do not have expected utility over their preference shocks. With our preference specification, we are able to derive a more complete characterization of the positive and normative implications of our model.

The external habit formation model has, when one concentrates purely on the resulting stochastic discount factor, a form of random risk aversion that is nested by our equicautious HARA utility specification if agents have common CRRA preferences over consumption less the external habit parameter, as in Campbell and Cochrane (1999).[5] Bekaert, Engstrom, and Grenadier (2010) develop and estimate a version of Campbell and Cochrane (1999) where the ratio of consumption to habit also has independent random variation. They estimate a (linearized) version of the model and find a substantial role for independent shocks to the consumption/habit ratio, which have the interpretation of shocks to risk aversion. Guo, Wang, and Yang (2013) and Cho (2014) further investigate estimates of variations of this model.

In contrast to the papers cited above, our recursive definition of preferences isolates the shocks to risk tolerance, leaving intertemporal preferences over the allocation of certainty equivalent consumption unchanged.

Santos and Veronesi (2017) consider a model with external habits in which agents experience idiosyncratic shocks to risk tolerance because they each have different exposures to changes in the external habit parameter. As in our model, rebalancing trade occurs in the aggregate risky asset and riskless bonds due to heterogeneous changes in agents' external habit parameters correlated with aggregates. These authors focus on the dynamics of leverage and asset trade that result from this assumption, as opposed to the impact on ex-ante asset prices.

Kozak (2015) uses time-varying aversion in a representative agent model with non-separable preferences to model variations on the market price of risk. Kim (2014) uses Epstein-Zin preferences with a representative agent with time-varying risk aversion to develop non-parametric estimates of risk aversion and finds strong evidence for its variability. Drechsler (2013) and Bhandari, Borovička, and Ho (2016) use models where agents have time varying concerns for model misspecification, which can also be interpreted as random risk aversion. Drechsler (2013) studies time varying returns, especially of volatility-related derivatives.

Barro et al. (2017) consider a model with Epstein-Zin utility in which agents have idiosyncratic shocks to their risk tolerance. These shocks are introduced to ensure a stationary distribution of consumption across agents in the model. These shocks are implemented in such a manner to ensure that they do not have an impact on asset prices. Lenel (2017) also uses an Epstein-Zin model with random risk aversion. His interest is in the joint explanation of the holding of bonds and risky assets of different (ex-post) agent types and their returns.

Rebalancing Trade. In our model, the two-mutual-fund separation theorem holds, so agents trade only the market portfolio of risky assets (aggregate risk) and riskless bonds. Agents have no need to trade individual risky assets, nor do they need to trade more complex claims to

[5] The alert reader of Campbell and Cochrane (1999) will recognize the non-linear adjustment on that model to zero out the precautionary saving effect and obtain constant interest rates.

aggregate risk. We refer to trade in shares of the aggregate endowment and riskless bonds as "rebalancing trade." How much trade is there of this type? There is a large empirical literature on rebalancing trade. For instance, Lo and Wang (2000) and Lo and Wang (2006) use a factor analysis on the weekly trading volume of equities. They show that the detrended cross-sectional trade volume data have an important first component, which can be interpreted as rebalancing trade, accounting about two-thirds of the cross sectional variation. Yet, as they emphasize, this is far from being consistent with the two mutual fund separation theorem, and instead favors at least a second factor explaining trade. There are also many recent studies of individual household portfolios, which take advantage of large administrative data sets coming from tax authorities, such as Calvet, Campbell, and Sodini (2009). In that paper, the authors find strong evidence of idiosyncratic active rebalancing of portfolios between risky and riskless assets by Swedish households. In the final section of our paper, we consider a specification of our model with shocks to hedging needs that motivate trade that is not rebalancing trade, but instead is trade in individual risky securities subject to diversifiable risk. We show that the volume of this alternative type of asset trade does not affect ex-ante asset prices.

Shocks to hedging needs. Vayanos and Wang (2012) and Vayanos and Wang (2013) survey theoretical and empirical work on asset pricing and trading volume using a unified three-period model similar in structure to ours. In their model, agents are ex-ante identical in period $t = 0$, and they consume the payout from a risky asset in period $t = 2$. In period $t = 1$, agents receive non-traded endowments whose payoffs at $t = 2$ are heterogeneous in their correlation with the payoff from the risky asset. This heterogeneity motivates trade in the risky asset at $t = 1$ due to investors' heterogeneous desires to hedge the risk of their non-traded endowments. Vayanos and Wang focus their analysis on the impact of various frictions (participation costs, transactions costs, asymmetric information, imperfect competition, funding constraints, and search) on the model's implications for three empirical measures of the relationship between trading volume and asset pricing.[6] Our focus differs from theirs in that we study the impact of the shocks that drive demand for trade at $t = 1$ on asset prices in a model without frictions and then consider the welfare implications of adding a trading friction in the form of a transactions tax. Yet we have shown that our setup is amenable to studying frictions on trading, as we have done with

[6]The first of these measures is termed *lambda* and is the regression coefficient of the return on the risky asset between periods $t = 0$ and $t = 1$ on liquidity demanders' signed volume. The second of these measures is termed *price reversal*, defined as the negative of the autocorrelation of the risky asset return between periods $t = 1$ and $t = 1$ and between $t = 1$ and $t = 2$. The third measure is the ex-ante expected returns on the risky asset between periods $t = 0$ and $t = 1$.

our study of transaction taxes.

Duffie, Gârleanu, and Pedersen (2005) study the relationship between trading volume and asset prices in a search model in which trade is motivated by heterogeneous shocks to agents' marginal utility of holding an asset. As they discuss, these preference shocks can be motivated in terms of random hedging needs; see also Uslu (2015). Again, trade in their framework is subject to a friction not considered here.

2 The Model

In this section, we describe our model environment and our specification of agents' preferences with random shocks to each agent's risk tolerance. We define optimal and equilibrium allocations and develop our asset pricing formulas. In the next section, we solve the model for a specific class of preferences and characterize the model's implications for asset prices and trading volume due to portfolio rebalancing.

Consider a three-period economy with $t = 0, 1, 2$ and a continuum of measure one of agents. Agents are all identical at time $t = 0$. Agents consume in periods $t = 0$ and $t = 2$. Shocks to agents' risk tolerance are realized at $t = 1$.

There is an aggregate endowment of consumption available at $t = 0$ of $\bar{C}_0$. Agents face uncertainty over the aggregate endowment of consumption available at time $t = 2$, denoted by $y \in Y$. To simplify notation, we assume that Y is a finite set.

Agents face idiosyncratic and aggregate shocks to their preferences that are realized at $t = 1$. Heterogeneity in agents' preferences at time $t = 1$ motivates trade at $t = 1$ in claims to the aggregate endowment at $t = 2$. Preference types at $t = 1$ are indexed by τ with support $\tau \in \{\tau_1, \tau_2, \ldots, \tau_I\}$.

Uncertainty is described as follows. At time $t = 1$, an aggregate state $z \in Z$ is realized. Again, to simplify notation, we assume that Z is a finite set and the probabilities of z being realized at $t = 1$ are denoted by $\pi(z)$. The distribution of agents across types τ depends on the realized value of z, with $\mu(\tau|z)$ denoting the fraction of agents with realized type τ at $t = 1$ in state z. In describing agents' preferences below, we assume that the probability that an individual has realized type τ at $t = 1$ if state z is realized is also given by $\mu(\tau|z)$.

The conditional distribution of the aggregate endowment at $t = 2$ may also depend on z, with $\rho(y|z)$ denoting the probability of y being realized at $t = 2$ conditional on z being realized at $t = 1$. We denote the conditional mean and variance of the aggregate endowment at $t = 2$

by $\bar{y}(z)$ and $\sigma_y^2(z)$, respectively.

We summarize the timing of the realization of uncertainty agents face in our model as in Figure 1.

Allocations: An allocation in this environment is denoted by $\vec{c}\,(y;z) = \{C_0, c(\tau, y; z)\}$ where C_0 is the consumption of each agent at $t = 0$ and $c(\tau, y; z)$ is the consumption at $t = 2$ of an agent whose realized type is τ if aggregate states z and y are realized.

Feasibility requires $C_0 = \bar{C}_0$ at $t = 0$ and, at $t = 2$,

$$\sum_{\tau} \mu(\tau|z)\, c(\tau, y; z) = y \text{ for all } y \in Y \text{ and } z \in Z. \tag{1}$$

2.1 Preferences

We describe agents' preferences at $t = 0$ (before z and their individual types are realized) over allocations $\vec{c}\,(y;z)$ by the utility function

$$V(C_0) + \beta \sum_{z} \left[\sum_{\tau} \mu(\tau|z) V \left(U_{\tau}^{-1} \left(\sum_{y} \left[U_{\tau}(c(\tau, y; z))\rho(y|z) \right] \right) \right) \right] \pi(z), \tag{2}$$

where V is some concave utility function. We refer to U_{τ} as agents' type-dependent subutility function.

Certainty Equivalent Consumption: It is useful to consider this specification of preferences in two stages as follows. In the first stage, consider the allocation of certainty equivalent consumption at $t = 1$ over states of nature z. For any allocation $\vec{c}\,(y;z)$, an agent whose realized type is τ at $t = 1$ has certainty equivalent consumption implied by the allocation to his or her type and the remaining risk over y in state z given by

$$C_1(\tau; z) \equiv U_{\tau}^{-1} \left(\sum_{y} U_{\tau}(c(\tau, y; z))\rho(y|z) \right). \tag{3}$$

Given this definition, in the second stage, we can write agents' preferences as of time $t = 0$ in equation (2) as expected utility over certainty equivalent consumption

$$V(C_0) + \beta \sum_{z} \left[\sum_{\tau} \mu(\tau|z) V\left(C_1(\tau; z) \right) \right] \pi(z). \tag{4}$$

Convexity of Upper Contour Sets: To ensure that agents' indifference curves define convex upper contour sets, we must restrict the class of subutility functions $U_\tau(c)$ that we consider to those for which, given z, certainty equivalent consumption at time $t = 1$ as defined in equation (3) is a concave function of the underlying allocation $c(\tau, y; z)$ for each given τ and z at $t = 2$. We have the following propositions characterizing such subutility functions.

Proposition 1. *Fix z and τ. Certainty equivalent consumption*

$$\mathcal{C}_1(\tau, \vec{c}; z) \equiv U_\tau^{-1}\left(\sum_{y \in Y} U_\tau(c(\tau, y; z))\rho(y|z)\right) \tag{5}$$

is a concave function of the vector $\vec{c} = \{c(\tau, y; z)\}_{y \in Y}$ if and only if risk tolerance $\mathcal{R}_\tau(c) \equiv -U_\tau''(c)/U_\tau'(c)$ is a concave function of c.

This condition is satisfied for the equicautious HARA subutility function that we consider as our leading example throughout the paper, where $\mathcal{R}_\tau(c)$ is linear in c.

Feasible Allocations of Certainty Equivalent Consumption: To help in the interpretation of the asset pricing formulas below and in solving the model, it is useful to restate the feasibility constraint in equation (1) in terms of allocations of certainty equivalent consumption. Given a realization of z and the corresponding distribution of agent types $\mu(\tau|z)$, we say that an allocation of certainty equivalent consumption across individuals with risk tolerances τ at $t = 1$, $\{C_1(\tau; z)\}$, is feasible if there exists an allocation of consumption at $t = 2$, $c(\tau, y; z)$, that is feasible as in (1) and that delivers that vector of certainty equivalent consumption via (3).

Let $\mathbf{C}_1(z)$ denote the set of feasible allocations of certainty equivalent consumption at $t = 1$ given a realization of z. Note that this set is convex as long as agents have subutility functions $U_\tau(c)$ with convex upper contour sets. The set $\mathbf{C}_1(z)$ can be interpreted as a production possibility set whose shape is affected by the aggregate shock z which determines the distribution of tolerance for risk across agents through $\mu(\tau|z)$ and the quantity of risk to be borne through $\rho(y|z)$. As we discuss below, the marginal cost of producing certainty equivalent consumption computed from this production possibility set plays an important role in asset pricing.

We next consider optimal allocations and the corresponding decentralization of those allocations as equilibria with complete asset markets.

2.2 Optimal Allocations

Consider a social planning problem of choosing an allocation $\vec{c}(y;z)$ to maximize welfare (2) subject to the feasibility constraints (1). We refer to the solution to this problem as the *optimal allocation*. It will be useful to consider the solution of the social planning problem in two stages.

The first stage is to compute the set of feasible allocations of certainty equivalent consumption at $t = 1$ given z, denoted by $\mathbf{C}_1(z)$, and then solve the planning problem of choosing a feasible allocation of certainty equivalent consumption $\{C_0, C_1(\tau;z)\}$ to maximize (4) subject to those feasibility constraints. To characterize the sets $\mathbf{C}_1(z)$, we also consider efficient allocations as of $t = 1$ given z.

We say that an allocation $\vec{c}(y;z)$ is *conditionally efficient* if, given a realization of z at $t = 1$, it solves the problem of maximizing the objective

$$\sum_{\tau} \lambda_\tau \left[\sum_{y \in Y} U_\tau(c(\tau, y; z))\rho(y|z) \right] \mu(\tau|z) \tag{6}$$

given constraints (1) given some vector of non-negative Pareto weights $\{\lambda_\tau\}$, which can depend on z. The allocation of certainty equivalent consumption corresponding to a conditionally efficient allocation is then given by equation (3). The frontier of the set of feasible allocations of certainty equivalent consumption $\mathbf{C}_1(z)$ is found by solving this Pareto problem for all possible non-negative vectors of Pareto weights $\{\lambda_\tau\}$.

Clearly, the optimal allocation is also conditionally efficient.

The second fundamental welfare theorem applies to this economy under our assumptions on preferences. Thus, corresponding to the socially optimal allocation is a decentralization of that allocation as an equilibrium allocation with complete markets in which agents can trade claims to consumption at $t = 2$ contingent on realized values of τ, y, and z. In what follows, we consider equilibrium with incomplete asset markets.

2.3 Equilibrium with Incomplete Asset Markets

We now consider equilibrium in an economy in which agents are not able to trade contingent claims on the realization of their type τ at $t = 1$. Instead, they can only trade claims contingent on aggregate states z and y. We are motivated to consider incomplete asset markets here by the possibility that the idiosyncratic realization of agents' preference types is private information and that the opportunity for agents to retrade at $t = 1$ prevents the implementation of incentive compatible insurance contracts on agents' reports of their realized preference type τ.

We consider a decentralization with two rounds of trading, one at $t = 0$ before agents' types are realized and one at $t = 1$ after the realization of agents' types. We assume that all agents start at time $t = 0$ endowed with equal shares of the aggregate endowment $\bar{C}_0$ at $t = 0$ and realized y at $t = 2$. In a first stage of trading at time $t = 0$, we assume that agents can trade bonds whose payoffs are certain claims to consumption at time $t = 2$ conditional on aggregate state z being realized at time $t = 1$. Let a single unit of such a bond pay off one unit of consumption at $t = 2$ in all states y given that z is realized at $t = 1$. Let $Q(z)\pi(z)$ denote the price at $t = 0$ of such a bond. Note that trade in such bonds at $t = 0$ is equivalent to trade in sure claims to certainty equivalent consumption at $t = 1$ since these bonds are sure claims to consumption at $t = 2$.

Let $B(\tau, z)$ denote the quantity of such bonds held by an agent with realized type τ in his or her portfolio. Note that in equilibrium, agents choose their portfolio of bonds at $t = 0$ before their type is realized. Hence, we must have $B(\tau, z) = B(z)$ independent of τ. The bond market clearing condition is given by $B(z) = 0$ for all z.

In a second stage of trading at $t = 1$, agents can trade their shares of the aggregate endowment or realized y at $t = 2$ and the payoff from their portfolio of bonds in exchange for a complete set of claims to consumption contingent on the realized value of y at $t = 2$. Let $p(y; z)$ denote the price at $t = 1$, given that aggregate state z has been realized at that date, of a claim to consumption at $t = 2$ in the event that endowment y is realized. In what follows, we choose to normalize asset prices at time $t = 1$ in each state z such that the price of a bond, i.e., a claim to a single unit of consumption at $t = 2$ for every realization of y, is equal to one. That is, in each equilibrium conditional on z, we choose the numeraire

$$\sum_y p(y; z)\rho(y|z)dy = 1. \tag{7}$$

At $t = 1$, given state z, the price of a share of the aggregate endowment at $t = 2$ relative to that of a bond is given by

$$D_1(z) = \sum_y p(y; z)y\rho(y|z). \tag{8}$$

Since the price of a bond at this date and in this state is equal to one, $D_1(z)$ is also the level of this share price at $t = 1$ given state z.

We can price arbitrary claims to consumption at $t = 2$ with payoffs $d(y; z)$ contingent on the realized aggregate states z and y as follows. Let

$$P_1(z;d) = \sum_y p(y;z)d(y;z)\rho(y|z) \tag{9}$$

denote the price at $t = 1$ of a security with payoffs $d(y; z)$ in period $t = 2$ given that state z is realized. Then the price of this security at $t = 0$ is

$$P_0(d) = \sum_z Q(z)P_1(z;d)\pi(z), \tag{10}$$

where $Q(z)$ are the equilibrium bond prices at date $t = 0$.

Each agent's budget constraint at the first stage of trading (at $t = 0$) is given by

$$C_0 + \sum_z Q(z)B(z)\pi(z) = \bar{C}_0. \tag{11}$$

Agents' budget sets at $t = 1$ are contingent on the aggregate state z and are given by

$$\sum_y p(y;z)c(\tau,y;z)\rho(y|z) \leq D_1(z) + B(\tau,z). \tag{12}$$

The timing of trading and the notation for asset prices in our model is illustrated in Figure 2.

We first use this decentralization to define a concept of equilibrium at time $t = 1$ conditional on a realization of z. Here we assume that at time $t = 1$, agents are each endowed with one share of the aggregate endowment y at $t = 2$ and a quantity of bonds $B(\tau, z)$ (here allowed to vary with type τ) that are sure claims to consumption at $t = 2$. We require that, given z, the initial endowment of bonds satisfies the bond market clearing condition $\sum_\tau \mu(\tau|z)B(\tau, z) = 0$.

Conditional Equilibrium Given z Realized at $t = 1$: An *equilibrium conditional on* z and an allocation of bonds $\{B(\tau; z)\}$ is a collection of asset prices $\{p(y; z)\}$ and feasible allocation $\{c(\tau, y; z)\}$ that maximizes agents' certainty equivalent consumption (3) given the allocation of bonds and budget constraints (12).

Clearly, from the two welfare theorems, every conditional equilibrium allocation is conditionally efficient, and every conditionally efficient allocation is a conditional equilibrium allocation for some initial endowment of bonds.

We now present our definition of equilibrium.

Incomplete Markets Equilibrium: An *equilibrium with incomplete asset markets* in this economy is a collection of asset prices $\{Q^e(z), p^e(y; z)\}$ and a feasible allocation $\vec{c}^e(y; z)$ and bond holdings at $t = 0$ $\{B^e(z)\}$ that satisfy the bond market clearing condition and that together solve the problem of maximizing agents' ex-ante utility (4) subject to the budget constraints (11) and (12).

Note that since all agents are ex-ante identical, at date $t = 0$, they all hold identical bond portfolios $B^e(z) = 0$. This implies that we can solve for the equilibrium asset prices and quantities in two stages starting from $t = 1$ given a realization of z. Specifically, the equilibrium allocation of consumption at $t = 2$ conditional on z being realized at $t = 1$ is the conditional equilibrium allocation of consumption given z at $t = 1$ and initial bond holdings $B(\tau, z) = B^e(z) = 0$ for all τ and z, and the allocation of certainty equivalent consumption at $t = 1$ given z, $\{C_1^e(\tau; z)\}$, is that implied by the conditional equilibrium allocation of consumption at $t = 2$. Likewise, equilibrium asset prices at $t = 1$, $p^e(y; z)$, are the conditional equilibrium asset prices at $t = 1$ given z. We refer to this conditional equilibrium as the *equal wealth conditional equilibrium* because in it all agents have identical portfolios comprising one share of aggregate y and zero bonds.

2.4 Preference Shocks and Asset Prices

To gain intuition for how preference shocks affect asset pricing and to solve the model in the next section, it is useful to follow a two-stage procedure in solving for equilibrium.

In the first stage, we take as given the realized value of z at $t = 1$ and the payoffs from agents' date $t = 0$ bond portfolios and solve for the conditional equilibrium prices at $t = 1$, $p(y; z)$, for contingent claims to consumption at $t = 2$ and the corresponding conditional equilibrium allocation of consumption $c(\tau, y; z)$. These prices and this allocation satisfy the budget constraints (12) with $B(\tau; z)$ given, and the standard first-order conditions

$$\frac{U_\tau'(c(\tau, y_1; z))}{U_\tau'(c(\tau, y_2; z))} = \frac{p(y_1; z)}{p(y_2; z)} \tag{13}$$

characterizing conditional efficiency for all types τ and all y_1, y_2.

Given a solution for contingent equilibrium prices $p(y; z)$, we can define for each type of agent a cost function for attaining a given level of certainty equivalent consumption at time $t = 1$ given z as

$$H_\tau(C_1; z) = \min_{c(y;z)} \sum_y p(y; z)c(y; z)\rho(y|z) \tag{14}$$

subject to the constraint that $c(y;z)$ delivers certainty equivalent consumption C_1 at $t = 1$ for an agent of type τ. Using these cost functions, in the second stage, we can then compute the date $t = 0$ bond prices that decentralize the equilibrium allocation of certainty equivalent consumption as follows.

Consider the problem for the consumer of choosing certainty equivalent consumption and bond holdings to maximize utility (4) subject to budget constraints (11) and (12). These budget constraints can be restated as

$$H_\tau(C_1^e(\tau;z);z) = D_1^e(z) + B^e(z) \tag{15}$$

with $D_1^e(z)$ defined in (8) as the price of a share at $t = 1$ in state z. This problem has first order conditions

$$Q^e(z) = \beta \sum_\tau \left[\frac{V'(C_1^e(\tau;z))}{V'(C_0^e)} \Big/ \frac{\partial}{\partial C_1} H_\tau(C_1^e(\tau;z);z) \right] \mu(\tau|z) \tag{16}$$

with

$$\frac{\partial}{\partial C_1} H_\tau(C_1^e(\tau;z);z) = \frac{U'_\tau(C_1^e(\tau;z))}{\sum_y U'_\tau(c^e(\tau,y;z))\rho(y|z)}. \tag{17}$$

Note that this is the "standard" risk adjustment due to Kreps-Porteus non-expected utility, with the added feature of random risk tolerance. To see this, first consider the case where there is no dispersion in risk tolerance at z, so that $\tau = \bar\tau(z)$ for all agents, obtaining the standard risk adjustment:

$$Q^e(z) = \beta \frac{V'(C_1^e(\bar\tau(z);z))}{V'(C_0^e)} \frac{\sum_y U'_{\bar\tau(z)}(c^e(\bar\tau(z),y;z))\rho(y|z)}{U'_{\bar\tau(z)}(C_1^e(\bar\tau(z);z))}.$$

Moreover, if $\tau = \bar\tau(z)$ for all agents and $V(\cdot) = U_{\bar\tau(z)}(\cdot)$, we have expected utility, and thus

$$Q^e(z) = \beta \frac{\sum_y V'(c^e(\bar\tau(z),y;z))\rho(y|z)}{V'(C_0^e)}.$$

Note that if we interpret certainty equivalent consumption in our model as analogous to consumption in incomplete market models such as Mankiw (1986) and Constantinides and Duffie (1996), then this formula is the standard pricing formula for a sure claim to consumption in the presence of idiosyncratic risk to consumption.

3 Solving the Model with HARA Subutility

The specification of preferences we use to solve our model has subutility U_τ of the equicautious HARA utility class defined as

$$U_\tau(c) = \left(\frac{\gamma}{1-\gamma}\right)\left(\frac{c}{\gamma}+\tau\right)^{1-\gamma} \quad \gamma \neq 1 \text{ for } \left\{c : \tau + \frac{c}{\gamma} > 0\right\} \tag{18}$$

$$U_\tau(c) = \log(c+\tau) \text{ for } \{c : \tau + c > 0\} \text{ for } \gamma = 1 \text{ for } \{c : \tau + c > 0\}\,, \text{ and} \tag{19}$$

$$U_\tau(c) = -\tau \, \exp\left(-c/\tau\right) \text{ as } \gamma \to \infty,\ \text{for all } c\,. \tag{20}$$

This utility function is increasing and concave for any values of τ and γ as long as consumption belongs to the sets described above for each of the cases. To see this, we compute the first and second derivative as well as the risk tolerance function:

$$U'_\tau(c) = \left(\frac{c}{\gamma}+\tau\right)^{-\gamma} > 0\,,\ U''_\tau(c) = -\left(\frac{c}{\gamma}+\tau\right)^{-\gamma-1} < 0 \text{ and} \tag{21}$$

$$\mathcal{R}_\tau(c) \equiv -\frac{U'_\tau(c)}{U''_\tau(c)} = \frac{c}{\gamma} + \tau \tag{22}$$

Note that the notation above assumes that γ is common across agents. Note also that $\gamma > 0$ gives decreasing absolute risk aversion and $\gamma < 0$ gives increasing absolute risk aversion. The sign of γ will turn out to be immaterial for the qualitative implications of the model. Note as well that τ can be positive or negative. We do require, however, that $c/\gamma + \tau > 0$ for these preferences to be defined.

When agents have subutility U_τ of the equicautious HARA utility class, the interpretation of preference type τ is that if $\tau > \tau'$, then at any level of consumption, an agent of type τ has higher risk tolerance than an agent of type τ'. Hence, the heterogeneity we consider with these preferences is purely in terms of the level of risk tolerance across agents. The Arrow-Pratt theorem then immediately implies that if, given z at $t = 1$, agents of type τ and τ' receive the same allocation at $t = 2$, i.e., if given z, $c(\tau, y; z) = c(\tau', y; z)$ for all y, then agents of type τ have higher certainty equivalent consumption at $t = 1$, i.e., $C_1(\tau; z) \geq C_1(\tau'; z)$. In this sense, for an individual agent, having type τ' realized at $t = 1$ is a negative shock relative to having type τ realized at $t = 1$ in that with preferences of type τ', it requires more resources for the agent to attain the same level of certainty equivalent consumption as an agent with preferences of type τ.

Note that the equicautious HARA utility class nests several commonly used preference specifications in the literature. In particular, we have that as $\gamma \to \infty$, these preferences display risk tolerance that is constant in consumption and hence constant absolute risk aversion, or CARA preferences. With $\tau = 0$, these preferences display constant relative risk aversion, or

CRRA preferences. With $\tau \neq 0$, these preferences are equivalent to CRRA preferences with an additive external habit parameter.

When agents have subutility functions of the equicautious HARA class (18), then our model is particularly tractable, and it is possible to derive specific implications of the model for the relationship between asset prices and transactions volumes at $t = 1$. The tractability of our model follows from four related properties of these preferences that are derived from the observation that all agents have linear risk tolerance with a common slope in consumption (determined by γ). We prove each of these properties in the appendix.

These four properties are (1) Gorman aggregation, (2) linearity of the frontier of the set of feasible allocations of certainty equivalent consumption, (3) a two-fund theorem, and (4) type-independent marginal cost of certainty equivalent consumption. We present and prove each of these properties next.

Gorman Aggregation: Given a realization of z at $t = 1$, Gorman aggregation holds in all conditional equilibria. That is, in all conditional equilibria at $t = 1$, asset prices $p(y; z)$ are independent of the allocation of bonds $B(\tau; z)$ at that date and also independent of moments of the distribution of types $\mu(\tau|z)$ other than the mean of this distribution defined by

$$\bar{\tau}(z) \equiv \sum_{\tau} \tau \mu(\tau|z). \tag{23}$$

This result allows us to solve for equilibrium asset prices at $t = 1$, $p(y; z)$, directly from the parameters of the environment. Specifically, in all conditional equilibria, $p(y; z) = \bar{p}(y; z)$ where

$$\frac{U'_{\tau}(c(\tau, y_1; z))}{U'_{\tau}(c(\tau, y_2; z))} = \frac{U'_{\bar{\tau}(z)}(y_1)}{U'_{\bar{\tau}(z)}(y_2)} \equiv \frac{\bar{p}(y_1; z)}{\bar{p}(y_2; z)} \tag{24}$$

for all types τ and all y_1, y_2. The level of asset prices $\bar{p}(y; z)$ is set from the normalization in equation (7). Thus, asset prices in any conditional equilibrium correspond to those in an economy with a representative agent with risk tolerance $\bar{\tau}(z)$. We establish this result in Proposition 2.

Linear Frontier of Feasible Allocations of Certainty Equivalent Consumption Given subutility functions of the equicautious HARA class (18), given z realized at $t = 1$, the set of allocations of certainty equivalent consumption $\mathbf{C}_1(z)$ has a linear frontier, the optimal final consumption is affine, and the Lagrange multipliers on the resource constraints (1) of the

Pareto problem (6) defining conditional efficiency are independent of the Pareto weights for the problem. More formally, we have the following proposition:

Proposition 2. *Assume all $U_\tau(\cdot)$ are of the equicautious HARA class, and fix a particular z. An allocation is conditionally efficient if and only if:*

(i) There are scalars $\hat{\phi}(\tau; z) \geq 0$, with $\sum_\tau \hat{\phi}(\tau; z)\mu(\tau|z) = 1$, for which the allocation of consumption satisfies

$$\frac{c(\tau, y; z)}{\gamma} + \tau = \hat{\phi}(\tau; z)\left(\frac{y}{\gamma} + \bar{\tau}\right) \text{ for all } y. \tag{25}$$

(ii) The Lagrange multipliers on the resource constraints (1) of Pareto problem (6) are independent of the weights $\{\lambda_\tau\}$ and are directly proportional to

$$\hat{p}(y|z) = \frac{U'_{\bar{\tau}(z)}(y)\ \rho(y|z)}{\sum_{\tilde{y}} U'_{\bar{\tau}(z)}(\tilde{y})\ \rho(\tilde{y}|z)}. \tag{26}$$

(iii) The conditionally efficient allocation of certainty equivalent consumption $C_1(\tau; z)$ satisfies the pseudo-feasibility constraint

$$\sum_\tau \mu(\tau|z) C_1(\tau; z) = \bar{C}_1(z), \tag{27}$$

where

$$\bar{C}_1(z) \equiv U^{-1}_{\bar{\tau}(z)}\left(\sum_y U_{\bar{\tau}(z)}(y)\rho(y|z)\right) \tag{28}$$

is the certainty equivalent consumption of an agent with the average risk tolerance $\bar{\tau}(z)$ in the market who consumes the aggregate endowment y at $t = 2$.

Note that this characterization of the set of feasible allocations of certainty equivalent consumption $\mathbf{C}_1(z)$ implies that the fully optimal allocation of certainty equivalent consumption $C_1^*(\tau; z)$ solves the problem of maximizing welfare (4) subject to the pseudo-resource constraint (27). If the utility function over certainty equivalent consumption $V(C)$ is strictly concave, then the solution to this social planning problem is to have all agents receive the same certainty equivalent consumption at date $t = 1$, i.e., $C_1^*(\tau; z) = \bar{C}_1(z)$ for all τ. This allocation is implemented by uncontingent transfers at $t = 1$ from risk-tolerant to risk-averse agents. As we will see below, $C_1^*(\tau; z) = \bar{C}_1(z)$ is also the equilibrium allocation of certainty equivalent consumption in an economy in which there is no dispersion in shocks to risk tolerance.

Since any equilibrium allocation is conditionally efficient, the first and second welfare theorem apply given realized identities τ at $t = 1$. Hence, Gorman aggregation follows since we can take the Lagrange multipliers $\hat{p}(y; z)$ to be the equilibrium prices $p(y; z)$ independent of the Pareto weights across agents at $t = 1$. Recall Gorman aggregation implies that the asset prices at $t = 1$ that decentralize the fully optimal allocation are the same as those in the equilibrium with incomplete asset markets.

We now turn to the characterization of portfolios in the equilibrium with incomplete market, and then return, as a consequence, to the determination of the allocation of certainty equivalent consumption in an equilibrium with incomplete markets.

Two-Mutual Fund Separation Theorem and Equilibrium Certainty Equivalent Consumption Given subutility functions of the equicautious HARA class (18), we obtain a two fund separation theorem for all conditional equilibria. That is, to decentralize any conditionally efficient allocation at $t = 1$ given z, it is sufficient to have agents trade only shares of the aggregate endowment y at $t = 2$ and a riskless bond. We show that each agent's equilibrium purchase of shares of aggregate risk is linear in the difference between that agent's realized risk tolerance and the average risk tolerance in the market. This, in turn, implies that the certainty equivalent consumption allocated to that agent is also linear in the difference between that agent's realized risk tolerance and the average risk tolerance in the market. These two observations allow us to establish a direct relationship between trade volume in shares of aggregate risk and the idiosyncratic risk to certainty equivalent consumption at $t = 1$ that agents face as of $t = 0$.

It is convenient to define the *representative agent absolute risk aversion* for each realization of z at $t = 1$ as the absolute risk aversion of an agent with budget feasible risk-free consumption $\bar{B}_1(z)$ and risk tolerance equal to the average risk tolerance in the market $\tau = \bar{\tau}(z)$ as

$$\bar{\mathcal{A}}(z) \equiv -\frac{U''_{\bar{\tau}(z)}\left(\bar{D}_1(z)\right)}{U''_{\bar{\tau}(z)}\left(\bar{D}_1(z)\right)} = \frac{1}{\frac{\bar{D}_1(z)}{\gamma} + \bar{\tau}(z)}\,. \tag{29}$$

Recall that $\bar{D}_1(z)$ is the price of a share of the aggregate endowment at $t = 1$ in state z defined in (8) using equilibrium asset prices $\bar{p}(y; z)$.

We then have the following proposition:

Proposition 3. *Let $\phi^e(\tau; z)$ denote the post-trade quantity of shares of the aggregate endowment held by an agent of type τ at $t = 1$ given realized z, and let $C_1^e(\tau; z)$ denote the certainty* *equivalent consumption allocated to that agent in the equilibrium with incomplete markets. Then*

(i) To implement the incomplete markets equilibrium allocation, the quantity of shares purchased at $t = 1$ by this agent is

$$\phi^e(\tau; z) - 1 = (\tau - \bar{\tau}(z))\,\bar{\mathcal{A}}(z)\ . \tag{30}$$

(ii) In equilibrium we have

$$C_1^e(\tau; z) = \bar{C}_1(z)\ +\ \big(\bar{C}_1(z) - \bar{D}_1(z)\big)\,(\tau - \bar{\tau}(z))\,\bar{\mathcal{A}}(z). \tag{31}$$

(iii) Certainty equivalent consumption for the representative agent is higher than the market value of y, i.e., $\bar{C}_1(z) > \bar{D}_1(z)$, which we can approximate as

$$\bar{C}_1(z) = \bar{y}(z) - \frac{1}{2}\bar{\mathcal{A}}(z)\sigma^2(z) + o(\sigma^2(z)) \text{ and } \bar{D}_1(z) = \bar{y}(z) - \bar{\mathcal{A}}(z)\sigma^2(z) + o(\sigma^2(z)), \tag{32}$$

where $\bar{y}(z)$ and $\sigma^2(z)$ is the variance of y using $\rho(\cdot|z)$.

Note that the term $\bar{C}_1(z) - \bar{D}_1(z)$ is a measure of the risk premium on the market portfolio. In particular, it is the gap between the certainty equivalent consumption of the representative agent in equilibrium $\bar{C}_1(z)$ and the certainty equivalent consumption that any agent would have at $t = 1$, if she sold her one share of the aggregate endowment at price $\bar{D}_1(z)$ and purchased instead a portfolio made up entirely of sure bonds. Note as well that equations (30) and (31) imply a linear relationship between each agent's purchases of risky shares $\phi^e(\tau; z) - 1$ and the deviation of that agent's certainty equivalent consumption from that of the representative agent $C_1^e(\tau; z) = \bar{C}_1(z)$, where the slope of that line is given by the risk premium on the market portfolio $\bar{C}_1(z) - \bar{D}_1(z)$.

From this proposition, the observed incomplete market equilibrium trade volume in shares at $t = 1$ given state z is given by

$$TV^e(z) \equiv \frac{1}{2}\sum_{\tau} |\phi^e(\tau; z) - 1|\mu(\tau|z) = \frac{1}{2}\bar{\mathcal{A}}(z)\sum_{\tau} |\tau - \bar{\tau}(z)|\mu(\tau|z). \tag{33}$$

This measure of trade volume is also a measure of the mean absolute deviation of agents' risk tolerances from the risk tolerance of the agent with average risk tolerance. In other words, observed share trade volumes are a direct measure of the dispersion in agents' risk tolerances. Equation (31) implies as well that dispersion in agents' risk tolerances drives dispersion in agents' equilibrium certainty equivalent consumption. Hence, equations (30) and (31) together imply that observed share trade volumes is a direct measure of dispersion (in terms of mean absolute deviation) of agents' certainty equivalent consumption.

To complete our characterization of date $t = 0$ bond prices in equation (16), we must also compute the marginal cost of certainty equivalent consumption. We do so next.

Type-Independent Marginal Cost of Certainty Equivalent Consumption. Given subutility functions of the equicautious HARA class (18), for any conditionally efficient allocation of consumption, together with the associated certainty equivalent consumptions, the marginal cost of delivering an additional unit of certainty equivalent consumption to any agent of type τ is defined as in equation (17) and independent of type, as the next proposition shows.

Proposition 4. *Assuming U_τ are of the equicautious HARA class, then in an equilibrium with incomplete markets,*

$$\frac{\partial}{\partial C_1} H_\tau(C_1^e(\tau; z); z) = \frac{U'_{\bar{\tau}(z)}(\bar{C}_1(z))}{\sum_y U'_{\bar{\tau}(z)}(y)\rho(y|z)} \tag{34}$$

for all values of τ.

Using the results from Propositions 2, 3, and 4 we have a complete solution of the model for the optimal and equilibrium allocations, their associated asset prices, and the implications of the model for equilibrium trading volumes. We summarize our solution of the model in the following proposition.

Proposition 5. *Let $V(C)$ be strictly concave and let agents have type-dependent subutility functions of the equicautious HARA class (18) with $\frac{y}{\gamma} + \bar{\tau}(z) > 0$ for all y and z.*

(i) Asset prices at $t = 1$ in any conditional equilibrium are given by $\bar{p}(y; z)$ defined in (24) with $\sum_y \bar{p}(y; z)\rho(y|1) = 1$ as the numeraire. The price of a share of the aggregate endowment at $t = 1$ given z is denoted $\bar{D}_1(z)$ and given by (8) at asset prices $\bar{p}(y, z)$. They depend exclusively on the representative agent's valuations.

(ii) The optimal allocation of certainty equivalent consumption is given by $C_0^* = \bar{C}_0$ and $C_1^*(\tau; z) = \bar{C}_1(z)$ defined in (28), while the allocation of certainty equivalent consumption in the equilibrium with incomplete asset markets is given by $C_0^e = \bar{C}_0$ and $C_1^e(\tau; z)$ given as in (31).

(iii) *Date $t = 0$ bond prices $Q^e(z)$ in equilibrium are given by (16) evaluated at the equilibrium allocation of certainty equivalent consumption (31) with common marginal cost of certainty equivalent consumption given as in (34). The prices for these bonds in the decentralization of the optimal allocation are given by (16) with all agents receiving common certainty equivalent consumption $C_1^*(\tau; z) = \bar{C}_1(z)$.*

(iv) *Agents can implement the incomplete markets equilibrium allocation of consumption at time $t = 2$, by trading at $t = 1$ their one share of the aggregate endowment for $\phi^e(\tau; z)$ shares of the aggregate endowment y given as in (30) and holding $\bar{D}_1(z)(1 - \phi^e(\tau; z))$ risk-free bonds. This leads to share turnover of $TV^e(z)$ as in (33).*

The restriction that $\frac{y}{\gamma} + \bar{\tau}(z) > 0$ for all possible values of y in the statement of this proposition is required to ensure that the HARA subutility is well defined for all agents in equilibrium for all values of y.

3.1 Solving the Model as an Endowment Shock Model

When agents have subutility functions of the equicautious HARA class (18), then the equilibrium allocations of certainty equivalent consumption in our model and the associated date $t = 0$ asset prices are equivalent to those of the following economy with idiosyncratic endowment shocks but no preference shocks. This equivalence result, which we demonstrate here, follows from the properties of the equicautious HARA preferences used above. We spell out this mapping of the model to an endowment shock economy to highlight the mathematical connection between the role of idiosyncratic risk in certainty equivalent consumption due to uninsured idiosyncratic risk tolerance shocks in shaping asset prices in our model to the role of idiosyncratic risk in consumption at $t = 1$ due to uninsured idiosyncratic endowment shocks in shaping asset prices in Mankiw (1986) and Constantinides and Duffie (1996). Of course, in our model, the equilibrium allocation of certainty equivalent consumption at $t = 1$ is implemented with a positive volume of asset trade, while there is no such trade in the endowment shock economy.

The endowment shock model is described as follows. Consider an economy with two time periods, $t = 0$ and $t = 1$. Let agents face aggregate uncertainty indexed by z and idiosyncratic uncertainty indexed by τ. Let the probability that state z is realized at time $t = 1$ be given by

$\tilde{\pi}(z)$ with change of measure

$$\tilde{\pi}(z) = \frac{J(z)\pi(z)}{\sum_{z'} J(z')\pi(z')}.$$

The term $J(z)$ is the inverse of the marginal cost of certainty equivalent consumption which, in equilibrium, is common to all agents and, from equation (34), is given by

$$J(z) \equiv \frac{\sum_y U'_{\bar{\tau}(z)}(y)\rho(y|z)}{U'_{\bar{\tau}(z)}(\bar{C}_1(z))} = \frac{\sum_y \left[\frac{y}{\gamma} + \bar{\tau}(z)\right]^{-\gamma} \rho(y|z)}{\left[\frac{\bar{C}_1(z)}{\gamma} + \bar{\tau}(z)\right]^{-\gamma}}, \tag{35}$$

which equals the expected marginal utility of the representative agent relative to that agent's marginal utility of his or her certainty equivalent consumption. This risk adjustment J comes from the Kreps-Porteus-Selden-Epstein-Zin non-expected utility, and hence it is a feature in all the asset pricing models with such preferences. Note that in the case with CARA subutility (i.e., $\gamma \to \infty$), we have $J(z) = 1$ for all z.[7]

Let the distribution of the idiosyncratic uncertainty faced by agents at $t = 1$ in state z be given by $\mu(\tau|z)$. Assume that an agent who has realized type τ in state z has endowment at $t = 1$:

$$Y_1(\tau; z) \equiv \bar{C}_1(z) \; + \; (\tau - \bar{\tau}(z)) \; \bar{\mathcal{A}}(z) \; \left(\bar{C}_1(z) - \bar{D}_1(z)\right).$$

Let the allocation of consumption at $t = 1$ be denoted by $C_1(\tau; z)$. This allocation must satisfy the pseudo-resource constraint (27). As before, let all agents be endowed with $Y_0 = \bar{C}_0$ at time $t = 0$.

Let agents have preferences over allocations given by

$$V(C_0) + \tilde{\beta} \sum_z \sum_\tau V(C_1(\tau; z))\mu(\tau|z)\tilde{\pi}(z)$$

with

$$\tilde{\beta} \equiv \beta \sum_{z'} J(z')\pi(z').$$

In the equilibrium of this endowment shock economy with incomplete asset markets, let agents choose consumption $C(0), C_1(\tau; z)$ and bond holdings $B(z)$ to maximize utility subject

[7] For values of $\gamma < \infty$, we have the Taylor approximation around the conditional mean realization of the endowment, $\bar{y}(z)$:

$$J(z) \approx 1 + \frac{\sigma_y^2(z)}{2} \frac{1/\gamma}{\left(\frac{\bar{y}(z)}{\gamma} + \bar{\tau}(z)\right)^2}. \tag{36}$$

Hence, holding γ fixed, $J(z)$ is increasing in the conditional variance of the endowment, $\sigma_y^2(z)$, and decreasing in the average risk tolerance across agents $\bar{y}(z)/\gamma + \bar{\tau}(z)$ if and only if $\gamma > 0$.

to budget constraints (11) with Y_0 replacing $\bar{C}_0$ at $t = 0$ and, at $t = 1$,

$$C_1(\tau; z) = Y_1(\tau; z) + B(z).$$

The bond market clearing conditions are given by $B(z) = 0$ for all z.

Proposition 6. *The equilibrium allocations $C_0, C_1(\tau; z)$ and date zero bond prices $Q(z)$ for the endowment shock economy are equivalent to the equilibrium allocations of certainty equivalent consumption and date zero bond prices $Q(z)$ for the corresponding taste shock economy.*

Proof. Note that with the change of measure to $\tilde{\pi}(z)$ and the rescaling of the discount factor $\tilde{\beta}$, the bond pricing conditions (16) are the same in the two economies. Direct calculation then shows that the equilibrium allocations and date zero bond prices in our preference shock economy are also equilibrium allocations and bond prices in this endowment shock economy and vice versa. □

This proposition is also useful in establishing a bound on the extent of downside idiosyncratic risk to certainty equivalent consumption that agents can face in this economy. This bound on the downside risk that agents can face does put a bound on the extent to which this idiosyncratic risk can affect asset prices at $t = 0$. Specifically, note that the parameter restrictions we need to ensure that our HARA utility is well defined imply that the lowest possible endowment $Y_1(\tau; z)$ that can be realized is $\bar{D}_1(z)$.

This lower bound has a simple economic interpretation: an agent in our preference shock economy always has the option at $t = 1$ to trade his or her endowment of one share, at price $\bar{D}_1(z)$, for a portfolio made up entirely of risk-free bonds, hence ensuring certainty equivalent consumption of $\bar{D}_1(z)$ independent of that agent's realized risk tolerance τ. Thus, in the equilibrium with incomplete asset markets, the gap between the certainty equivalent consumption of the agent with the lowest realized risk tolerance and the average level of certainty equivalent consumption in the economy is always bounded above by the measure of the aggregate risk premium given by $\bar{C}_1(z) - \bar{D}_1(z)$. This bound restricts the downside risk that agents face ex-ante and hence the premia they are willing to pay at $t = 0$ to avoid the impact of this preference risk on their certainty equivalent consumption at $t = 1$.

4 Trade Volumes and Asset Prices

In Proposition 5, we provided a complete characterization of equilibrium allocations and asset prices under the assumption that agents have subutility functions of the equicautious HARA class. We also characterized trade volumes in asset markets at $t = 1$ under the assumption that agents trade only shares of the aggregate endowment and risk-free bonds. In this section, we study the implications of our model for the joint distribution of trade volumes and asset prices in greater detail.

We first discuss our model's implications for trading volume and expected excess returns as of date $t = 0$. As shown in equation (31), in equilibrium, agents' certainty equivalent consumption is exposed to idiosyncratic shocks to their risk tolerance. We consider the impact of these idiosyncratic shocks to agents' certainty equivalent consumption on asset pricing in terms of multiplicative expected excess returns.

The price at $t = 0$ of a riskless bond, i.e., a claim to a single unit of consumption at $t = 2$ for each possible realization of τ, z, and y, is given by $P_0(1) = \sum_z Q(z)\pi(z)$. We use the inverse of this price to define the risk-free interest rate at $t = 0$ between periods $t = 0$ and $t = 1$ as $\bar{R}_0 = 1/P_0(1)$. Note that this formula follows from our normalization of the riskless bond price at $t = 1$ to one for all realized z.

Consider a security with payoffs $d(y|z)$ at $t = 2$. The time $t = 0$ multiplicative expected excess return of a claim with payoffs d at $t = 2$ is denoted by $\mathcal{E}_{0,2}(d)$, and is defined as

$$\mathcal{E}_{0,2}(d) \equiv \frac{\mathbb{E}_0\left[d(y,z)\right]}{P_0(d)} \Big/ \frac{1}{P_0(1)}. \tag{37}$$

Analogously, the multiplicative expected excess return of a claim with payoffs d at $t = 2$ bought at $t = 1$ in state z is denoted by $\mathcal{E}_{1,2}(d)$, and is defined as

$$\mathcal{E}_{1,2}(z;d) \equiv \frac{\mathbb{E}_1\left[d(y,z)|z\right]}{P_1(z;d)} \Big/ \frac{1}{P_1(z;1)} = \frac{\mathbb{E}_1\left[d(y,z)|z\right]}{P_1(z;d)} \tag{38}$$

since we use the normalization $P_1(z;1) = 1$ for all z. The expressions for $P_1(z;d)$ are given by equation (9), where $\bar{p}(y;z)$ are given by the representative agent marginal utilities defined in (24). The expectations $\mathbb{E}[\cdot]$ are taken with respect to the statistical distribution, i.e., using the probability distributions π and ρ.

From equations (16) and (34), we have

$$Q^e(z) = \beta \frac{V'(\bar{C}_1(z))}{V'(\bar{C}_0)} J(z)L(z) \tag{39}$$

with

$$L(z) \equiv \sum_{\tau} \frac{V'(C_1^e(\tau; z))}{V'(\bar{C}_1(z))} \mu(\tau; z) \tag{40}$$

and $J(z)$ is defined as in equation (35). The random variable L will be key to describing the effect of trade volume on $t = 0$ asset prices.

This gives a complete characterization of asset prices for the incomplete market economy. Next we turn to an analysis of these asset prices based on these expressions.

Trade Volumes, $L(z)$, and Asset Prices: In the incomplete markets equilibrium, there is a direct connection between asset prices and the dispersion of the preference shocks τ realized at $t = 1$ in state z. This connection comes through the term $L(z)$ in Q^e in equations (39) and (40). Under the assumption that $V''' > 0$, the term $L(z)$ is equal to one if there is no dispersion in τ and is strictly increasing in the dispersion in τ. Specifically,

$$V'''(\cdot) \geq 0 \text{ implies } L(z) = \sum_{\tau} \frac{V'(C_1^e(\tau; z))}{V'(\bar{C}_1(z))} \mu(\tau; z) \geq 1$$

so that $L(z)$ is the extra valuation in state z for a prudent agent facing rebalancing risk in state z. Using a Taylor expansion, so that the remainder is of smaller order than the conditional variance of τ we have

$$\begin{aligned} L(z) &= 1 + \frac{1}{2} \frac{V'''(\bar{C}_1(z))}{V'(\bar{C}_1(z))} \sum_{\tau} \left[C_1^e(\tau, z) - \bar{C}_1(z)\right]^2 \mu(\tau; z) + o(\sigma^2(\tau|z)) \\ &= 1 + \frac{1}{2} \frac{V'''(\bar{C}_1(z))}{V'(\bar{C}_1(z))} \left[\bar{C}_1(z) - \bar{D}_1(z)\right]^2 \sum_{\tau} \left[\phi^e(\tau; z) - 1\right]^2 \mu(\tau; z) + o(\sigma^2(\tau|z)). \end{aligned} \tag{41}$$

Hence, if $V'''(\bar{C}_1(z)) > 0$, then our approximation to $L(z)$ is directly proportional to the variance of individual share trades times the square of the aggregate consumption risk premium as measured by $\left(\bar{C}_1(z) - \bar{D}_1(z)\right)^2 = \frac{1}{2}\bar{\mathcal{A}}(z)\sigma^2(z) + o(\sigma^2(z))$ using the approximation in Proposition 3.

As an example, consider the case of a uniform distribution of τ. For a uniform distribution of τ, the mean absolute deviation of τ from $\bar{\tau}(z)$ is directly proportional to the standard deviation of τ and hence, in this case, to a second order approximation, data on the square of trading volume in state z are a valid proxy for the term $\sum_{\tau} \left(\phi^e(\tau; z) - 1\right)^2 \mu(\tau|z) + o(\sigma^2(\tau|z))$ in our approximation to $L(z)$.

Of course, the previous result that the square of trading volume is directly proportional to the dispersion of agents' marginal utilities of certainty equivalent consumption is special to the case of uniform shocks to risk tolerance. More generally, if one had data on the distribution of trade sizes, one could potentially map data on trade volumes to empirical proxies for $L(z)$ using the relevant distributional assumptions. Moreover, since $|x-1|$ and $(x-1)^2$ are both convex functions, if we consider a mean-preserving spread in the distribution of risk tolerance, both trade volume and $L(z)$ increase. In this sense, both trade volumes and $L(z)$ are increasing in the dispersion of idiosyncratic risk tolerance, and hence $L(z)$ measures the exposure to rebalancing risk of state z.

4.1 Comparative Statics on Trade Volumes and Asset Prices

We now develop three results regarding the impact of trading volumes on asset pricing. The first one is a result about interest rates, the second one is a comparison of risk premia across economies with different patters of trading volume, and the third one is a comparison of the risk premia on different assets in the same economy.

For these results, it is useful to collect two properties of asset prices. Define

$$Q^*(z) = \beta \frac{V'(\bar{C}_1(z))}{V'(\bar{C}_0)} J(z).$$

Note that this bond prioe $Q^*(z)$ is the bond price that would obtain in the decentralization of the optimal allocation. Likewise, it is the equilibrium bond price in an economy in which there is no dispersion in realized risk tolerances and hence no trade volume at $t=1$.

Then we have that

$$P_0(d) = \sum_z Q^*(z)L(z)\pi(z)P_1(z;d) \tag{42}$$

for all assets with dividend d. This expression, together with the previous definitions of expected excess returns, gives the following expression for the (inverse) time $t=0$ expected excess return as a weighted average of the time $t=1$ expected excess returns:

$$\frac{1}{\mathcal{E}_{0,2}(d)} = \sum_z \frac{Q^*(z)L(z)\pi(z)}{\sum_{z'} Q^*(z')L(z')\pi(z')} \frac{\mathbb{E}_1\left[d(y,z)|z\right]}{\mathbb{E}_0\left[d(y,z)\right]} \frac{1}{\mathcal{E}_{1,2}(z;d)}, \tag{43}$$

which we summarize in the following proposition.

Proposition 7. *Take a payoff $d(y,z)$ at $t=2$. The time $t=0$ (inverse) excess expected return $\mathcal{E}_{0,2}(d)$ of a this payoff is the risk-neutral complete market expected value of the product of three random variables realized at $t=1$, i.e., functions of the realization of z. These are exposure to* *rebalancing risk, $L/\mathbb{E}_0^*[L]$, updates in the asset expected payoffs $\mathbb{E}_1[d]/\mathbb{E}_0[d]$, and the (inverse) excess expected returns $\mathcal{E}_{1,2}(d)$, i.e.,*

$$\frac{1}{\mathcal{E}_{0,2}(d)} = \mathbb{E}_0^* \left[\frac{L}{\mathbb{E}_0^*[L]} \, \frac{\mathbb{E}_1[d]}{\mathbb{E}_0[d]} \, \frac{1}{\mathcal{E}_{1,2}(d)} \right], \tag{44}$$

where $\mathbb{E}_0^*[x]$ is the time $t = 0$ complete-market risk-neutral expected value of $t = 1$ random variable $x(z)$, i.e.,

$$\mathbb{E}_0^*[x] \equiv \frac{\sum_z x(z)\, Q^*(z)\pi(z)}{\sum_{z'} Q^*(z')\pi(z')}. \tag{45}$$

Note that the expression $\mathbb{E}_0^*[L]$ is the time $t = 0$ value of a riskless bond paying 1 at $t = 1$.

The expression (43), or its equivalent form (44), is key to show the two results on excess returns. Note that the only term in this asset pricing formula that involves the dispersion of $\mu(\cdot|z)$ is the term $L(z)$. The expressions for $Q^*(z)$, $\pi(z)$, $P_1(z;d)$, $\mathcal{E}_{1,2}(z;d)$, and $\mathbb{E}_1[d|z]/\mathbb{E}_0[d]$ are not functions of the shape of $\mu(\cdot|z)$ other than the mean of this distribution, and hence they are *independent of trade volume*. The ratio $\mathbb{E}_1[d|z]/\mathbb{E}_0[d]$ measures the exposure of the cash flow d to the state z. Hence, expected excess returns at $t = 0$ depend on trade volume only through the correlation of trade volume with future expected returns $\mathcal{E}_{1,2}(z;d)$ or with expected cash flows $\mathbb{E}_1[d|z]/\mathbb{E}_0[d]$.

Trade Volume and Interest Rates. We have the following comparative static result regarding the dispersion of shocks to risk tolerance, and hence trade volume, and time $t = 0$ bond prices $P_0(1) = \sum_z Q(z)\pi(z)$.

Proposition 8. *Consider two economies in which agents have the same preferences with $V'''(\cdot) > 0$ and face the same distribution of endowments, $\bar{C}_0$, $\pi(z)$, and $\rho(y|z)$. Assume that the distribution of shocks to risk tolerance in the two economies $\mu(\tau|z)$ and $\mu'(\tau;z)$ are such that, for all j, $\bar{\tau}(z_j) = \bar{\tau}'(z_j)$. Then these two economies have the same equilibrium values of $\bar{C}_1(z)$ and $J(z)$, but, for each state z, the economy with the higher dispersion in shocks to risk tolerance as measured by a higher value of $L(z)$ has the higher equilibrium bond price at $t = 0$, $Q^e(z)$.*

Proof. The proof is by direct calculation. □

Given our previous result that trade volume and dispersion in certainty equivalent consumption are both increasing in the dispersion of shocks to risk tolerance, we have that, for each state z, the economy that has the higher trade volume has the higher equilibrium bond price at $t = 0$, $Q^e(z)$.

Trading Volume and Expected Returns (Risk Premia) across Economies. We now compare an economy with a constant dispersion of risk tolerance across different states z at $t = 1$ with one in which the marketwide risk tolerance is negatively correlated with the dispersion of risk tolerance. We find that if V displays prudence (i.e. if $V''' > 0$), then any cash flow with systematic risk has a higher risk premium in the economy in which dispersion is negatively correlated with risk tolerance.

Denote by $\tilde{\mu}(\cdot|z)$ the distribution of $(\tau - \bar{\tau}(z))\bar{\mathcal{A}}(z)$ conditional on z. We consider the following assumptions:

$$\text{If } z' > z, \text{ then } \bar{\tau}(z') > \bar{\tau}(z) \text{ and} \tag{46}$$

$$\text{If } z' > z, \text{ then } \tilde{\mu}(\cdot|z') \text{ is less dispersed (in a second-order stochastic sense) than } \tilde{\mu}(\cdot|z). \tag{47}$$

In words, states with higher marketwide risk tolerance have a lower dispersion of risk tolerance and thus a lower volume of trade at $t = 1$. We say that an asset has systematic payoff exposure if d is increasing in y, and

$$d(y', z') > d(y, z) \text{ for all } z, z' \text{ and } y' > y\,. \tag{48}$$

With this notation at hand, we can state the following result:

Proposition 9. *Let the distribution of y not vary with z, so $\rho(y|z) = \bar{\rho}(y)$ for all states z. Consider two economies where shock z indexes marketwide risk tolerance as in (46). The first economy has constant dispersion on the idiosyncratic risk tolerance across states at time $t = 1$, so $L(z) = L_1(z)$ is constant for all z. The second economy has $\mu(\cdot|z)$ more dispersed for lower marketwide risk tolerance as defined in (47), so $L(z) = L_2(z)$ is decreasing in z. We fix the same asset $d(\cdot)$ with a systematic payoff exposure as defined in (48) in both economies. If investors are prudent (i.e., they have precautionary savings motives, or $V''' > 0$), then the second economy (where the cross-sectional dispersion in risk tolerance is negatively correlated with the marketwide risk tolerance) has a higher $t = 0$ expected excess return $\mathcal{E}_0(d)$.*

We can describe the second economy as one where the cross-sectional dispersion in risk tolerance is negatively correlated with the marketwide risk tolerance, or equivalently as one in which trade volume is positively correlated with the marketwide risk aversion. Thus, in this second economy, times of high marketwide risk aversion (times in which the risky assets sell at a low price) are times in which lots of investors want to rebalance.

This proposition parallels the results in Mankiw (1986) and Constantinides and Duffie (1996). In both papers, the authors consider the level of excess expected returns when investors have uninsurable labor risk whose dispersion is correlated with the level of aggregate consumption. While their results are mathematically parallel to ours, it is important to note that there is no trade in assets in their models.

Trading Volume and Expected Returns (Risk Premia) in the Cross Section of Assets. We now compare the risk premia across risky assets in the same incomplete markets economy. In this case, we find that if V displays prudence (i.e., if $V''' > 0$), then assets with cash flows that load more onto the time $t = 1$ states z with higher dispersion of risk tolerances have higher prices or lower expected returns. Since trade volume is also given by a measure of dispersion of risk tolerances, this result means that assets whose cash flows load on states at $t = 1$ with high trade volume have low expected excess returns.

To make this result precise, we fix an economy with incomplete markets and compare the excess expected returns of assets with different exposures to the idiosyncratic dispersion of risk tolerance. We assume that the average risk tolerance and the distribution of the endowment y conditional on z are both constant across z.

Proposition 10. *Consider an economy with incomplete markets with the same marketwide risk tolerance $\bar{\tau}(z)$ and the same conditional distribution of aggregate risk $\rho(y|z) = \bar{\rho}(y)$ for all states z. Assume that the states z are ordered in terms of dispersion of idiosyncratic risk tolerance as in (47), so that $L(z)$ decreases with z. Consider two cash flows, $\tilde{d}$ and d, in the same economy, where $\tilde{d}$ loads more than d in states with higher dispersion of risk tolerance in the following way: $\tilde{d}(y, z) = \delta(y)\tilde{e}(z)$ and $d(y, z) = \delta(y)e(z)$ with $\tilde{e}(z)/e(z)$ decreasing in z. Then, the time $t = 1$ conditional expected excess returns are the same for both assets and all states z, i.e., $\mathcal{E}_{1,2}\big(z; \tilde{d}\big) = \mathcal{E}_{1,2}\big(z; d\big)$, and the time $t = 0$ expected excess returns for the asset with higher exposure to trade volume are smaller, i.e., $\mathcal{E}_{0,2}\big(\tilde{d}\big) < \mathcal{E}_{0,2}\big(d\big)$.*

This result gives conditions under which trade volume acts as a pricing factor, i.e., the conditions under which the cross-sectional expected excess returns on assets (i.e., $\mathcal{E}_{0,2}(\tilde{d})$ versus $\mathcal{E}_{0,2}(d)$) depend on the correlation of returns with trade volume. In this case, the asset with dividend $\tilde{d}$, which has higher value when trade volume is high, and thus ex ante is a better hedge against the rebalance risk, has a higher price, i.e., it has a lower $t = 0$ expected excess return. The higher price of the asset with dividend $\tilde{d}$ is due to higher exposure of its dividends to trade volume, as captured by the term $\mathbb{E}_1[\tilde{d}|z]/\mathbb{E}_0[\tilde{d}] = \tilde{e}(z)/\mathbb{E}_0[\tilde{e}(z)]$.

5 Taxes on Trading and Ex-ante Welfare

In this section, we consider the implications for welfare of a tax on trade in shares of the aggregate endowment at $t = 1$ (a *Tobin tax*). We show that while a Tobin tax on trade has a zero first-order effect on welfare if imposed on the socially optimal allocation, it has a *first-order negative* welfare effect if imposed on the equilibrium allocation. In other words, a small Tobin subsidy to trade increases ex-ante welfare in equilibrium. The basic logic of this result is that a Tobin tax exacerbates the inefficient sharing of idiosyncratic preference risk in equilibrium. Agents who have negative risk tolerance shocks suffer a negative shock to certainty equivalent consumption in equilibrium. The Tobin tax also falls on them in terms of its tax incidence because their demand for shares is relatively inelastic. Hence, the tax exacerbates the inefficient sharing of risk in certainty equivalent consumption in equilibrium.

In the online appendix, we complement our analysis of a simple linear tax of trade rebated lump sum, with the analysis of the optimal non-linear tax or subsidy scheme. We use a standard mechanism design approach, assuming that the realization of individual risk tolerance is private information for each investor. We think of this assumption as the natural explanation of why we assume that these risks are uninsurable. We use the optimal non-linear scheme to judge the sense in which a subsidy to trade is a general feature of the optimal policy. We conclude that, consistent with the results on Tobin taxes, the optimal non-linear tax/subsidy is one that induces more trade.

5.1 A Tobin Tax

In our analysis of a Tobin tax, we assume that there are two asset markets — one at $t = 0$ for bonds that pay off at $t = 1$ and one at $t = 1$ in which agents trade shares of the aggregate endowment for sure claims to consumption at $t = 2$. Assume that trade in shares at $t = 1$ is taxed. Specifically, assume that there is a tax of ω per share traded such that if the seller receives price $\bar{D}_1(z)$ for selling a share of the aggregate endowment at $t = 1$, the buyer pays $\bar{D}_1(z) + \omega$, and the total revenue collected through this tax, equal to ω times the volume of shares traded, is rebated lump sum to all agents.

With this notation, we define a conditional equilibrium with a Tobin tax as follows.

Conditional Equilibrium with a Tobin Tax. An *equilibrium conditional on* z *with a Tobin tax* ω is a share price $\{\bar{D}_1(z;\omega)\}$, transactions tax revenue rebate $T(z;\omega)$, post-trade holdings of share $s(\tau;z;\omega)$, and corresponding allocation of consumption at $t=2$, $c(\tau,y;z;\omega)$, that satisfy the market clearing condition

$$\sum_{\tau} s(\tau;z;\omega)\mu(\tau|z) = 0, \tag{49}$$

and budget constraints,

$$c(\tau,y;z;\omega) = ys(\tau;z;\omega) - \left(\bar{D}_1(z;\omega)+\omega\right)\left(s(\tau;z;\omega)-1\right) + T(z;\omega) + B(\tau,z)$$

if $s(\tau;z;\omega) \geq 1$ and

$$c(\tau,y;z;\omega) = ys(\tau;z;\omega) - \bar{D}_1(z;\omega)\left(s(\tau;z;\omega)-1\right) + T(z;\omega) + B(\tau,z)$$

if $s(\tau;z;\omega) < 1$. The rebate $T(z;\omega)$ satisfies the government budget constraint $T(z;\omega) = \omega TV(z;\omega)$ where trade volume $TV(z;\omega)$ is given by

$$TV(z;\omega) = \sum_{\tau:s(\tau;z;\omega)>1} \left(s(\tau;z;\omega)-1\right)\mu(\tau|z). \tag{50}$$

Equilibrium share holdings $s(\tau;z;\omega)$ maximize each agent's certainty equivalent consumption (3) among all share holdings and allocations of consumption that satisfy the budget constraints given the initial bond holdings, the share price, the tax, and the tax rebate.

We denote by $C_1^i(\tau;z;\omega)$ the time $t=1$ certainty equivalent consumption for agent τ in state z for the conditional equilibrium with a Tobin tax ω with initial bond holdings $B^i(\tau,z)$ for $i \in \{*,e\}$ corresponding to those in the decentralization of the optimal allocation and equal wealth equilibrium, respectively. Likewise, let s^i denote the shareholdings in the conditional equilibrium with a Tobin tax with initial bond holdings $B^i(\tau,z)$ for $i \in \{*,e\}$. Let $\tilde{H}_\tau^i(C_1;z;\omega)$ denote the minimum cost to an agent of type τ of attaining certainty equivalent consumption C_1 at $t=1$ in state z by trading in shares and risk-free bonds subject to Tobin tax ω.

Consider the following calculation of the change in ex-ante welfare from a marginal increase in the transactions tax ω starting from $\omega=0$. Here we must compute

$$\left.\frac{dW^i}{d\omega}\right|_{\omega=0} = \beta\sum_z \pi(z)\sum_\tau \mu(\tau|z)V'(C_1^i(\tau;z;0))\frac{d}{d\omega}C_1(\tau;z;0), \tag{51}$$

where $C_1^i(\tau;z;0)$ is the initial allocation of certainty equivalent consumption (with $\omega=0$) corresponding to either the optimal allocation or the equilibrium allocation.

Using that any conditionally efficient allocation of certainty equivalent consumption must satisfy the pseudo-feasibility constraint (27), we must have that

$$\sum_{\tau} \mu(\tau|z) \frac{d}{d\omega} C_1^*(\tau; z; 0) = \sum_{\tau} \mu(\tau|z) \frac{d}{d\omega} C_1^e(\tau; z; 0) = 0. \tag{52}$$

(This restriction follows from the observation that all perturbations to the initial conditionally efficient allocation of certainty equivalent consumption must remain inside the set of feasible allocations of certainty equivalent consumption.) This result implies that a Tobin tax, at the margin, simply redistributes certainty equivalent consumption across agents, regardless of whether the initial allocation of certainty equivalent consumption corresponds to either the optimal or the equilibrium allocation.

Since in the optimal allocation, $C_1^*(\tau; z) = \bar{C}_1(z)$ for all τ, the formula (51) together with equation (52) then immediately implies the standard result that a share transactions tax has no first-order impact on welfare starting from the optimal allocation since all types of agents share the same initial marginal utilities of certainty equivalent consumption in each state z.

In contrast, in the incomplete markets economy, the baseline equilibrium allocation of certainty equivalent consumption at $t = 1$ is not socially efficient. The restriction (52) that the aggregate change in certainty equivalent consumption must be zero gives us that the total change in ex-ante welfare in equation (51) can be written as

$$\frac{dW}{d\omega} = \beta \sum_{z} \pi(z) \mathbb{C}ov \left(V'(C_1(\tau; z)) \, , \, \frac{d}{d\omega} C_1(\tau; z) \, | \, z \right), \tag{53}$$

where $\mathbb{C}ov(\cdot, \cdot \,|\, z)$ denotes the covariance of two variables dependent on τ conditional on z. As shown in equation (31), certainty equivalent consumption for an agent with realized type τ in state z at $t = 1$ is strictly increasing in the risk tolerance τ of that agent. If V is strictly concave, the marginal utility of certainty equivalent consumption for an agent with realized type τ in state z at $t = 1$, $V'(C_1^e(\tau; z))$, is strictly decreasing in the risk tolerance of that agent. Thus, the first-order impact on welfare of a tax on trading in shares imposed on the equilibrium allocation is then determined by whether it is agents with high or low marginal utilities of certainty equivalent consumption in the initial equilibrium allocation who bear the cost of the tax net of the lump sum transfer of tax revenue. In other words, the welfare implications of a Tobin tax depend on the incidence of that tax.

To study the incidence of a Tobin tax, we must solve for the the changes in certainty equivalent consumption by type τ that arise from the direct effect of the tax on agents' consumption at $t = 2$ and the changes in certainty equivalent consumption that arise indirectly from the change in the share price and the lump sum transfer induced by the tax. We do so by differentiating the cost minimization problem that determines $\tilde{H}^i_\tau(C_1; z; \omega)$. Using the envelope theorem, together with the observation that with equicautious HARA subutility, the cost functions $\tilde{H}^i_\tau$ coincide with the cost functions H_τ defined in equation (14) when $\omega = 0$, we get that

$$\frac{d}{d\omega} C^i_1(\tau; z; 0) = \begin{cases} J(z) \left[(s^i(\tau; z; 0) - 1) \left(-\frac{\partial \bar{D}^i_1(z;0)}{\partial \omega} - 1 \right) + TV^i(z; 0) \right] & \text{if } s^i(\tau; z; 0) > 1 \\ J(z)\, TV^i(z; 0) & \text{if } s^i(\tau; z; 0) = 1 \\ J(z) \left[(s^i(\tau; z; 0) - 1) \left(-\frac{\partial \bar{D}^i_1(z;0)}{\partial \omega} \right) + TV^i(z; 0) \right] & \text{if } s^i(\tau; z; 0) < 1 \end{cases} \quad (54)$$

for all τ, z, and $i \in \{e, *\}$, and where $J(z)$ is given by expression (35).

From equation (54), we see that, in general, the incidence of a Tobin tax on each type of agent depends on the quantity of shares that they trade. A Tobin tax lowers the equilibrium share price, with $\frac{\partial \bar{D}^i_1(z;0)}{\partial \omega} \in [-1, 0]$. Thus, a Tobin tax tends to lower the certainty equivalent consumption of agents with extremely low values of τ. That is because these agents wish to sell a large number of shares, and thus the impact of the tax on the certainty equivalent consumption of these agents through the impact of the tax on the price at which these agents can sell their shares is larger than the gain to these agents from the lump sum transfer of tax revenue. A Tobin tax also lowers the certainty equivalent consumption of agents with extremely high values of τ. In contrast, a Tobin tax benefits agents with values of τ close to the mean value $\bar{\tau}(z)$, as these agents do not wish to trade shares but do benefit from the lump sum transfer of tax revenue. This observation that agents with values of τ close to the mean value of τ tend to benefit from a Tobin tax and those with extreme values of τ, either low or high, tend to lose from a Tobin tax does not directly allow us to compute the covariance term that determined the change in welfare in equation (53). Instead, we make progress on computing that covariance as follows.

First observe that the total gain in certainty equivalent consumption for those agents that are buyers of shares (and hence have values of $\tau > \bar{\tau}(z)$) is positive if and only if the endogenous fall in the share price that results from the tax is sufficiently large in magnitude, i.e.,

$$-\frac{\partial \bar{D}^i_1(z; 0)}{\partial \omega} \geq \sum_{\tau : s(\tau; z; \omega) \leq 1} \mu(\tau | z). \quad (55)$$

That is, buyers of shares, in the aggregate, benefit from a Tobin tax if the magnitude of the fall in the share price in response to the tax is larger than the fraction of agents who are sellers of shares. Since the equilibrium change in the share price that results from the imposition of a Tobin tax depends on the elasticities of demand of buyers and sellers of shares, this result is the version of the classic public finance result that the sellers of a good (here, shares) subject to a tax bear the incidence of the tax if their individual demands are relatively inelastic, and thus the equilibrium price falls by more than the weight of these sellers in the population. To derive inequality (55), use equations (50) and (54) to compute

$$\sum_{\tau : s(\tau;z;\omega)>1} \frac{d}{d\omega} C_1^i(\tau;z;0)\mu(\tau|z) = J(z)\,TV^i(z;0)\,\left[\left(-\frac{\partial \bar{D}_1^i(z;0)}{\partial \omega} - 1\right) + \sum_{\tau : s(\tau;z;\omega)>1} \mu(\tau|z)\right]$$

and recall that both $J(z)$ and trade volume are positive.

That inequality (55) is satisfied in our model is simply a reflection of the classic result in finance that an agent's elasticity of demand for risky shares is increasing in that agent's risk tolerance. In fact, given our assumption of equicautious HARA preferences, we can derive a very simple formula for the equilibrium decline in the share price that occurs in response to the imposition of a Tobin tax. To derive this formula, we first compute the changes in demand for shares by each type τ of agent as a function of the change in price and lump sum transfer induced by the tax. The implied equilibrium change in price then follows from the share market clearing condition.

Consider the derivatives of agents' demands for shares with respect to a change in the price of shares and a lump sum transfer. The first-order condition for the risky asset trade is

$$\mathbb{E}\left[U_\tau'\left(\, y + \left(S\left(\tau; z, D, T\right) - 1\right)(y - D) + T\,\right)(y - D)\;|\; z\right] = 0, \; . \tag{56}$$

where the expectation is taken with respect to the random variable y. Differentiating this first-order condition and evaluating it at the equal wealth equilibrium, we obtain the following result.

Lemma 1. *Let $S(\tau; zD, T)$ be defined as the solution of (56) evaluated at the equilibrium price $D = \bar{D}_1(z)$ for equal wealth and at transfer $T = 0$. Then:*

$$\frac{\partial S(\tau;z;D,T)}{\partial D} = \phi^e(\tau,z)\frac{\mathbb{E}\left[U_{\bar{\tau}(z)}'(y)\,|z\right]}{\mathbb{E}\left[U_{\bar{\tau}(z)}''(y)\,(y-D)^2\,|z\right]} + \left(\phi^e(\tau,z) - 1\right)\,\frac{\mathbb{E}\left[U_{\bar{\tau}(z)}''(y)\,(y-D)\,|z\right]}{\mathbb{E}\left[U_{\bar{\tau}(z)}''(y)\,(y-D)^2\,|z\right]}$$

$$\frac{\partial S(\tau;z;D,T)}{\partial T} = -\frac{\mathbb{E}\left[U_{\bar{\tau}(z)}''(y)\,(y-D)\,|z\right]}{\mathbb{E}\left[U_{\bar{\tau}(z)}''(y)\,(y-D)^2\,|z\right]}.$$

Three comments about this lemma are in order, as these results play an important role in our calculation of the welfare impact of a Tobin tax. First, observe that the derivative of share trades of an agent with risk tolerance τ with respect to the price is increasing in the risk tolerance of that agent. That is, agents who are more risk tolerant adjust their share demand more in response to a change in price than do agents who are less risk tolerant. Second, observe that the second term in the derivative of agents' demand for shares with respect to a change in price D, the term which reflects the income effects on demand from a change in price, cancels out when aggregated across all agents, since the market for shares clears. This result is also an implication of the result that equicautious HARA preferences satisfy Gorman aggregation. Third, the result that the derivative of agents' demand for shares with respect to a transfer is common across all agents is simply an implication of the result that equicautious HARA preferences satisfy Gorman aggregation.

These second and third features of the demand for shares in our economy allow us to compute the change in price that arises from a change in the Tobin tax simply as a function of trade volume in shares and the fractions of agents who are buyers and sellers of shares. Other parameters of preferences do not enter into this calculation. In particular, using Lemma 1 and these conditions, we derive the following characterization for the impact of prices of a Tobin tax.

Proposition 11. *Let $\bar{D}(z;\omega)$ be equilibrium price of a claim to the aggregate endowment with a tax on trade ω introduced in the equal wealth equilibrium. Assume, to simplify, that there are no marginal investors, i.e., μ has no mass point at $\tau = \bar{\tau}$. Then the price $\bar{D}(z;\omega)$ received by sellers decreases by the fraction of shares held post-trade by buyers times the Tobin tax, i.e.,*

$$\frac{dD(z;0)}{d\omega} = -\sum_{\tau>\bar{\tau}} \phi^e(\tau;z)\mu(\tau|z) = -\left[TV^e(0;z) + \sum_{\tau>\bar{\tau}} \mu(\tau|z)\right] \in (-1\,,\,0)\ . \tag{57}$$

From equations (55) and (57), we have that, in the aggregate, buyers of shares (relatively risk-tolerant individuals) benefit from a Tobin tax if and only if the initial equilibrium trade volume exceeds the difference between the measure of buyers of shares and the measure of sellers of shares. This observation gives as an immediate result that if the distribution of preference shocks $\mu(\tau|z)$ is symmetric, so that the measures of buyers and sellers are equal, then, on average, those experiencing negative shocks to risk tolerance (sellers of shares) lose certainty equivalent consumption and those experiencing positive shocks to risk tolerance (buyers of shares) gain certainty equivalent consumption. In fact, the next proposition, proved in the online appendix, shows that the average gains (or losses) for buyers (and sellers) of the risky asset after the introduction of a small transaction tax ω are an extremely simple function of trade volume prior to the introduction of taxes.

Proposition 12. *Assume that $\mu(\cdot|z)$ is symmetric around $\bar{\tau}$. Then the average consumption equivalent gain among all buyers (respectively, losses among sellers) of the risky asset is proportional to the square of trade volume:*

$$\text{Avg. Gain Buyers} \equiv \sum_{\tau > \bar{\tau}} \frac{d}{d\omega} C_1^e(\tau; z; 0) \frac{\mu(\tau|z)}{\sum_{\tau' > \bar{\tau}} \mu(\tau'|z)} \, \omega = +2\, J(z) \left[TV^e(z)\right]^2 \omega,$$

$$\text{Avg. Loss Sellers} \equiv \sum_{\tau < \bar{\tau}} \frac{d}{d\omega} C_1^e(\tau; z; 0) \frac{\mu(\tau|z)}{\sum_{\tau' < \bar{\tau}} \mu(\tau'|z)} \, \omega = -2\, J(z) \left[TV^e(z)\right]^2 \omega \, .$$

Three comments about this proposition are in order. First a corollary of this proposition is that in the case of a symmetric distribution $\mu(\cdot|z)$ with only *two values of* τ, there is a first-order welfare loss of introducing a Tobin tax ω. This is because the marginal utility of buyers of risky assets is discretely below the marginal utility of sellers. Second, since this result gives a strict inequality, it suggests that in the case of two values of τ, one can relax the assumption of symmetry of $\mu(\cdot|z)$. Indeed, Proposition 13 shows that. Third, and more subtly, the result in Proposition 12 does *not* imply that, assuming symmetry, there is a first-order loss in welfare for a Tobin tax ω. The reason why this is not sufficient is that, in general, there is also a redistribution of certainty equivalent consumption among sellers and among buyers, as those who have intermediate values of τ sell or buy only a small quantity of shares and hence do not suffer from tax-induced changes in share prices while still benefiting from the lump sum transfer of tax revenue. Proposition 14 imposes extra conditions on the utility function V so that these potential redistributional effects do not overturn the result that a Tobin tax has a first-order negative impact on welfare.

We now prove our result that a Tobin tax imposes a first-order welfare loss in an economy with only two possible realizations of τ. We then present this result in an economy with a symmetric distribution of shocks to risk tolerance $\mu(\tau|z)$.

Proposition 13. *Fix a value of z. Consider an economy with only two types of agents, $\tau \in \{\tau_1, \tau_2\}$ with $\tau_1 < \bar{\tau}(z) < \tau_2$, and thus $\phi^e(\tau_1; z) < 1 < \phi^e(\tau_2; z)$. Assume that V is strictly concave. Then, when agents have equicautious HARA preferences, a Tobin tax on asset trade imposed on the equal wealth equilibrium has a negative first-order ex-ante welfare effect if and* *only if*

$$(\phi^e(\tau_2; z) - 1)\mu(\tau_2; z) > \mu(\tau_1; z) - \mu(\tau_2; z). \tag{58}$$

Note that $\mu(\tau_2|z) > 1/2$ is a sufficient condition for the Tobin tax to induce a first-order welfare loss. Also, symmetry of the distribution of τ implies that $\mu(\tau_2|z) = 1/2$ and hence satisfies condition (58). Inequality (58) is, of course, again our condition that buyers of shares (relatively risk-tolerant individuals) in the aggregate benefit from a Tobin tax if and only if the initial equilibrium trade volume exceeds the difference between the measure of buyers of shares and the measure of sellers of shares. With only two types of agents, this condition is sufficient to sign the covariance term in equation (53) and thus prove our result that a Tobin tax has a first-order negative impact on welfare.

Now we extend the result to the case of a general symmetric distribution $\mu(\cdot, z)$ and where V is concave with derivatives that alternate signs.

Proposition 14. *Fix a state z. Assume that there are no marginal investors, i.e., $\mu(\cdot|z)$ has no mass point at $\tau = \bar{\tau}$, and that the distribution of τ is symmetric, i.e., $\mu(\bar{\tau} - a; z) = \mu(\bar{\tau} + a; z)$ for all a. Furthermore, assume that the ex-ante utility V is analytical, strictly increasing, and strictly concave, with all derivatives evaluated at $\bar{C}_1(z)$ alternating signs, i.e., the utility function is "proper":*

$$sign\left(\frac{\partial^{n+1}V(C)}{\partial C^{n+1}}\right) = -sign\left(\frac{\partial^{n}V(C)}{\partial C^{n}}\right) \quad \text{evaluated at } C = \bar{C}_1(z), \text{ and all } n = 1, 2, 3, \ldots \tag{59}$$

Then, when agents have equicautious HARA preferences, a Tobin tax on asset trade on the equal wealth equilibrium has a negative first-order ex-ante welfare effect for each z. Moreover, approximating the change on ex-ante utility in terms of moments of τ, and using the first leading term, we obtain

$$\frac{d}{d\omega}W^e(0; z) \approx J(z)\, V''(\bar{C}_1(z)) \left[\bar{A}(z)\right]^2 \left(\bar{C}_1(z) - \bar{D}_1(z)\right) TV^e(z)\, Var(\tau|z), \tag{60}$$

where TV^e is the trade volume in the equal wealth equilibrium.

A few comments are in order. Functions with alternative signs as in (59) are called "completely monotone" and satisfy several properties. First, the assumption that the derivatives of V change sign includes the case of polynomials, such as quadratic utility. Second, since V is concave, this assumption is consistent with V displaying prudence, a key property that we use above in the asset pricing implications. Third, although we did not emphasize this in the statement of the proposition, in the proof we show that every extra term in the approximation corresponding to higher order derivatives is negative. Fourth, the most commonly used utility functions satisfy the condition that derivatives of higher order change signs; for instance, they include all the HARA utility functions. Fifth, Pratt and Zeckhauser (1987) have shown that completely monotone utility functions are "proper," so that independent risks exacerbate each other (i.e., a decision maker with a proper utility function finds that an undesirable lottery cannot be made desirable by adding an independent undesirable lottery to it). Sixth, Caballe and Pomansky (1996) studied the characterization and properties of the choices of decision makers using this type of utility function.

Finally, we can evaluate the expression (60) for the particular case where U_τ is CARA (so that $\gamma \to \infty$) and $y \sim N(\mu, \sigma_y^2(z))$. In this case, $\bar{C}_1(z) - \bar{D}_1(z) = .5\sigma_y^2(z)/\bar{\tau}(z)$, and $J(z) = 1$. Additionally, if $\mu(\cdot|z)$ is uniform, then $Var(\tau|z) = [TV^e(z)]^2 [\bar{\tau}(z)]^2 64/12$. Hence, writing the expected utility $W^e(\omega; z)$ in ex-ante equivalent terms, i.e., defining $C_w^e(\omega; z)$ as $V(C_w^e(\omega; z)) = \sum_\tau V(C_a^e(\tau; z; \omega))\mu(\tau)$, we get

$$\frac{d}{d\omega} W^e(0; z) \approx \frac{16}{3} \frac{V''(\bar{C}_1(z))}{V'(\bar{C}_1(z))} \frac{\sigma_y^2(z)}{2\bar{\tau}(z)} [TV^e(z)]^3 .$$

Thus, the welfare loss of a Tobin tax is proportional to the curvature of the utility function V, the representative agent time $t = 1$ risk premium $0.5\sigma_y^2(z)/\bar{\tau}(z)$, and the *cube of the trade volume.*

Note that in this economy, when a Tobin tax has a first-order negative impact on welfare if applied to the equilibrium allocation with incomplete markets, then a Tobin subsidy to trade must have a positive first-order impact on ex-ante welfare. This observation raises the question of what the optimal subsidy to trade looks like. We take up this question in the online appendix where we present a mechanism design approach to study the optimal non-linear taxes and subsidies to trade.

6 Other Motives for Trade

To this point, we have considered a model of the relationship between trade volumes and asset prices based on the idea that agents face the risk of uninsurable idiosyncratic shocks to their risk tolerance. As we have seen above, a negative idiosyncratic shock to an agent's risk tolerance leads that agent to sell shares of aggregate risk at $t = 1$ and to receive lower certainty equivalent consumption in equilibrium. In particular, we showed that when agents have HARA subutility functions $U_\tau(\cdot)$, their trading of shares of aggregate risk and their equilibrium certainty equivalent consumption are both linear and increasing in the difference between their risk tolerance and the risk tolerance of the representative agent (see equations 30 and 31). This result gave us very tractable formulas for both trade volumes at $t = 1$ and asset prices as of $t = 0$. This result is also central to the normative implications of our model for a Tobin tax: that such a tax has a first-order negative impact on welfare because it is borne by sellers of aggregate risk who are also experiencing a negative shock to their certainty equivalent consumption.

In this section, we consider the implications of two alternative specifications of our model for trade volumes and asset prices when agents' desire to trade is driven by shocks to their needs to hedge the risk in their non-traded endowments of consumption at $t = 2$ in addition to or in place of shocks to their risk tolerance. In particular, in both alternative specifications of our model, we assume that agents are identical at time $t = 0$, but that they are each endowed at $t = 1$ with an idiosyncratic claim to consumption at $t = 2$. We assume that this claim is non-traded, but that there are sufficient traded securities so that agents can trade these securities at $t = 1$ to hedge the risk in their non-traded endowment.

In the first alternative specification of our model, we retain the shocks to agents' risk tolerance and add risk in agents' non-traded endowments. We assume that this endowment risk is diversifiable risk. Thus, this risk can be hedged by offsetting trade of non-systematic securities (or an insurance contract) without paying a risk premium. We specify this non-traded diversifiable risk so that it results in no shock to agents' budget constraint for certainty equivalent consumption in equation (15). Hence, uncertainty over non-traded endowments in this case does not introduce risk over certainty equivalent consumption and thus does not affect asset prices at $t = 0$. But realizations of non-traded endowments induce agents to trade assets with non-systematic risk to hedge their endowment risk. The conclusion from this alternative specification of our model is that the positive and normative results regarding the relationship between trade volume and asset prices from our baseline model carry over as long as one focuses on trade in systematic risk, which we have termed "portfolio rebalancing," and not on trade in assets that are claims to diversifiable risk.

In the second alternative specification of our model, for simplicity, we assume that agents all have common preferences at $t = 1$. Instead of preference shocks, agents receive at $t = 1$ an idiosyncratic shock to their non-traded endowment to consumption at $t = 2$ that has exposure to aggregate risk and which has an impact on their budget constraint for certainty equivalent consumption at $t = 1$. Agents thus have two motives to trade aggregate risk and riskless assets at $t = 1$. One is to hedge their endowment of non-traded exposure to aggregate risk. The other is due to endogenous changes in their risk tolerance that arise with changes in their wealth at $t = 1$. Under the assumption that agents have HARA preferences, we develop simple extensions of our formulas for equilibrium trading volume and certainty equivalent consumption (equations 30 and 31). These extensions allow us to apply the asset pricing formulas we have developed in section 4. In contrast, the normative results regarding Tobin taxes do not directly apply because it is no longer the case that one can show that sellers of aggregate risk at $t = 1$ also experience lower certainty equivalent consumption.

6.1 Model with Idiosyncratic Non-Systematic Exposure Shocks

In this model there are K securities, with payoffs $d^k(y, z)$ at time $t = 2$. Agents receive at time $t = 1$ a random endowment of one of the securities. Each security pays $d^k(y, z) = y + \epsilon^k$ where ϵ^k is, conditionally on z, statistically independent of y. Moreover, we assume that conditional on (z, y)

$$\frac{1}{K}\sum_{k=1}^{K}\epsilon^k = 0 \tag{61}$$

with probability one, so each ϵ^k is a diversifiable risk. We let the $k = 0$ security be the market portfolio, i.e., $d^0(y, z) = y$. We assume that at time $t = 1$, conditional on the realization of z and τ, each agent draws an idiosyncratic random variable k denoting the security that the agent is assigned. Additionally, we assume that each agent is endowed the same amount $\bar{\alpha}$ of this security. In particular for each τ, we let $\varrho(k|\tau, z)$ be the probability that an agent with risk tolerance τ when the state is z will receive $\bar{\alpha}$ of the security $k \geq 1$. Additionally, each agent is assigned $1 - \bar{\alpha}$ of the market portfolio or security $k = 0$. We assume that at time $t = 1$, agents can only trade in bonds, and in the $k = 0, 1, \ldots, K$ securities.[8]

The equilibrium in this model is, essentially, the same as the one in the benchmark model. In particular, interest rates and risk premia for any security with systematic risk are exactly the same as in the benchmark model. This is because the price of any security $k \geq 0$ is the same as the the price of the market portfolio, i.e., $P_1(z; d^k) = P_1(z; d^0)$, since for the representative agent, $P_1(z; \epsilon^k) = 0$.[9] Second, at time $t = 1$ we can decompose the trade of agents in two parts,

[8] Alternatively, we can have a richer model where $\bar{\alpha}$ is a random variable. Such model has the same implications, so we avoid its extra notational complexity by assuming a degenerate distribution.

[9] Given this property of the prices, we could have equally assumed that the holdings are known as of time $t = 0$, but that the market for these transactions takes place at time $t = 1$.

since agents will hedge the non-systematic risk included in the $k \geq 1$ security assigned to them. They will do so because they strictly prefer not to bear the non-systematic risk, which can be avoided at no cost. What they trade against the $k \geq 1$ security is not uniquely determined. We assume that they sell $\bar{\alpha}(K-1)/K$ of it and buy $\bar{\alpha}/K$ of the remaining $K-1$ securities. In this way, given our assumption about $\{\epsilon^k\}_{k=1}^K$, agents will eliminate their idiosyncratic risk and effectively will end up with one unit of the market portfolio, as in our benchmark case. Additionally, depending on the realization of τ, agents will buy (or sell) additional units of the market portfolio against bonds, exactly as in the benchmark model. We refer to the second type of trade as "rebalancing trade."

Thus, agents trade in the $K+1$ securities has a factor structure, with all agents selling, for large K, most of the quantity $\bar{\alpha}$ of the security k that was assigned to them and buying $\bar{\alpha}/K$ of each of the other securities. The volume of rebalancing trade relative to the total trade volume depends on the values of $\bar{\alpha}$ and K. Hence, all the results of the benchmark model apply to this version when trade volume is interpreted as rebalancing trade volume. Yet, rebalancing trade accounts for a fraction of the total trade.

6.2 Model with Idiosyncratic Systematic Exposure Shocks

In this model, to simplify, we assume that there is no idiosyncratic shock to agents' risk tolerance, so that for each z as of $t=1$, agents use the same utility function to value $t=2$ payoffs. As in the benchmark case, we use HARA utility function, so $\tau = \bar{\tau}(z)$ with probability one. Instead, each agent receives a wealth shock equal to ω at time $t=1$. An agent with a wealth shock ω gets units of assets that are worth $\omega \bar{D}_1(z)$. Thus, we will index agents by ω as opposed to τ. Abusing notation, we let $\mu(\cdot|z)$ denote the conditional distribution of ω for each z. Agents are identical as of time $t=0$, i.e., they all face the same ω risk at time $t=1$. Since the shock is idiosyncratic, we have that $\sum_\omega \mu(\omega|z) = 0$ for all z. Using the properties of the HARA utility function, we have that $\bar{C}_1(z), \bar{p}(y;z), \bar{D}_1(z)$ depend only on the preferences of the representative agent and the distribution of the aggregate endowment at $t=2$, and hence are identical to those in our baseline model.

Using essentially, the same type of arguments as in the analysis of the baseline case, given

z at time $t = 1$, the conditionally efficient allocation of consumption at $t = 2$ is given by[10]

$$\frac{c(\omega, y; z)}{\gamma} + \bar{\tau}(z) = \hat{\phi}(\omega; z) \left(\frac{y}{\gamma} + \bar{\tau}(z) \right) \text{ for some } \sum_{\omega} \hat{\phi}(\omega; z) = 1. \tag{62}$$

Since the equilibrium with incomplete markets is conditionally efficient, the equilibrium allocation is linear in y with some exposure $\phi^e(\omega; z)$ determined by the wealth of an ω agent. Given the observation that an agent with realized type ω has to finance his or her purchases of securities with wealth $(1 + \omega)\bar{D}_1(z)$, we obtain the following expression for ϕ^e and the equilibrium certainty equivalent C_1^e of an ω agent as

$$\phi^e(\omega; z) = 1 + \omega\, \bar{D}_1(z)\, \frac{\bar{\mathcal{A}}(z)}{\gamma} \text{ and} \tag{63}$$

$$C_1^e(\omega; z) = \bar{C}_1(z) + \omega \bar{D}_1(z) \left[1 + \left(\bar{C}_1(z) - \bar{D}_1(z) \right) \frac{\bar{\mathcal{A}}(z)}{\gamma} \right]. \tag{64}$$

The effect on the certainty equivalence is clear: higher wealth increases certainty equivalent consumption in a linear fashion. How much the agent trades depends upon which security the non-traded endowment $\omega \bar{D}_1(z)$ resembles. Consider two different cases. In the first case, we assume that the agent receives a *non-traded endowment* that pays off $\omega \bar{D}_1(z)$ units of consumption at $t = 2$ with certainty. In this case, the agent hedges his or her exposure by trading in the market portfolio. Here, the rebalancing trade of the ω agent is

$$\phi^e(\omega; z) - 1 = \omega\, \bar{D}_1(z)\, \frac{\bar{\mathcal{A}}(z)}{\gamma}. \tag{65}$$

In this case, there is trade at $t = 1$ only because of the effect of wealth on risk tolerance. Note that in the CARA case (obtained as $\gamma \to \infty$), these shocks do not lead to trade. For finite γ, trade is driven by the fact that risk tolerance changes endogenously with wealth. For $\gamma > 0$ absolute risk tolerance increases with wealth, and hence the agent receiving a positive wealth shock ω will like to buy more shares of the market portfolio. The reverse is true if $\gamma < 0$.

In the second case, we assume that the shock consists of the agents receiving ω units of the *market portfolio*. In this case, the rebalancing trade of an ω agent is

$$\phi^e(\omega; z) - (1 + \omega) = -\, \omega\, \bar{\tau}(z)\, \frac{\bar{\mathcal{A}}(z)}{\gamma}. \tag{66}$$

Note that the sign of this rebalancing trade depends on the product of $\gamma \bar{\tau}(z)$, where each of the terms can be positive or negative. The difference between the first and the second case arises

[10]As before, we define conditionally efficient allocations as those maximizing the weighted average of the time $t = 1$ expected utility with weights $\hat{\lambda}(\omega; z)$.

because now there are two effects. As in the first case, there is trade due to impact of wealth on risk aversion, given by $\bar{\mathcal{A}}(z)/\gamma$. The other effect is the direct effect on desired trade that comes from the need to hedge the aggregate risk in the non-traded endowment. Importantly, in either of the two cases, the extent of rebalancing trade is proportional to the size of the wealth shock, although the expression and quantitative size depends on the case. Additionally, the effect on the certainty equivalence is exactly the same and monotone in ω.

From (64) it follows, as expected, than in equilibrium an agent with higher value of ω has higher certainty equivalent consumption, since this agent has higher wealth. Whether this agent will sell the extra shares ω is s more subtle question as explained above, and depends on whether the shock consists of a non-traded or a traded security. For instance, in the case in which the shock is in terms of labor income (i.e., a non-traded security), so agents hedge it with their rebalancing trade $\phi^e(\omega; z) - 1$, the rebalancing trade volume is

$$TV^e(z) \equiv \frac{1}{2} \sum_{\omega} |\phi^e(\omega; z) - 1| \, \mu(\omega; z) = \frac{\bar{D}_1(z)}{2} \frac{\bar{\mathcal{A}}(z)}{|\gamma|} \sum_{\omega} |\omega| \mu(\omega; z). \tag{67}$$

Continuing with the case of hedging labor income, and using (64) and (63), we can rewrite certainty equivalent consumption as a function of trade as in the benchmark model, whose expression we write next to it to facilitate the comparison:

$$C_1^e(\omega; z) = \bar{C}_1(z) + \left(\bar{C}_1(z) - \bar{D}_1(z) + \frac{\gamma}{\bar{\mathcal{A}}(z)} \right) [\phi^e(\omega; z) - 1]$$
$$C_1^e(\tau; z) = \bar{C}_1(z) + \big(\bar{C}_1(z) - \bar{D}_1(z) \big) \, [\phi^e(\tau; z) - 1] \, .$$

In both expressions, the last term in square brackets is the rebalancing trade of the agent with either ω or τ realization. For both specifications of our model, the difference between the certainty equivalent consumption of an individual agent and that for the representative agent is proportional to the rebalancing trade of that agent. This is the crucial aspect of the model, from which we obtain the same relationship between trade volume on asset prices as in the benchmark model. In particular, we obtain the decomposition of risk premium of Proposition 7, the comparative statics of trade volume and interest rates of Proposition 8, and the comparative statics of trade volume and risk premium of Propositions 9 and 10.

One difference between the benchmark case with shocks to risk tolerance and this version of the model is the sensitivity of certainty equivalent consumption to rebalancing trade. In the benchmark case with shocks to risk tolerance, it is $\bar{C}_1 - \bar{D}_1$, which is approximately given by a risk premium; see (32). Instead, in the hedging model, there is an extra term $\bar{C}_1 - \bar{D}_1 + \gamma/\bar{\mathcal{A}}$, which can be substantially larger given a positive value of γ and a negative value of $\bar{\tau}(z)$. Note that this configuration corresponds to the one-period version of the preferences in Campbell and Cochrane (1999). Indeed, in this case, we can let $V = U_{\bar{\tau}}$ and consider the case of expected utility.

References

Albuquerque, Rui, Eichenbaum Martin, Victor Xi Luo, and Sergio Rebelo. 2016. "Valuation risk and asset pricing." *Journal of Finance* 71 (6): 2861–2904. 7

Barro, Robert J., Jesús Fernández-Villaverde, Oren Levintal, and Andrew Mollerus. 2017. "Safe assets." 8

Bekaert, Geert, Eric Engstrom, and Steven R. Grenadier. 2010. "Stock and bond returns with moody investors." *Journal of Empirical Finance* 17 (5): 867–894. 8

Ben-Tal, Aharon and Marc Teboulle. 1986. "Expected utility, penalty functions, and duality in stochastic nonlinear programming." *Management Science* 32 (11): 1445–1466. 53

Bhandari, Anmol, Jaroslav Borovička, and Paul Ho. 2016. "Identifying ambiguity shocks in business cycle models using survey data." Working Paper 22225, National Bureau of Economic Research. 8

Caballe, Jordi and Alexey Pomansky. 1996. "Mixed risk aversion." *Journal of Economic Theory* 71 (2): 485–513. 41

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini. 2009. "Fight or flight? Portfolio rebalancing by individual investors." *Quarterly Journal of Economics* 124 (1): 301. 9

Campbell, John Y. and John H. Cochrane. 1999. "By force of habit: A consumption-based explanation of aggregate stock market behavior." *Journal of Political Economy* 107 (2): 205–251. 8, 47

Campbell, John Y., Sanford J. Grossman, and Jiang Wang. 1993. "Trading volume and serial correlation in stock returns." *Quarterly Journal of Economics* 108 (4): 905–939. 7

Cho, Sungjun. 2014. "What drives stochastic risk aversion?" *International Review of Financial Analysis* 34: 44–63. 8

Constantinides, George M. and Darrell Duffie. 1996. "Asset pricing with heterogeneous consumers." *Journal of Political Economy* 104 (2): 219–240. 1, 3, 17, 24, 32

Diamond, Peter A. 1998. "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates." *American Economic Review* 88 (1): 83–95. 74

Drechsler, Itamar. 2013. "Uncertainty, time-varying fear, and asset prices." *Journal of Finance* 68 (5): 1843–1889. 8

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen. 2005. "Over-the-counter markets." *Econometrica* 73 (6): 1815–1847. 10

Gordon, Stephen and Pascal St-Amour. 2004. "Asset returns and state-dependent risk preferences." *Journal of Business & Economic Statistics* 22 (3): 241–252. 7

Grant, Simon, Atsushi Kajii, Ben Polak, and Zvi Safra. 2010. "Generalized utilitarianism and Harsanyi's impartial observer theorem." *Econometrica* 78 (6): 1939–1971. 3

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2018. "Time varying risk aversion." *Journal of Financial Economics* 128 (3): 403–421. 7

Guo, Hui, Zijun Wang, and Jian Yang. 2013. "Time varying risk return tradeoff in the stock market." *Journal of Money, Credit and Banking* 45 (4): 623–650. 8

Kim, Kun Ho. 2014. "Counter-cyclical risk aversion." *Journal of Empirical Finance* 29: 384–401. 8

Kozak, Serhiy. 2015. "Dynamics of bond and stock returns." Working Paper 1277, Ross School of Business. 8

Kreps, David and Evan L Porteus. 1978. "Temporal resolution of uncertainty and dynamic choice theory." *Econometrica* 46 (1): 185–200. 2

Lenel, Moritz. 2017. "Safe assets, collateralized lending and monetary policy." Working paper 17-010, Stanford Institute for Economic Policy Research. 8

Lo, Andrew W and Jiang Wang. 2000. "Trading volume: definitions, data analysis, and implications of portfolio theory." *Review of Financial Studies* 13 (2): 257–300. 9

———. 2006. "Trading volume: Implications of an intertemporal capital asset pricing model." *Journal of Finance* 61 (6): 2805–2840. 9

Lucas, Robert E. Jr. 1980. "Equilibrium in a pure currency economy." *Economic Inquiry* 18 (2): 203–220. 7

Mankiw, N. Gregory. 1986. "The equity premium and the concentration of aggregate shocks." *Journal of Financial Economics* 17 (1): 211–219. 1, 3, 17, 24, 32

Mirrlees, James A. 1971. "An exploration in the theory of optimum income taxation." *Review of Economic Studies* 38 (2): 175–208. 74

Mongin, Philippe and Marcus Pivato. 2015. "Social preference and social welfare under risk and uncertainty." In *Oxford Handbook of Well-Being and Public Policy*, edited by Matthew Adler and Marc Fleurbaey. Oxford University Press. 3

Pratt, John W. and Richard J. Zeckhauser. 1987. "Proper risk aversion." *Econometrica* 55 (1): 143–154. 41

Santos, Tano and Pietro Veronesi. 2017. "Habits and leverage." Research Paper 16-220, University of Chicago Booth School of Business. 8

Seade, Jesus K. 1977. "On the shape of optimal tax schedules." *Journal of Public Economics* 7 (2): 203–235. 79

Selden, Larry. 1978. "A new representation of preferences over 'certain x uncertain' consumption pairs: The 'ordinal certainty equivalent' hypothesis." *Econometrica* 46 (5): 1045–1060. 2, 6

Steffensen, Mogens. 2011. "Optimal consumption and investment under time-varying relative risk aversion." *Journal of Economic Dynamics and Control* 35 (5): 659–667. 7

Uslu, Semih. 2015. "Pricing and liquidity in decentralized asset markets." Johns Hopkins Carey School of Business. 10

Vayanos, Dimitri and Jiang Wang. 2012. "Liquidity and asset returns under asymmetric information and imperfect competition." *Review of Financial Studies* 25 (5): 1339–1365. 9

———. 2013. "Market liquidity: Theory and empirical evidence." In *Handbook of the Economics of Finance*, edited by George Constantinides, Milton Harris, and Rene Stulz, chap. 19. North Holland. 9

Figure 1: Event tree for three-period model

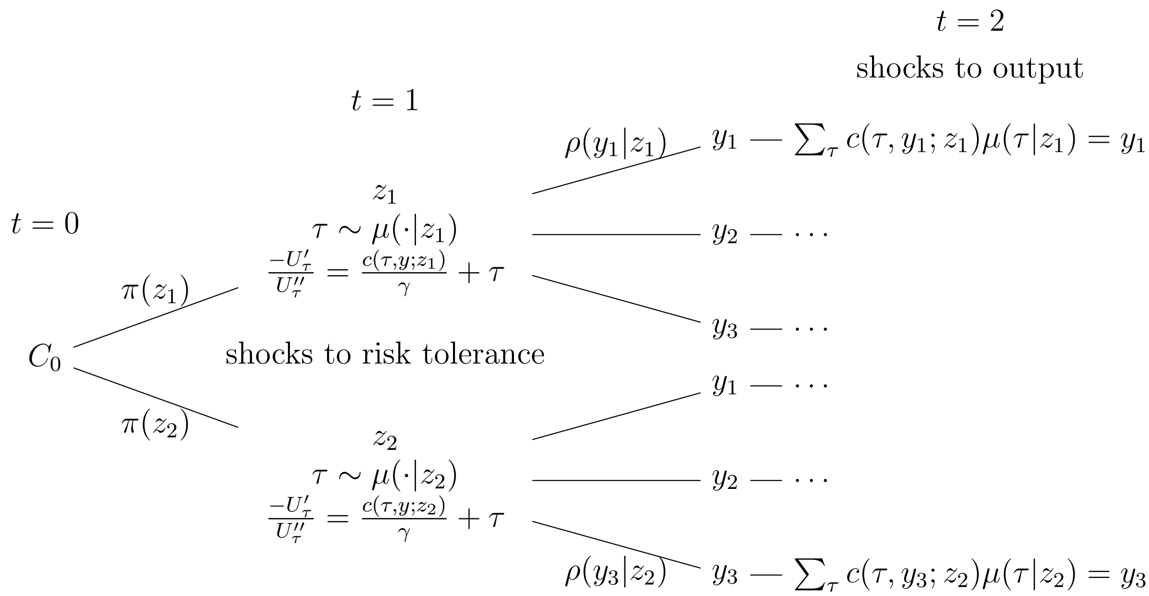


Figure for the case of two values for $z \in \{z_1, z_2\}$ and three values for $y \in \{y_1, y_2, y_3\}$.

Figure 2: Time line of three-period model

time $t = 0$	time $t = 1$	time $t = 2$
aggregate shocks:	$z \sim \pi(\cdot)$	$y \sim \rho(\cdot \vert z)$
idiosyncratic shocks:	$U_\tau(\cdot)$ w/risk tolerance shock $\tau \sim \mu(\cdot \vert z)$	
C_0	certainty equivalent(s): $\bar{C}(z), C^e(\tau; z)$	$c(\tau, y; z)$
price asset $P_0(d)$	rebalance portfolio, price assets $P_1(z; d)$	payoff $d(y; z)$

A Online Appendix

This appendix has two sections. In the first section, we present proofs of propositions. In the second section we present our results regarding a mechanism design approach to optimal non-linear Tobin taxes on asset trade.

B Proofs

Proof of Proposition 1. To simplify the notation of the proof, we omit the values of z and τ from all the expressions. Take two vectors $\vec{c}_a$ and $\vec{c}_b$ and a scalar $\theta \in (0,1)$. We want to show that

$$\theta U^{-1}\left(\sum_{y\in Y} U(c_a(u))\rho(y)\right) + (1-\theta)U^{-1}\left(\sum_{y\in Y} U(c_b(y))\rho(y)\right)$$
$$\leq U^{-1}\left(\sum_{y\in Y} U\left(\theta c_a(y) + (1-\theta)c_b(y)\right)\rho(y)\right)$$

if and only if U has concave risk tolerance.

Theorem 1 in Ben-Tal and Teboulle (1986) shows that for any vector x and conformable random vector W of the same dimension, the function $v(x) = U^{-1}\left(E\left[U(x^T W)\right]\right)$ is concave in x if and only if the risk tolerance of U is a concave function, where the expectation is taken with respect to the distribution of the random vector W. We use this result for the case with two-dimensional x and W where the two dimensions of the random vector W are perfectly correlated, taking as many values as the cardinality of Y, with probabilities $\rho(y)$. Moreover the realizations of each of the dimensions of W are as follows. The first dimension of W coincides with the values of $\vec{c}_a$ and the second with the values of $\vec{c}_b$. This means that $\Pr\{W = (c_a(y), c_b(y))\} = \rho(y)$ for all $y \in Y$. Moreover, take $x_a = (1,0)$ and $x_b = (0,1)$. We thus have

$$
\begin{aligned}
v(x_a) = v((1,0)) &= U^{-1}\left(E\left[U(x_a^T W)\right]\right) = U^{-1}\left(\sum_{y\in Y} U(c_a(y))\rho(y)\right)\\
v(x_b) = v((0,1)) &= U^{-1}\left(E\left[U(x_b^T W)\right]\right) = U^{-1}\left(\sum_{y\in Y} U(c_b(y))\rho(y)\right)\\
v(\theta x_a + (1-\theta)x_b) = v((\theta, 1-\theta)) &= U^{-1}\left(E\left[U((\theta x_b + (1-\theta)x_b)^T W)\right]\right)\\
&= U^{-1}\left(\sum_{y\in Y} U(\theta c_b(y) + (1-\theta)c_b(y))\rho(y)\right).
\end{aligned}
$$

Hence if v is concave, so is $\mathcal{C}_1$. Likewise, if $\mathcal{C}_1$ is concave, so must be v. □

Proof of Proposition 2. Given that agents have the same beliefs ρ and that they use expected utility, an allocation is conditionally efficient if and only if it maximizes the following objective function for each y:

$$
\sum_{\tau} \hat{\lambda}_\tau U_\tau\left(c(\tau,y;z)\right)\rho(y|z)\mu(\tau|z) + \hat{p}(y;z)\sum_{\tau}\left(y - c(\tau;y;z)\right)\mu(\tau|z),
$$

where $p(y|z)$ is the multiplier of the feasibility constraint for each y. The first-order conditions of this problem are

$$
U'_\tau\left(c(\tau,y;z)\right) = \left(\frac{c(\tau,y;z)}{\gamma} + \tau\right)^{-\gamma} = \frac{\hat{p}(y;z)}{\rho(y|z)}\frac{1}{\lambda_\tau},
$$

where the first equality uses the assumption that utility is of the equicautious HARA class. We can rewrite this expression as

$$
\frac{c(\tau,y;z)}{\gamma} + \tau = \left(\frac{p(y;z)}{\rho(y|z)}\right)^{-\frac{1}{\gamma}} (\lambda_\tau)^{\frac{1}{\gamma}}.
$$

Multiplying this expression by $\mu(\tau|z)$, adding across τ, and using feasibility, we obtain

$$
\frac{y}{\gamma} + \bar{\tau} = \left(\frac{\hat{p}(y;z)}{\rho(y|z)}\right)^{-\frac{1}{\gamma}} \sum_{\tau} (\lambda_\tau)^{\frac{1}{\gamma}} \mu(\tau|z),
$$

or

$$
U'_{\bar{\tau}(z)}(y)\,\rho(y|z) = \left(\frac{y}{\gamma} + \bar{\tau}\right)^{-\gamma}\rho(y|z) = \hat{p}(y;z)\left[\sum_{\tau}(\lambda_\tau)^{\frac{1}{\gamma}}\mu(\tau|z)\right]^{-\gamma},
$$

which gives the expression for the Lagrange multipliers. Using again the first-order conditions of each agent, her consumption is

$$
\lambda_\tau\left(\frac{c(\tau,y;z)}{\gamma} + \tau\right)^{-\gamma}\rho(y|z) = \hat{p}(y;z) = \left(\frac{y+\bar{\tau}}{\gamma}\right)^{-\gamma}\rho(y|z)\left[\sum_{\tau}(\lambda_\tau)^{\frac{1}{\gamma}}\mu(\tau|z)\right]^{\gamma}
$$

or

$$\left(\frac{c(\tau,y;z)}{\gamma}+\tau\right)=\left(\frac{y}{\gamma}+\bar{\tau}\right)\frac{(\lambda_\tau)^{\frac{1}{\gamma}}}{\sum_{\tau'}(\lambda_{\tau'})^{\frac{1}{\gamma}}\mu(\tau|z)}.$$

Defining

$$\hat{\phi}(\tau;z)\equiv\frac{(\lambda_\tau)^{\frac{1}{\gamma}}}{\sum_{\tau'}(\lambda_{\tau'})^{\frac{1}{\gamma}}\mu(\tau|z)},$$

we can write our first desired result:

$$\left(\frac{c(\tau,y;z)}{\gamma}+\tau\right)=\hat{\phi}(\tau;z)\left(\frac{y}{\gamma}+\bar{\tau}\right).$$

Then raising both sides to $(1-\gamma)$ and multiplying them by $\frac{\gamma}{1-\gamma}\rho(y|z)$, we get

$$\left(\frac{\gamma}{1-\gamma}\right)\left(\frac{c(\tau,y;z)}{\gamma}+\tau\right)^{1-\gamma}\rho(y|z)=\hat{\phi}(\tau;z)^{1-\gamma}\left(\frac{\gamma}{1-\gamma}\right)\left(\frac{y}{\gamma}+\bar{\tau}\right)^{1-\gamma}\rho(y|z).$$

Using that this proportionality must hold for all y and adding across all y, we have

$$\sum_{y\in Y}\left(\frac{\gamma}{1-\gamma}\right)\left(\frac{c(\tau,y;z)}{\gamma}+\tau\right)^{1-\gamma}\rho(y|z)$$
$$=\hat{\phi}(\tau;z)^{1-\gamma}\sum_{y\in Y}\left(\frac{\gamma}{1-\gamma}\right)\left(\frac{y}{\gamma}+\bar{\tau}\right)^{1-\gamma}\rho(y|z).$$

Recall that $U_\tau^{-1}(u)=\gamma\left(\left[\left(\frac{1-\gamma}{\gamma}\right)u\right]^{\frac{1}{1-\gamma}}-\tau\right)$. So multiply each side by $(1-\gamma)/\gamma$ and raise the resulting expressions to the $(1/(1-\gamma))$, and subtract τ from both sides and multiply by γ to obtain

$$\gamma\left(\left[\sum_{y\in Y}\left(\frac{c(\tau,y;z)}{\gamma}+\tau\right)^{1-\gamma}\rho(y|z)\right]^{\frac{1}{1-\gamma}}-\tau\right)=\gamma\left(\hat{\phi}(\tau;z)\left[\sum_{y\in Y}\left(\frac{y}{\gamma}+\bar{\tau}\right)^{1-\gamma}\rho(y|z)\right]^{\frac{1}{1-\gamma}}-\tau\right).$$

Use the definition of $C_1(\tau;z)$ and multiply both sides by $\mu(\tau|z)$ and add across $\tau's$ to obtain

$$\sum_\tau C_1(\tau;z)\mu(\tau|z)=\gamma\left(\sum_\tau\hat{\phi}(\tau;z)\left[\sum_{y\in Y}\left(\frac{y}{\gamma}+\bar{\tau}\right)^{1-\gamma}\rho(y|z)\right]^{\frac{1}{1-\gamma}}\mu(\tau|z)-\sum_\tau\tau\mu(\tau|z)\right).$$

Using that $\sum_\tau\hat{\phi}(\tau;z)\mu(\tau|z)=1$ and the definition of $\bar{\tau}$, we get

$$\sum_\tau C_1(\tau;z)\mu(\tau|z)=\gamma\left(\left[\sum_{y\in Y}\left(\frac{y}{\gamma}+\bar{\tau}\right)^{1-\gamma}\rho(y|z)\right]^{\frac{1}{1-\gamma}}-\bar{\tau}\right)=\bar{C}_1(z),$$

which establishes the last desired result. □

Proof of Proposition 3. In an incomplete market equilibrium, all agents have the same wealth at time $t = 1$, since they are identical as of time $t = 0$, and hence the budget constraint is

$$\bar{D}_1(z) \equiv \sum_y y\bar{p}(y; z) = \sum_y c(\tau, y; z)\bar{p}(y; z).$$

Using that the allocation is conditionally efficient, we have $c(\tau, y; z)/\gamma + \tau = \hat{\phi}(\tau; z)\,(y/\gamma + \bar{\tau})$. Replacing consumption in the budget constraint then gives

$$\bar{D}_1(z) = \sum_y \left[\hat{\phi}(\tau; z)\,(y + \gamma\bar{\tau}) - \gamma\tau\right]\bar{p}(y; z) = \phi(\tau; z)\bar{D}_1(z) + \gamma\left(\hat{\phi}(\tau; z)\bar{\tau} - \tau\right),$$

where we have used the definition of $\bar{D}_1(z)$ again, as well as the normalization of prices. Solving for $\hat{\phi}$ gives

$$\hat{\phi}(\tau; z) = \frac{\bar{D}_1(z)/\gamma + \tau}{\bar{D}_1(z)/\gamma + \bar{\tau}} \text{ or } \hat{\phi}(\tau; z) - 1 = \frac{\tau - \bar{\tau}}{\bar{D}_1(z)/\gamma + \bar{\tau}}.$$

Setting $\phi^e(\tau; z) = \hat{\phi}(\tau; z)$ we obtain the desired result for ϕ^e.

Using the equality $c(\tau, y; z)/\gamma + \tau = \hat{\phi}(\tau; z)\,(y/\gamma + \bar{\tau})$ with $\hat{\phi}(\tau; z) = \frac{\bar{D}_1(z)/\gamma+\tau}{\bar{D}_1(z)/\gamma+\bar{\tau}}$ into the definition of certainty equivalent consumption, we obtain the expression in equation (31).

We omit the subindex z in several of the expressions to simplify the notation. For the same reason, we omit the index $\bar{\tau}(z)$ for the function $U_{\bar{\tau}(z)}(\cdot)$. First we show that $\bar{C}_1 > \bar{D}_1$. To show this, we construct the function $F(\theta) = \mathbb{E}\left[U\left((1-\theta)y + \theta\bar{C}_1\right)\right]$. Note that F is strictly concave in θ since it is the expected value of the composition of a strictly concave function with a linear function. In particular, $F''(\theta) = \mathbb{E}\left[U''\left((1-\theta)y + \theta\bar{C}_1\right)\left(\bar{C}_1 - y\right)^2\right] < 0$. Direct computation gives $F(0) = \mathbb{E}[U(y)]$ and $F(1) = U\left(\bar{C}_1\right)$, and thus by definition of $\bar{C}_1$ we have $F(0) = F(1)$. Summarizing, F is a strictly concave function that attains the same value at $\theta = 0$ and $\theta = 1$, and hence the maximum of F is attained in $\theta \in (0, 1)$. Thus, F must be strictly increasing at $\theta = 0$, or $F'(0) > 0$. Direct computation gives $F'(0) = \mathbb{E}\left[U'(y)\left(\bar{C}_1 - y\right)\right]$ and using $F'(0) > 0$ we have $\bar{C}_1\,\mathbb{E}[U'(y)] > \mathbb{E}[U'(y)\,y]$, which after rearranging and using the definition of $\bar{D}_1$, gives the desired result.

Now we find the expression for the first order expansions. The one for $\bar{C}_1$ can be obtained by differentiating $\bar{C}_1$ with respect to σ_y^2. We begin by expanding $E[U'(y)] = U'(\bar{y}) + (1/2)U'''(\bar{y})\sigma_y^2 + o(\sigma_y^2)$ where $o(\sigma_y^2)$ denotes an expression of order smaller than σ_y^2. And $E[U'(y)y] = U'(\bar{y})\bar{y} + (1/2)[U'''(\bar{y})\bar{y} + 2U''(\bar{y})]\sigma_y^2 + o(\sigma_y^2)$. Using these expansions, $\bar{D}_1$ is given by

$$\bar{D}_1(\sigma_y^2) = \frac{U'(\bar{y})\bar{y} + (1/2)[U'''(\bar{y})\bar{y} + 2U''(\bar{y})]\sigma_y^2 + o(\sigma_y^2)}{U'(\bar{y}) + (1/2)U'''(\bar{y})\sigma_y^2 + o(\sigma_y^2)}.$$

Differentiating this expression with respect to σ_y^2 and evaluating at $\sigma_y^2 = 0$, we obtain

$$\bar{D}_1'(0) = \frac{U''(\bar{y})U'(\bar{y})}{U'(\bar{y})^2} = \frac{U''(\bar{y})}{U'(\bar{y})}.$$

Thus, a Taylor expansion of $\bar{D}_1$ gives $\bar{D}_1 = \bar{y} + U''(\bar{y})/U'(\bar{y})\sigma_y^2 + o(\sigma_y^2)$.

Finally, we can evaluate $U''(c)/U'(c)$ at either $c = \bar{y}$ or in $c = \bar{y} + a\sigma_y^2$ for any differentiable function of a of σ_y^2. Thus, we can evaluate the desired expressions at either $\bar{y}$ or $\bar{y} - \bar{\mathcal{A}}\sigma_y^2$, and the difference is of an order smaller than σ_y^2. □

Proof of Proposition 4. We can rewrite equation (31) for the equilibrium certainty equivalent consumption as

$$\begin{aligned} C_1^e(\tau; z) &= \bar{C}_1(z)\left[\frac{\frac{\bar{D}_1(z)}{\gamma} + \tau}{\frac{\bar{D}_1(z)}{\gamma} + \bar{\tau}(z)}\right] - (\tau - \bar{\tau}(z))\left(\frac{\bar{D}_1(z)}{\frac{\bar{D}_1(z)}{\gamma} + \bar{\tau}(z)}\right) \\ &= \bar{C}_1(z)\phi^e(\tau; z) - (\tau - \bar{\tau}(z))\left(\frac{\bar{D}_1(z)}{\frac{\bar{D}_1(z)}{\gamma} + \bar{\tau}(z)}\right), \end{aligned}$$

where the second equation uses the expression for ϕ^e. Dividing by γ and adding τ to both sides, we obtain

$$\begin{aligned} \frac{C_1^e(\tau; z)}{\gamma} + \tau &= \frac{\bar{C}_1(z)}{\gamma}\phi^e(\tau; z) - (\tau - \bar{\tau}(z))\left(\frac{\frac{\bar{D}_1(z)}{\gamma}}{\frac{\bar{D}_1(z)}{\gamma} + \bar{\tau}(z)}\right) + \tau \\ &= \phi^e(\tau; z)\left(\frac{\bar{C}_1(z)}{\gamma} + \bar{\tau}(z)\right). \end{aligned}$$

From here we get

$$U_\tau'(C_1^e(\tau; z)) = (\phi^e(\tau; z))^{-\gamma} U_{\bar{\tau}(z)}'(\bar{C}_1(z)).$$

Likewise, we can use the relationship between consumption in a conditionally efficient allocation in (25) and the weights corresponding to the incomplete market equilibrium $\hat{\phi}(\tau; z) = \phi^e(\tau; z)$ in (30) to obtain

$$\frac{c^e(\tau, y; z)}{\gamma} + \tau = \phi^e(\tau; z)\left(\frac{y}{\gamma} + \tau\right).$$

Then, raising both sides to $-\gamma$, multiplying by $\rho(y|z)$, and adding across y, we obtain

$$\sum_y U_\tau'\left(c^e\left(\tau, y; z\right)\right)\rho(y|z) = (\phi^e(\tau; z))^{-\gamma}\sum_y U_{\bar{\tau}(z)}'\left(y\right)\rho(y|z).$$

Taking the ratio of the two expressions for the marginal utility, we cancel the factor $(\phi^e(\tau; z))^{-\gamma}$ and obtain the desired result. □

Proof of Proposition 9. By the assumption that d has constant sensitivity to z in both economies, i.e., that $\mathbb{E}(d|z)/\mathbb{E}(d) = 1$ is independent of z, we have that, since in the first economy $L(z) = 1$ for all z, using (43),

$$\frac{1}{\mathcal{E}_{0,2}(d)} = \sum_z \frac{Q^*(z)\pi(z')}{\sum_{z'} Q^*(z')\pi(z')} \left[\frac{1}{\mathcal{E}_{1,2}(z;d)}\right],$$

and for the second economy we have

$$\frac{1}{\mathcal{E}_{0,2}(d)} = \sum_z \frac{Q^*(z)L_2(z)\pi(z')}{\sum_{z'} Q^*(z')L_2(z')\pi(z')} \left[\frac{1}{\mathcal{E}_{1,2}(z;d)}\right].$$

For both economies, the terms $\mathcal{E}_{1,2}(z;d) \geq 1$ are the same and equal to the one for a representative agent economy. Using the assumption that the cash flow d has systematic exposure, these excess returns are increasing in marketwide risk aversion, and hence decreasing in z, so that its reciprocal i $1/\mathcal{E}_{1,2}(z;d)$ is increasing in z. By assumption $L_2(z) \geq 1$ and decreasing in z. Thus, the induced distribution $Q^*(z)L_2(z)\pi(z')/[\sum_{z'} Q^*(z')L_2(z')\pi(z')]$ for the second economy is stochastically lower than the distribution $Q^*(z)\pi(z')/[\sum_{z'} Q^*(z')\pi(z')]$ corresponding to the first economy. Hence, $1/\mathcal{E}_{0,2}(d)$, is smaller for the second economy than for the first economy, and thus its reciprocal, $\mathcal{E}_{0,2}(d)$, is higher for the second economy than for the first economy. $\square$

Proof of Proposition 10. Direct computation gives $\mathbb{E}_1[\tilde{d}|z]/\mathbb{E}_0[\tilde{d}] = \tilde{e}(z)/\mathbb{E}_0[\tilde{e}(z)]$ and $\mathbb{E}_1[d|z]/\mathbb{E}_0[d] = e(z)/\mathbb{E}_0[e(z)]$. Using the specification for d and $\tilde{d}$ and that $\rho(y|z) = \bar{\rho}(y)$, we have

$$\mathcal{E}_{1,2}(z;\tilde{d}) = \frac{\tilde{e}(z)\sum_y \delta(y)\bar{\rho}(y)}{\tilde{e}(z)\sum_y \bar{p}(y)\delta(y)\bar{\rho}(y)} = \frac{e(z)\sum_y \delta(y)\bar{\rho}(y)}{e(z)\sum_y \bar{p}(y)\delta(y)\bar{\rho}(y)} = \mathcal{E}_{1,2}(z;d) \equiv \bar{\mathcal{E}}_{1,2}.$$

Using that $\rho(y|z) = \bar{\rho}(y)$, we have that $Q^*(z) = \bar{Q}^*$. Thus, we have, using (43),

$$\begin{aligned}
\frac{1}{\mathcal{E}_{0,2}(\tilde{d})} &= \frac{1}{\bar{\mathcal{E}}_{1,2}} \sum_z \frac{\bar{Q}^* L(z)\pi(z)}{\sum_{z'} \bar{Q}^* L(z')\pi(z')} \frac{\tilde{e}(z)}{\mathbb{E}_0\left[\tilde{e}(z)\right]} = \frac{1}{\bar{\mathcal{E}}_{1,2}} \sum_z \frac{L(z)\pi(z)}{\sum_{z'} L(z')\pi(z')} \frac{\tilde{e}(z)}{\mathbb{E}_0\left[\tilde{e}(z)\right]} \\
&= \frac{1}{\bar{\mathcal{E}}_{1,2}} \mathbb{E}_0\left[\frac{L(z)}{\mathbb{E}_0[L(z)]} \frac{\tilde{e}(z)}{\mathbb{E}_0\left[\tilde{e}(z)\right]}\right] \\
&= \frac{1}{\bar{\mathcal{E}}_{1,2}} \left\{1 + \mathbb{C}ov_0\left[\frac{L(z)}{\mathbb{E}_0[L(z)]}, \frac{\tilde{e}(z)}{\mathbb{E}_0[\tilde{e}(z)]}\right]\right\}.
\end{aligned}$$

Likewise:

$$\frac{1}{\mathcal{E}_{0,2}(d)} = \frac{1}{\bar{\mathcal{E}}_{1,2}} \left\{1 + \mathbb{C}ov_0\left[\frac{L(z)}{\mathbb{E}_0[L(z)]}, \frac{e(z)}{\mathbb{E}_0[e(z)]}\right]\right\}.$$

Since, by assumption, $\tilde{e}(z)/e(z)$ decreases with z, then

$$\mathbb{C}ov_0\left[\frac{L(z)}{\mathbb{E}_0[L(z)]}, \frac{e(z)}{\mathbb{E}_0[e(z)]}\right] \leq \mathbb{C}ov_0\left[\frac{L(z)}{\mathbb{E}_0[L(z)]}, \frac{\tilde{e}(z)}{\mathbb{E}_0[\tilde{e}(z)]}\right].$$

Then, $\frac{1}{\mathcal{E}_{0,2}(\tilde{d})} \geq \frac{1}{\mathcal{E}_{0,2}(d)}$ or $\mathcal{E}_{0,2}(\tilde{d}) \leq \mathcal{E}_{0,2}(d)$. □

Proof of Proposition 11. To render the notation manageable, we suppress the z index for all variables and let $D(\omega) = \bar{D}_1(z;\omega)$. Under the assumption of no marginal investors, for a small tax ω, then $S(\tau) > 1$ if $\tau > \bar{\tau}$ and $S(\tau) < 1$ if $\tau < \bar{\tau}$. We can differentiate market clearing to obtain

$$0 = \sum_{S(\tau)>1} \left\{ \frac{\partial S(\tau, D, 0)}{\partial D} \left[\frac{dD(0)}{d\omega} + 1 \right] + \frac{\partial S(\tau, D, 0)}{\partial T} TV \right\} \mu(\tau)$$
$$+ \sum_{\tau<\bar{\tau}} \left\{ \frac{\partial S(\tau, D, 0)}{\partial D} \frac{dD(0)}{d\omega} + \frac{\partial S(\tau, D, 0)}{\partial T} TV \right\} \mu(\tau).$$

Rearranging, we have

$$0 = \sum_{S(\tau)>1} \frac{\partial S(\tau, D, 0)}{\partial D} \mu(\tau) + \frac{dD(0)}{d\omega} \sum_{\tau} \frac{\partial S(\tau, D, 0)}{\partial D} \mu(\tau) + TV \sum_{\tau} \frac{\partial S(\tau, D, 0)}{\partial T} \mu(\tau)$$

or

$$\frac{dD(0)}{d\omega} = \frac{\sum_{S(\tau)>1} \frac{\partial S(\tau,D,0)}{\partial D} \mu(\tau)}{-\sum_{\tau} \frac{\partial S(\tau,D,0)}{\partial D} \mu(\tau)} + TV \frac{\sum_{\tau} \frac{\partial S(\tau,D,0)}{\partial T} \mu(\tau)}{-\sum_{\tau} \frac{\partial S(\tau,D,0)}{\partial D} \mu(\tau)}.$$

Using the characterization of the partial derivative of $S(\tau)$ with respect to D in the previous lemma, we have

$$\sum_{\tau} \frac{\partial S(\tau, D, 0)}{\partial D} \mu(\tau) = \sum_{\tau} \phi(\tau) \mu(\tau) \frac{\mathbb{E}\left[U'_{\bar{\tau}}(y)\right]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]} + \sum_{\tau} (S(\tau)-1)\mu(\tau) \frac{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)\right]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}$$
$$= \frac{\mathbb{E}\left[U'_{\bar{\tau}}(y)\right]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}.$$

Likewise, using the partial derivative of $S(\tau)$ with respect to T in the previous lemma, we have

$$\sum_{\tau} \frac{\partial S(\tau, D, 0)}{\partial T} \mu(\tau) = -\frac{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)\right]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]},$$

and finally,

$$\sum_{S(\tau)>1} \frac{\partial S(\tau, D, 0)}{\partial D} \mu(\tau) = \sum_{S(\tau)>1} \phi(\tau) \mu(\tau) \frac{\mathbb{E}\left[U'_{\bar{\tau}}(y)\right]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}$$
$$+ \sum_{S(\tau)>1} (S(\tau)-1)\mu(\tau) \frac{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)\right]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]},$$

or using the expression for TV, we have

$$\sum_{S(\tau)>1} \frac{\partial S(\tau, D, 0)}{\partial D} \mu(\tau) = \frac{\mathbb{E}\left[U'_{\bar{\tau}}(y)\right]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]} \left(\sum_{S(\tau)>1} \phi(\tau) \mu(\tau) \right) + TV \frac{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)\right]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}.$$

Thus, we have

$$\begin{aligned}
\frac{dD(0)}{d\omega} &= \frac{\frac{\mathbb{E}[U'_{\bar{\tau}}(y)]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}\left(\sum_{S(\tau)>1}\phi(\tau)\mu(\tau)\right) + TV\frac{\mathbb{E}[U''_{\bar{\tau}}(y)(y-D)]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}}{-\frac{\mathbb{E}[U'_{\bar{\tau}}(y)]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}} - TV\frac{\frac{\mathbb{E}[U''_{\bar{\tau}}(y)(y-D)]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}}{-\frac{\mathbb{E}[U'_{\bar{\tau}}(y)]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}} \\
&= \frac{\frac{\mathbb{E}[U'_{\bar{\tau}}(y)]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}\left(\sum_{S(\tau)>1}\phi(\tau)\mu(\tau)\right)}{-\frac{\mathbb{E}[U'_{\bar{\tau}}(y)]}{\mathbb{E}\left[U''_{\bar{\tau}}(y)(y-D)^2\right]}} \\
&= -\sum_{S(\tau)>1}\phi(\tau)\mu(\tau) = -\sum_{\tau>\bar{\tau}}\phi(\tau)\mu(\tau) = -\sum_{\phi(\tau)>1}\phi(\tau)\mu(\tau) \in (-1\,,\,0)
\end{aligned}$$

since $\sum_\tau \phi(\tau)\mu(\tau) = 1$ and $\phi(\tau) \geq 0, \mu(\tau) \geq 0$ for all τ. $\square$

Proof of Proposition 13. Again we omit z to render the notation simpler. We want to compute

$$\frac{d}{d\omega}C_1^e(\tau_1; z) = \frac{\mathbb{E}[U'_{\bar{\tau}}(y)]}{U_{\bar{\tau}}(\bar{C})}\left[-(\phi(\tau_1)-1)\frac{\partial}{\partial\omega}\bar{D}(0) + TV\right].$$

We have from the previous proposition:

$$\frac{\partial}{\partial\omega}\bar{D}(0) = -\mu(\tau_2)\phi(\tau_2) \text{ and } TV = (\phi(\tau_2)-1)\mu(\tau_2),$$

thus,

$$\begin{aligned}
\frac{d}{d\omega}C_1^e(\tau_1; z) &= \frac{\mathbb{E}[U'_{\bar{\tau}}(y)]}{U_{\bar{\tau}}(\bar{C})}\left[(\phi(\tau_1)-1)\mu(\tau_2)\phi(\tau_2) + (\phi(\tau_2)-1)\mu(\tau_2)\right] \\
&= \frac{\mathbb{E}[U'_{\bar{\tau}}(y)]}{U_{\bar{\tau}}(\bar{C})}\mu(\tau_2)\left[\phi(\tau_1)\phi(\tau_2)-1\right] \\
&= \frac{\mathbb{E}[U'_{\bar{\tau}}(y)]}{U_{\bar{\tau}}(\bar{C})}\phi(\tau_1)(1-\phi(\tau_1))\left[1-\mu(\tau_2)\frac{1+\phi(\tau_1)}{\phi(\tau_1)}\right],
\end{aligned}$$

so

$$\frac{d}{d\omega}C_1^e(\tau_1; z) = \frac{\mathbb{E}[U'_{\bar{\tau}}(y)]}{U_{\bar{\tau}}(\bar{C})}\phi(\tau_1)(1-\phi(\tau_1))\left[1-\mu(\tau_2)\frac{1+\phi(\tau_1)}{\phi(\tau_1)}\right],$$

and thus,

$$\begin{aligned}
\frac{dW^e}{d\omega} &= \beta\sum_z \pi(z)\left[V'(C_1^e(\tau_1; z)) - V'(C_1^e(\tau_2; z))\right]\frac{d}{d\omega}C_1^e(\tau_1; z)\mu(\tau_1; z) \text{ thus} \\
\frac{dW^e(0)}{d\omega} &= \beta\sum_z \pi(z)\left[V'(C_1^e(\tau_1; z)) - V'(C_1^e(\tau_2; z))\right] \times \\
&\quad \frac{\mathbb{E}[U'_{\bar{\tau}}(y)|z]}{U_{\bar{\tau}}(\bar{C}_1(z))}\phi(\tau_1; z)(1-\phi(\tau_1; z))\left[1-\mu(\tau_2; z)\frac{1+\phi(\tau_1; z)}{\phi(\tau_1; z)}\right]\mu(\tau_1; z).
\end{aligned}$$

Since $V'' < 0$ and $C_1^e(\tau_1; z) < C_1^e(\tau_2; z)$, then

$$\frac{dW^e(0;z)}{d\omega} < 0 \iff \mu(\tau_2; z) > \frac{\phi(\tau_1; z)}{1 + \phi(\tau_1; z)}. \tag{68}$$

□

Proof of Proposition 14. Again we omit z to render the notation easier to follow. Recall that in an equal wealth equilibrium

$$\begin{aligned}
C_1^e(\tau) - \bar{C}_1(0) &= (\tau - \bar{\tau})\frac{\bar{C}_1 - \bar{D}_1}{\bar{\tau} + \bar{D}_1/\gamma} \equiv \chi(\tau - \bar{\tau}) \\
\phi^e(\tau) - 1 &= \frac{\tau - \bar{\tau}}{\bar{\tau} + \bar{D}_1/\gamma} \equiv \eta(\tau - \bar{\tau}) \\
TV^e &= \int_{\bar{\tau}}^{\tau_H} (\phi^e(\tau) - 1)\mu(\tau)d\tau = \eta \int_{\bar{\tau}}^{\tau_H} (\tau - \bar{\tau})\mu(\tau)d\tau \\
\bar{D}_1'(0) &= -\int_{\bar{\tau}}^{\tau_H} \phi^e(\tau)\mu(\tau)d\tau = -\left[\eta \int_{\bar{\tau}}^{\tau_H} (\tau - \bar{\tau})\mu(\tau)d\tau + \int_{\bar{\tau}}^{\tau_H} \mu(\tau)d\tau\right],
\end{aligned}$$

where $\eta \equiv 1/(\bar{\tau} + \bar{D}_1/\gamma)$ and $\chi \equiv (\bar{C}_1 - \bar{D}_1)/(\bar{\tau} + \bar{D}_1/\gamma)$. Also, since we assume that V is analytical, we can write

$$V'(c) = \sum_{n=0}^{\infty} \frac{V^{n+1}(\bar{C}_1)}{n!} \left(c - \bar{C}_1\right)^n \text{ for any } c,$$

where $V^{n+1}\left(\bar{C}_1\right) \equiv \partial^n V\left(\bar{C}_1\right)/\partial c^n$.

Using the expression (54), the ex-ante change on welfare of a small Tobin tax can be written as

$$\begin{aligned}
\frac{d}{d\omega}W^e(0;z) &= J(z)TV^e \int_{\tau_L}^{\tau_H} V'(C_1^e(\tau))\mu(\tau)d\tau \\
&- J(z)\bar{D}_1'(0) \int_{\tau_L}^{\tau_H} V'(C_1^e(\tau))\left(\phi^e(\tau) - 1\right)\mu(\tau)d\tau \\
&- J(z) \int_{\bar{\tau}}^{\tau_H} V'(C_1^e(\tau))\left(\phi^e(\tau) - 1\right)\mu(\tau)d\tau \\
&= J(z)\eta\left[\int_{\bar{\tau}}^{\tau_H} (\tau - \bar{\tau})\mu(\tau)d\tau\right]\left[\sum_{n=0}^{\infty} \chi^n \frac{V^{n+1}(\bar{C}_1)}{n!} \int_{\tau_L}^{\tau_H} (\tau - \bar{\tau})^n \mu(\tau)d\tau\right] \\
&- J(z)\bar{D}_1'(0)\,\eta\left[\sum_{n=0}^{\infty} \chi^n \frac{V^{n+1}(\bar{C}_1)}{n!} \int_{\tau_L}^{\tau_H} (\tau - \bar{\tau})^{n+1}\mu(\tau)d\tau\right] \\
&- J(z)\eta\left[\sum_{n=0}^{\infty} \chi^n \frac{V^{n+1}(\bar{C}_1)}{n!} \int_{\bar{\tau}}^{\tau_H} (\tau - \bar{\tau})^{n+1}\mu(\tau)d\tau\right].
\end{aligned}$$

We can rewrite this expression as

$$\frac{d}{d\omega}W^{e}(0;z)=J(z)\eta\chi^{n}\frac{V^{n+1}(\bar{C}_{1})}{n!}\times$$
$$\sum_{n=0}^{\infty}\left\{\left[\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})\mu(\tau)d\tau\right]\left[\int_{\tau_L}^{\tau_H}(\tau-\bar{\tau})^{n}\mu(\tau)d\tau\right]-\bar{D}_1'(0)\left[\int_{\tau_L}^{\tau_H}(\tau-\bar{\tau})^{n+1}\mu(\tau)d\tau\right]\right.$$
$$\left.-\left[\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})^{n+1}\mu(\tau)d\tau\right]\right\}.$$

Now we analyze the term for each of the derivatives $V(\bar{C})$. For $n=0,2,4$, we obtain

$$0=\eta V^{1}(\bar{C})\left\{\left[\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})\mu(\tau)d\tau\right]-\bar{D}_1'(0)\times 0-\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})\mu(\tau)d\tau\right\}.$$

We split the contribution of the remaining term into those with even and odd order of the derivatives of V. Using the symmetry of μ for these values of n, we have

$$\left[\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})\mu(\tau)d\tau\right]\left[\int_{\tau_L}^{\tau_H}(\tau-\bar{\tau})^{n}\mu(\tau)d\tau\right]-\bar{D}_1'(0)\times 0-\left[\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})^{n+1}\mu(\tau)d\tau\right]$$
$$=\frac{1}{2}\left[\int_{\tau_L}^{\tau_H}|\tau-\bar{\tau}|\mu(\tau)d\tau\right]\left[\int_{\tau_L}^{\tau_H}|\tau-\bar{\tau}|^{n}\mu(\tau)d\tau\right]-\frac{1}{2}\left[\int_{\tau_L}^{\tau_H}|\tau-\bar{\tau}|^{n+1}\mu(\tau)d\tau\right]<0,$$

since $E[xy]=E[x]E[y]+Cov(x,y)$ can be applied to $x=|\tau-\bar{\tau}|$ and $y=|\tau-\bar{\tau}|^{n}$, which are clearly positively correlated. Finally, since this term is multiplied by $V^{n+1}(\bar{C})$, which for these n is positive by hypothesis, the terms with $n=2,4,6,...$ have a negative contribution to $\frac{d}{d\omega}W^{e}(0;z)$.

For $n=1,3,5,...$ we have, using the symmetry of μ,

$$\left[\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})\mu(\tau)d\tau\right]\times 0-D'(0)\left[\int_{\tau_L}^{\tau_H}(\tau-\bar{\tau})^{n+1}\mu(\tau)d\tau\right]\left[\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})^{n+1}\mu(\tau)d\tau\right]$$
$$=-\bar{D}_1'(0)\int_{\tau_L}^{\tau_H}(\tau-\bar{\tau})^{n+1}\mu(\tau)d\tau-\frac{1}{2}\int_{\tau_L}^{\tau_H}(\tau-\bar{\tau})^{n+1}\mu(\tau)d\tau$$
$$\left[\frac{1}{2}+\eta\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})\mu(\tau)d\tau\right]\int_{\tau_L}^{\tau_H}(\tau-\bar{\tau})^{n+1}\mu(\tau)d\tau-\frac{1}{2}\int_{\tau_L}^{\tau_H}(\tau-\bar{\tau})^{n+1}\mu(\tau)d\tau$$
$$=\eta\left[\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})\mu(\tau)d\tau\right]\int_{\tau_L}^{\tau_H}(\tau-\bar{\tau})^{n+1}\mu(\tau)d\tau>0,$$

where we use that symmetry implies that $-\bar{D}_1'(0)=1/2+\eta\int_{\bar{\tau}}^{\tau_H}(\tau-\bar{\tau})\mu(\tau)d\tau$. Finally, since this term is multiplied by $V^{n+1}(\bar{C})$, which for these n is negative by hypothesis, the terms with $n=1,3,4,...$ have a negative contribution to $\frac{d}{d\omega}W^{e}(0;z)$.

The expression for the approximation is obtained by using the term for $n=1$. $\square$

C Mechanism design approach to optimal Tobin taxes

We now consider a mechanism design approach in which we assume that agents' realized type τ is private information at $t = 1$. This is a natural assumption for the risk tolerance parameter τ and a reasonable justification for the assumption of a lack of time $t = 0$ insurance against the realization of time $t = 1$ value of τ in previous sections, i.e., a reasonable justification for the assumption of incomplete markets. We discuss several specifications of this mechanism design problem. First we briefly discuss the case in which a mechanism designer is able to control the consumption of an agent. For this case, we show that the optimal allocation is not incentive compatible. We then consider a version of the problem in which the mechanism designer can allocate claims to consumption at $t = 2$, but agents can trade a complete set of contingent claims themselves at $t = 1$. This restriction implies that the planner must choose among conditionally efficient allocations of consumption. In this case, we show that, if the distribution of risk tolerance τ has a density, then the equal wealth equilibrium allocation is the only conditionally efficient allocation that is incentive compatible. This result implies that, if τ has a density and agents can trade securities at $t = 1$, then the solution to this mechanism design problem must be to trade off the conditional efficiency of the allocation of risk at $t = 1$ against the risk agents perceive as of $t = 0$ of shocks to τ affecting their desire to hold such risk. We then turn in the next section to the main case where the designer must use an investor-specific portfolio of shares of the aggregate endowment at $t = 2$ and uncontingent transfers (bonds). There we explore the extent to which the solution to this mechanism design problem resembles a Tobin subsidy to trade.

We now specify the mechanism design problem in which the planner can fully control agents' consumption. Consider a given allocation of consumption at $t = 2$ contingent on agents' announced type τ' at $t = 1$ and the realized value of y at $t = 2$ denoted by $c(\tau', y)$. For simplicity, we suppress reference to the aggregate shock z realized at $t = 1$. The certainty equivalent consumption obtained by an agent of type τ who announces type τ' at $t = 1$ is given by

$$\mathbb{C}(\tau, \tau') = U_\tau^{-1}\left[\sum_y U_\tau(c(\tau', y))\rho(y)\right].$$

An allocation $\{c(\tau' y)\}$ for all τ, y is *incentive compatible* if

$$\mathbb{C}(\tau, \tau) \geq \mathbb{C}(\tau, \tau') \text{ for all } \tau, \tau'. \tag{69}$$

We then have the following result.

Lemma 2. *The optimal allocation is not incentive compatible.*

Proof. This lemma follows directly from the definition of risk tolerance. In the first-best allocation, we have all agents receiving the same certainty equivalent consumption $\mathbb{C}(\tau, \tau) = \mathbb{C}(\tau', \tau')$. But if $\tau' > \tau$ and there is any uncertainty in the allocation of consumption to type τ, then we have that the agent with higher risk tolerance obtains a higher certainty equivalent consumption from the allocation assigned to type τ than does that type, i.e. , $\mathbb{C}(\tau, \tau) < \mathbb{C}(\tau, \tau')$. But then incentive compatibility requires that $\mathbb{C}(\tau, \tau) < \mathbb{C}(\tau', \tau')$, which is a contradiction. □

This lemma highlights the fundamental tension in this economy. Risk sharing requires equating the allocation of certainty equivalent consumption across agents, but incentive compatibility implies that agents with higher risk tolerance must receive higher certainty equivalent consumption.

We now consider what incentive compatible allocations of consumption can be achieved if the planner makes transfers to agents at time $t = 1$ based on their announced risk tolerance and then lets agents trade at $t = 1$ based on these post-transfer endowments. The allocations implemented in this way are conditionally efficient. We now show that, when τ has a distribution with a density μ, then the equal wealth equilibrium allocation is the *only* incentive compatible allocation among conditionally efficient allocations. That is, the planner cannot improve on this allocation through a mechanism that makes transfers to agents based on their reported risk tolerance and allows agents to engage in trade after these transfers.

Lemma 3. *Assume that there is a continuum of types of agents τ, and let $\mu(\tau)$ denote the strictly positive density of agents of type τ. Then the only conditionally efficient allocation that is incentive compatible is the equal wealth equilibrium allocation.*

Proof. Recall that conditionally efficient allocations $c(\tau, y)$ take the form

$$c(\tau, y) = \phi(\tau)y + \gamma(\bar{\tau}\phi(\tau) - \tau).$$

One necessary condition for incentive compatibility is

$$\frac{\partial}{\partial \tau'} \mathbb{C}(\tau, \tau') \bigg|_{\tau = \tau'} = 0,$$

which can be written as

$$\sum_y U'_\tau(c(\tau, y)) \left[y + \gamma\bar{\tau}\right] \rho(y) \phi'(\tau) = \sum_y U'_\tau(c(\tau, y)) \gamma \rho(y)$$

or

$$\frac{\sum_y U'_\tau(c(\tau,y))\left[\frac{y}{\gamma}+\bar{\tau}\right]\rho(y)}{\sum_y U'_\tau(c(\tau,y))\rho(y)} = \frac{1}{\phi'(\tau)}.$$

Using the form of conditionally efficient consumption given above, together with the specification of $U_\tau(\cdot)$, we have

$$\frac{1}{\phi'(\tau)} = \frac{\sum_y \left(\frac{y}{\gamma}+\bar{\tau}\right)^{-\gamma}\left(\frac{y}{\gamma}+\bar{\tau}\right)\rho(y)}{\sum_y \left(\frac{y}{\gamma}+\bar{\tau}\right)^{-\gamma}\rho(y)} = \frac{D_1^*}{\gamma}+\bar{\tau}.$$

This result, together with the requirement that the shares $\int \phi(\tau)\mu(\tau)d\tau = 1$ integrate to one, implies that

$$\phi(\tau) = \frac{\frac{D_1^*}{\gamma}+\tau}{\frac{D_1^*}{\gamma}+\bar{\tau}},$$

which is the form for $\phi(\tau)$ given in equation (30) for the equal wealth conditionally efficient allocation. □

This lemma can be interpreted as a justification for our focus on the incomplete markets equilibrium.

C.1 Optimal Non-linear Tax on Trade

We now study a mechanism design problem in which the planner allocates to each agent a portfolio of shares of the aggregate endowment y at $t = 2$ and an uncontingent transfer which can be interpreted as a bond. The incentive compatible mechanisms we study correspond to a menu of uncontingent bonds and risky equity from which the investor must choose only one point on the menu, so that they are not allowed to retrade. These mechanisms can be interpreted as the allocations implemented by a non-linear tax/subsidy on trade in assets at $t = 1$. Below we compare the features of this optimal tax/subsidy with the Tobin tax/subsidy analyzed in the main paper.

The specification of our mechanism design problem is as follows. To simplify the notation, assume that there is only one possible value of z and suppress that in the notation. Then the mechanism design problem is one of choosing a menu of shares $\mathcal{S}(\tau)$ and bonds $\mathcal{B}(\tau)$ as functions of agents' reported risk tolerance τ with corresponding allocation of consumption at $t = 2$, $c(\tau, y) = \mathcal{B}(\tau) + \mathcal{S}(\tau)y$. This allocation of portfolios induces an allocation of certainty equivalent consumption as a function of agents' true risk tolerance τ and reported risk tolerance τ' given by

$$\mathbb{C}(\tau, \tau') = U_{\tau}^{-1}\left(\mathbb{E}\left[U_{\tau}\left(\mathcal{B}(\tau') + \mathcal{S}(\tau')y\right)\right]\right), \tag{70}$$

where the expectation is taken with respect to y as of $t = 1$.

The *second-best* allocation of shares and uncontingent transfers solves the problem of maximizing ex-ante welfare,

$$W = \mathbb{E}\left[V\left(\mathbb{C}(\tau, \tau)\right)\right],$$

subject to the constraints that $\mathbb{C}(\tau, \tau')$ is given by equation (70) for all τ, τ', the resource constraints on portfolios,

$$1 = \mathbb{E}\left[\mathcal{S}(\tau)\right], \tag{71}$$

$$0 = \mathbb{E}\left[\mathcal{B}(\tau)\right], \tag{72}$$

and the incentive constraints (69). In the objective of this problem and in the constraints (71) and (72), the expectation operator $\mathbb{E}$ is respect to τ as of $t = 0$.

We first consider the case in which there are only two values of τ. We then turn to the problem in which τ is a continuous random variable with a density function.

Two Types τ. With only two types of τ, the incentive compatibility constraints (69) simplify to a single constraint that the agent with high risk tolerance does not want to report that he or she has low risk tolerance $\mathbb{C}(\tau_2, \tau_2) \geq \mathbb{C}(\tau_2, \tau_1)$. The result that this is the only one of the two incentive constraints that is binding follows from the single crossing property of agents' indifference curves over shares s and bonds b that can be shown when agents have equicautious HARA preferences. Moreover, given that we know that the solution of this problem without the incentive constraint is not incentive compatible, we have that the incentive constraint must bind as an equality.

The solution to this mechanism design problem has several features in common with the allocation that arises if a small Tobin subsidy is imposed on the equal wealth equilibrium allocation.

First, the solution to this mechanism design problem must offer the risk-averse agents higher certainty equivalent consumption than they achieve in the equal wealth equilibrium allocation (denoted with a superindex e), i.e., $\mathbb{C}(\tau_1, \tau_1) > \mathbb{C}^e(\tau_1)$, and vice versa for the risk-tolerant agents, $\mathbb{C}(\tau_2, \tau_2) < \mathbb{C}^e(\tau_2)$.

Second, the solution to the mechanism design problem allocates aggregate risk to the risk-tolerant agent, i.e., $\mathcal{S}(\tau_2) > \mathcal{S}(\tau_1)$, and bonds $\mathcal{B}(\tau_2) < \mathcal{B}(\tau_1)$ such that the risk-tolerant agents are indifferent between these two portfolios, $\mathbb{C}(\tau_2, \tau_2) = \mathbb{C}(\tau_2, \tau_1)$.

Third, if we add the assumption that agents have CARA preferences, then the solution to this mechanism design problem implies more trade than in the equal wealth equilibrium. That is, in the solution to the mechanism design problem, the risk-tolerant agent has a higher exposure to aggregate risk than would be the case in the equal wealth equilibrium and vice-versa for the risk-averse agent.

These three results are summarized in the following proposition.

Proposition 15. *Assume that $\tau \in \{\tau_1, \tau_2\}$. Then*

(i) The second-best allocation has higher certainty equivalent consumption for the risk-averse agents than these agents receive in the equal wealth equilibrium $\mathbb{C}(\tau_1, \tau_1) > \mathbb{C}^e(\tau_1)$ and vice versa for the risk-tolerant agents $\mathbb{C}(\tau_2, \tau_2) < \mathbb{C}^e(\tau_2)$.

(ii) Assume also that U_{τ_i} are both CARA, so that $\gamma \to \infty$. In any allocation that has a binding incentive constraint and that is conditionally efficient, ex-ante welfare can be improved by a small deviation. In particular, the allocation that improves ex-ante welfare by a small deviation satisfies

$$\frac{\mathcal{S}(\tau_1)}{\tau_1} < \frac{\phi^e(\tau_1)}{\tau_1} = \frac{1}{\bar{\bar{\tau}}} = \frac{\phi^e(\tau_2)}{\tau_2} < \frac{\mathcal{S}(\tau_2)}{\tau_2}. \tag{73}$$

(iii) Assume that U_{τ_i} are both CARA, so that $\gamma \to \infty$. If the set of feasible and incentive compatible allocations is convex, then the second-best allocation also has more dispersed risk exposure, i.e., (73) holds for the second best.

(iv) A sufficient condition for the convexity of the feasible set when both U_{τ_i} are CARA is that $\varphi'''(\cdot) \leq 0$ and $\mu_2 \leq \mu_1$, where φ is defined in (74), and it is a function solely of the distribution of y.

Proof. (i) Consider first the observation that $\mathbb{C}(\tau_1, \tau_1) > \mathbb{C}^e(\tau_1)$ and vice versa for the risk-tolerant agents. This result follows from the observation that the incentive constraint is slack at the equal wealth equilibrium allocation with discrete types since the risk-tolerant agents are able to purchase the equilibrium allocation of the risk averse agents but choose not to. Because this constraint is slack in the equilibrium allocation, it is possible to strictly improve ex-ante welfare by transferring bonds from the risk-tolerant agents to the risk-averse agents and thus bring their certainty equivalent consumption closer together. Hence, similar to the case of the Tobin subsidy, the optimal incentive compatible mechanism effects a transfer of certainty equivalent consumption from risk-tolerant to risk-averse agents via the incidence of the mechanism.

(ii) To simplify the notation in the remainder of the proof, we use the subindex 1 and 2 instead of the arguments τ_1 and τ_2 for all variables. Also we use that for CARA utility we can write the certainty equivalent as $\mathbb{C}_i \equiv \mathcal{B}_i + \tau_i \varphi(\mathcal{S}_i/\tau_i)$ where $\varphi(\cdot)$ is defined in expression (74). We impose the binding IC constraint

$$\mathcal{B}_2 + \tau_2 \varphi(\mathcal{S}_2/\tau_2) = \mathcal{B}_1 + \tau_2 \varphi(\mathcal{S}_1/\tau_2)$$

and the feasibility constraint

$$\begin{aligned} 0 &= \mathcal{B}_2 \mu_2 + \mathcal{B}_1 \mu_1 \\ 1 &= \mathcal{S}_2 \mu_2 + \mathcal{S}_1 \mu_1. \end{aligned}$$

Using these three constraints, we can parameterize the set of allocations that are feasible and have binding IC, which involve four variables $(\mathcal{B}_1, \mathcal{S}_1, \mathcal{B}_2, \mathcal{S}_2)$, as a one-dimensional manifold. In particular, we write them as a function of $\mathcal{B}_1$.

The certainty equivalent consumption for each type is thus:

$$\begin{aligned} \mathbb{C}_2 &= \mathcal{B}_2 + \tau_2 \varphi(\mathcal{S}_2/\tau_2) \\ \mathbb{C}_1 &= \mathcal{B}_1 + \tau_1 \varphi(\mathcal{S}_1/\tau_1). \end{aligned}$$

We can thus write ex-ante welfare as a function of $\mathcal{B}_1$:

$$E[V](\mathcal{B}_1) = V(\mathbb{C}_1(\mathcal{B}_1))\mu_1 + V(\mathbb{C}_2(\mathcal{B}_1))\mu_2.$$

We show that, evaluated at the feasible allocation with a binding IC constraint that is conditionally efficient, then $\mathcal{S}_2/\tau_2 = \mathcal{S}_1/\tau_1 = x$ and $\frac{dE[V](\mathcal{B}_1)}{d\mathcal{B}_1} > 0$, and hence a decrease in $\mathcal{S}_1$ and an increase in $\mathcal{S}_2$ improve ex-ante welfare.

We consider the deviations in $(\mathcal{B}_1, \mathcal{S}_2, \mathcal{B}_2, \mathcal{S}_2)$ that are feasible and where the IC constraint holds with equality, i.e.,

$$\begin{aligned} d\mathcal{B}_2 &= -\frac{\mu_1}{\mu_2} d\mathcal{B}_1 \\ d\mathcal{S}_2 &= -\frac{\mu_1}{\mu_2} d\mathcal{S}_1 \\ d\mathcal{B}_2 + \varphi'(\mathcal{S}_2/\tau_2) d\mathcal{S}_2 &= d\mathcal{B}_1 + \varphi'(\mathcal{S}_1/\tau_2) d\mathcal{S}_1. \end{aligned}$$

Substituting feasibility into the binding IC constraint, we have

$$
\begin{aligned}
-\frac{\mu_1}{\mu_2} d\mathcal{B}_1 - \varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} d\mathcal{S}_1 &= d\mathcal{B}_1 + \varphi'(\mathcal{S}_1/\tau_2) d\mathcal{S}_1 \\
-\left(\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)\right) d\mathcal{S}_1 &= d\mathcal{B}_1 \left(1 + \frac{\mu_1}{\mu_2}\right) \\
-\frac{\left(1 + \frac{\mu_1}{\mu_2}\right)}{\left(\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)\right)} d\mathcal{B}_1 &= d\mathcal{S}_1.
\end{aligned}
$$

The change in ex-ante utility is

$$
\begin{aligned}
dE[V] &= V'(\mathbb{C}_1) d\mathbb{C}_1 \mu_1 + V'(\mathbb{C}_2) d\mathbb{C}_2 \mu_2 \\
&= V'(\mathbb{C}_1)\left[d\mathcal{B}_1 + \varphi'(\mathcal{S}_1/\tau_1) d\mathcal{S}_1\right]\mu_1 + V'(\mathbb{C}_2)\left[d\mathcal{B}_2 + \varphi'(\mathcal{S}_2/\tau_2) d\mathcal{S}_2\right]\mu_2.
\end{aligned}
$$

Substituting feasibility, we have

$$
\begin{aligned}
dE[V] &= V'(\mathbb{C}_1)\left[d\mathcal{B}_1 + \varphi'(\mathcal{S}_1/\tau_1) d\mathcal{S}_1\right]\mu_1 - V'(\mathbb{C}_2)\left[d\mathcal{B}_1 + \varphi'(\mathcal{S}_2/\tau_1) d\mathcal{S}_1\right]\mu_1 \\
&= \left[V'(\mathbb{C}_1) - V'(\mathbb{C}_2)\right] d\mathcal{B}_1 \mu_1 + \left[V'(\mathbb{C}_1)\varphi'(\mathcal{S}_1/\tau_1) - V'(\mathbb{C}_2)\varphi'(\mathcal{S}_2/\tau_2)\right] d\mathcal{S}_1 \mu_1.
\end{aligned}
$$

Substituting the IC constraint we have

$$
\frac{dE[V]}{\mu_1} = \left[V'(\mathbb{C}_1) - V'(\mathbb{C}_2)\right] d\mathcal{B}_1 - \frac{\left(1 + \frac{\mu_1}{\mu_2}\right)\left[V'(\mathbb{C}_1)\varphi'(\mathcal{S}_1/\tau_1) - V'(\mathbb{C}_2)\varphi'(\mathcal{S}_2/\tau_2)\right]}{\left(\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)\right)} d\mathcal{B}_1.
$$

Denoting by $x_1 = \mathcal{S}_1/\tau_1$ and $x_2 = \mathcal{S}_2/\tau_2$, we have

$$
\frac{1}{\mu_1}\frac{dE[V]}{d\mathcal{B}_1} = \left[V'(\mathbb{C}_1) - V'(\mathbb{C}_2)\right] - \frac{\left(1 + \frac{\mu_1}{\mu_2}\right)\left[V'(\mathbb{C}_1)\varphi'(x_1) - V'(\mathbb{C}_2)\varphi'(x_2)\right]}{\left(\varphi'(x_2)\frac{\mu_1}{\mu_2} + \varphi'(x_1\tau_1/\tau_2)\right)}.
$$

Note that if $x_1 = x_2 = x$, we have

$$
\frac{1}{\mu_1}\frac{dE[V]}{d\mathcal{B}_1} = \left[V'(\mathbb{C}_1) - V'(\mathbb{C}_2)\right] - \frac{\left(1 + \frac{\mu_1}{\mu_2}\right)}{\left(\frac{\mu_1}{\mu_2} + \frac{\varphi'(x\tau_1/\tau_2)}{\varphi'(x)}\right)}\left[V'(\mathbb{C}_1) - V'(\mathbb{C}_2)\right] > 0,
$$

since $\mathbb{C}_1 < \mathbb{C}_2$ and hence $V'(\mathbb{C}_1) > V'(\mathbb{C}_2)$. Using the binding IC and feasibility constraints, we have, starting at $x_2 = x_1 = x$ (which characterize the constraint efficient allocations), that ex-ante welfare can be improved by increasing $\mathcal{B}_1$ and since

$$
d\mathcal{S}_1 = -\frac{\left(1 + \frac{\mu_1}{\mu_2}\right)}{\left(\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)\right)} < 0
$$

by decreasing $\mathcal{S}_1$.

(iii) Now we show that if $\mu_2/\mu_1 \leq 1$ and $\phi''' \leq 0$, then in the optimal allocation, $x_2 > 1/\bar{\tau} > x_1$. To prove this, we show that the set of feasible allocations is convex. If this set is convex, then $E[V](\mathcal{B}_1)$ must be concave, since V is a concave function. Hence, if $\partial E[V](\mathcal{B}_1)/\partial \mathcal{B}_1 > 0$ evaluated at $x_1 = x_2 = 1/\bar{\tau}$, then the optimal must have $\mathcal{B}_1$ larger than that amount, and thus the x's must be more dispersed. To establish the convexity of the feasible set, note that the feasibility constraints are linear, so that they define, as inequalities, a convex set. The remaining constraint is incentive compatibility, which can be written as

$$\begin{aligned}
&G(\mathcal{B}_2, \mathcal{B}_1, \mathcal{S}_2, \mathcal{S}_1) \geq 0 \text{ where} \\
&G(\mathcal{B}_2, \mathcal{B}_1, \mathcal{S}_2, \mathcal{S}_1) \equiv \mathcal{B}_2 + \tau_2\varphi(\mathcal{S}_2/\tau_2) - \mathcal{B}_1 - \tau_2\varphi(\mathcal{S}_1/\tau_2).
\end{aligned}$$

If G is a concave function, then the set of values for which $G \geq 0$ is convex. Since G is linear in $\mathcal{B}_2$ and $\mathcal{B}_1$, it suffices to show that it is concave in $\mathcal{S}_1, \mathcal{S}_2$. We substitute the feasibility constraint for S in G, obtaining

$$G(\mathcal{B}_2, \mathcal{B}_1, \mathcal{S}_2, (1-\mathcal{S}_2\mu_2)/\mu_1) \equiv \mathcal{B}_2 + \tau_2\varphi\left(\frac{\mathcal{S}_2}{\tau_2}\right) - \mathcal{B}_1 - \tau_2\varphi\left(\frac{(1-\mathcal{S}_2\mu_2)/\mu_1}{\tau_2}\right).$$

Thus, G is concave if and only $\frac{d^2}{d\mathcal{S}_2^2}G \leq 0$. Direct computation gives

$$\begin{aligned}
\frac{d}{d\mathcal{S}_2}G(\mathcal{B}_2, \mathcal{B}_1, \mathcal{S}_2, (1-\mathcal{S}_2\mu_2)/\mu_1) &\equiv \varphi'\left(\frac{\mathcal{S}_2}{\tau_2}\right) + \varphi'\left(\frac{(1-\mathcal{S}_2\mu_2)/\mu_1}{\tau_2}\right)\frac{\mu_2}{\mu_1} \\
\frac{d^2}{d\mathcal{S}_2^2}G(\mathcal{B}_2, \mathcal{B}_1, \mathcal{S}_2, (1-\mathcal{S}_2\mu_2)/\mu_1) &\equiv \frac{1}{\tau_2}\left[\varphi''\left(\frac{\mathcal{S}_2}{\tau_2}\right) - \varphi''\left(\frac{(1-\mathcal{S}_2\mu_2)/\mu_1}{\tau_2}\right)\left(\frac{\mu_2}{\mu_1}\right)^2\right].
\end{aligned}$$

Since $\varphi'' \leq 0$ and we have assumed that $\frac{\mu_2}{\mu_1} \leq 1$,

$$\begin{aligned}
\frac{d^2}{d\mathcal{S}_2^2}G(\mathcal{B}_2, \mathcal{B}_1, \mathcal{S}_2, (1-\mathcal{S}_2\mu_2)/\mu_1) &= \frac{1}{\tau_2}\left[\varphi''\left(\frac{\mathcal{S}_2}{\tau_2}\right) - \varphi''\left(\frac{\mathcal{S}_1}{\tau_2}\right)\left(\frac{\mu_2}{\mu_1}\right)^2\right] \\
&\leq \frac{1}{\tau_2}\left[\varphi''\left(\frac{\mathcal{S}_2}{\tau_2}\right) - \varphi''\left(\frac{\mathcal{S}_1}{\tau_2}\right)\right] \leq 0,
\end{aligned}$$

where the last inequality follows by $\varphi''' \leq 0$ and if $\mathcal{S}_2 \geq \mathcal{S}_1$.

The last step is to show that $\mathcal{S}_2 \geq \mathcal{S}_1$. For this, as a contradiction, assume that $\mathcal{S}_2 < \mathcal{S}_1$.

In this case, note that

$$\frac{1}{\mu_1}\frac{dE[V]}{d\mathcal{B}_1} = [V'(\mathbb{C}_1) - V'(\mathbb{C}_2)] - \frac{\left(1+\frac{\mu_1}{\mu_2}\right)[V'(\mathbb{C}_1)\varphi'(\mathcal{S}_1/\tau_1) - V'(\mathbb{C}_2)\varphi'(\mathcal{S}_2/\tau_2)]}{\left(\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)\right)}$$
$$\geq [V'(\mathbb{C}_1) - V'(\mathbb{C}_2)] - \frac{\left(1+\frac{\mu_1}{\mu_2}\right)[V'(\mathbb{C}_1)\varphi'(\mathcal{S}_1/\tau_2) - V'(\mathbb{C}_2)\varphi'(\mathcal{S}_2/\tau_2)]}{\left(\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)\right)},$$

where the inequality follows from the concavity of φ and from $\tau_2 > \tau_1$. Rearranging the right-hand side, we can write

$$\frac{1}{\mu_1}\frac{dE[V]}{d\mathcal{B}_1} \geq V'(\mathbb{C}_1)\left[1 - \frac{\left(1+\frac{\mu_1}{\mu_2}\right)\varphi'(\mathcal{S}_1/\tau_2)}{\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)}\right] - V'(\mathbb{C}_2)\left[1 - \frac{\left(1+\frac{\mu_1}{\mu_2}\right)\varphi'(\mathcal{S}_2/\tau_2)}{\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)}\right]$$
$$= V'(\mathbb{C}_1)\frac{\mu_1}{\mu_2}\left[\frac{\varphi'(\mathcal{S}_2/\tau_2) - \varphi'(\mathcal{S}_1/\tau_2)}{\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)}\right] - V'(\mathbb{C}_2)\left[\frac{\varphi'(\mathcal{S}_1/\tau_2) - \varphi'(\mathcal{S}_2/\tau_2)}{\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)}\right]$$
$$= \left[V'(\mathbb{C}_1)\frac{\mu_1}{\mu_2} + V'(\mathbb{C}_2)\right]\left[\frac{\varphi'(\mathcal{S}_2/\tau_2) - \varphi'(\mathcal{S}_1/\tau_2)}{\varphi'(\mathcal{S}_2/\tau_2)\frac{\mu_1}{\mu_2} + \varphi'(\mathcal{S}_1/\tau_2)}\right].$$

Thus, if $\mathcal{S}_1 > \mathcal{S}_2$, then $\varphi'(\mathcal{S}_2/\tau_2) - \varphi'(\mathcal{S}_1/\tau_2) > 0$ and $\frac{dE[V]}{d\mathcal{B}_1} > 0$, which means that $\mathcal{S}_1 > \mathcal{S}_2$ cannot be optimal.

□

A few comments on this proposition relating it to a small Tobin subsidy are in order. First, with U_τ CARA, the equilibrium allocation as well as all the conditionally efficient allocations satisfy $\phi^e(\tau_1)/\tau_1 = 1/\bar{\tau} = \phi^e(\tau_2)/\tau_2$ as in equation (73). Second, note that a small Tobin subsidy for the case of two types is essentially the same as the deviation in item (ii). Third, if equation (73) holds, the volume of trade in shares at $t = 1$ as measured by $\mathbb{E}|\mathcal{S}(\tau) - 1|$ is higher in the second-best allocation than in the equal wealth competitive equilibrium. Fourth, below, we give examples of distributions for which $\varphi''' \leq 0$. For instance, $\varphi''' = 0$ when y is normally distributed, a case that it is often used together with CARA utility functions U_τ. Fifth, there are many ways to decentralize the second-best allocation.[11]

We now give a description of how to decentralize this allocation with a non-linear tax/subsidy on trade implemented with simple policies.

[11] One way is to give agents a menu of portfolios containing only the two second-best portfolios. Alternatively, we can define the largest set of portfolios that decentralize the second-best allocation.

Decentralization of second-best allocation in the case of two values of τ. In the case with only two values of τ, the solution to the mechanism design problem can be implemented with a piecewise linear menu of share and bond portfolios. In particular, each agent has a piecewise linear budget set for shares and bonds that arises from a fixed fee (in terms of bonds) to enter the market that is specific to buyers and sellers of shares, a Tobin subsidy of the shares purchased by agents of type τ_2 from agents of type τ_1 (so that the buying price is lower than the selling price), with a cap on the subsidy limiting it to the share sales $1 - \mathcal{S}(\tau_1)$ mandated by the mechanism (or equivalently, a tax on further sales of shares by agents of type τ_1 beyond the quantity $1 - \mathcal{S}(\tau_1)$).

In the case of a continuum of τ's, there are many more incentive constraints, and hence the menu is essentially uniquely characterized. We turn to that case next.

The case of a continuum of values of τ. We now consider the version of our mechanism design problem in which the distribution of agents' types τ has a density $\mu(\tau)$. As we saw in Lemma 3, in this case, the equilibrium allocation is the only incentive compatible conditionally efficient allocation. Thus, a planner must necessarily trade off the conditional efficiency of the allocation of portfolios at $t = 1$ against the risk sharing properties of the allocation evaluated as of $t = 0$. We refer to the solution of this problem as the *second-best allocation*.

We obtain a more limited set of results relative to what we found with the two-type case. In particular, we focus on developing results regarding the allocation of aggregate risk in the second-best allocation and the pricing of that risk as reflected in agents' marginal rate of substitution between shares and bonds at $t = 1$ as a function of their realized risk tolerance τ. We show that agents with extremely low and high values of τ take on more aggregate risk in the second-best allocation than is the case in the equal wealth equilibrium, while agents with intermediate values of τ take on less risk. We explore the extent to which one can interpret this allocation of aggregate risk as arising from a non-linear subsidy to trade, both in terms of the portfolios of the agents with extreme values of τ and in terms of overall trade volumes.

We specialize our analysis to the case of CARA preferences. We first describe the mechanism design problem and then present our results.

When agents have CARA preferences, the certainty equivalent consumption of the investor with shares $\mathcal{S}(\tau')$ and transfer $\mathcal{B}(\tau')$, once the realization of her risk tolerance τ is known to her, is given by $\mathbb{C}(\tau,\tau') = \tau\varphi(\mathcal{S}(\tau')/\tau) + \mathcal{B}(\tau')$, where the function φ is given by

$$\varphi(x) \equiv -\log E\left[e^{-xy}\right] = -\log\int e^{-xy}\rho(y)dy. \tag{74}$$

In what follows, for some results we must also specialize the model to the case in which the endowment at $t = 2$, y, has a normal distribution. Note that if $y \sim N(\mu_y, \sigma_y^2)$, then $\varphi(x) = x\mu_y - \frac{\sigma_y^2}{2}x^2$.

Properties of φ. Define φ as in (74), then

$$\begin{aligned} &\varphi(0) = 0\,,\ \varphi(1) > 0\,,\ \varphi'(x) > 0 \text{ for all } x \text{ if } y \geq 0\ a.s.\,,\ \varphi'(0) = \mu_y\ , \\ &\varphi''(x) < 0 \text{ for all } x\,,\ \varphi''(0) = -\sigma_y^2. \end{aligned} \tag{75}$$

These properties are obtained as follows. Let $\kappa(t) \equiv \log E[e^{ty}]$ be the cumulant generating function of y. We then have

$$\varphi(x) = -\kappa(-x). \tag{76}$$

The n^{th} derivatives of these two functions are related by

$$\frac{\partial^n}{\partial x^n}\varphi(x) = (-1)^{(n+1)}\frac{\partial^n}{\partial x^n}\kappa(-x) \text{ for } n = 1, 2, ... \tag{77}$$

Since the cumulant generating function is a convex function, then $\varphi(x)$ is concave. Alternatively computing $\varphi''(x)$, we obtain

$$\varphi''(x) = -\left[\frac{E[e^{-yx}y^2]}{E[e^{-yx}]} - \left(\frac{E[e^{-yx}]y}{E[e^{-yx}]}\right)^2\right] \equiv -Var_x(y) < 0, \tag{78}$$

where $Var_x(y)$ denotes the variance of the distribution of y computed with the slanted density $\phi(y)\exp(-yx)/[\int \exp(-y'x)\rho(y')dy']$. For the first derivative, direct computation gives $\varphi'(x) = E\left[e^{-xy}y\right]/E\left[e^{-xy}\right]$. It is well known that the first two derivatives of the cumulant generating function are the expected value and variance, i.e., $\kappa'(0) = \mu_y$ and $\kappa''(0) = \sigma_y^2$.

Below we give examples of φ and its associated function Φ for some distributions.

Normal case. Suppose that y is normal $N(\mu_y, \sigma_y^2)$. Then

$$\Phi(x) = \varphi(x) - \varphi'(x)\,x = \frac{\sigma_y^2}{2}x^2\,,\ \Phi'(x) = \sigma_y^2 x > 0 \text{ and } \Phi''(x) = \sigma_y^2 > 0\,.$$

Poisson. If y is Poisson with mean μ_y then

$$\varphi(x) = -\mu_y\left(e^{-x} - 1\right) \text{ and } \Phi''(x) = \mu_y e^{-x}\left[1 - x\right], \tag{79}$$

so $\Phi''(x) > 0$ for $x < 1$ and $\Phi''(x) < 0$ for $x > 1$. Thus in this case, Φ' is not monotone.

Binominal. Suppose y is distributed as the outcome of n trials each with success with probability p. In this case,

$$\varphi(x) = -\log\left(1 - p + pe^{-x}\right) \text{ and } \Phi''(x) = \frac{n(1-p)e^x}{\left[(1-p)e^x + p\right]^3}\left[x(1-p)e^x + p(1+x)\right] > 0. \quad (80)$$

In this case Φ' is monotone.

Exponential. Suppose that y is exponential with parameter λ. In this case,

$$\varphi(x) = \log\left(\frac{\lambda + x}{\lambda}\right) \text{ and } \Phi''(x) = \frac{\lambda - x}{(\lambda + x)^3}, \quad (81)$$

so $\Phi'' > 0$ if $x < \lambda$ and $\Phi'' < 0$ if $x > \lambda$.

The planner wants to maximize ex-ante expected utility, where we assume that investors evaluate expected utility over their certainty equivalent consumption using utility function V, which we assume to be strictly increasing and strictly concave. Thus, the planner seeks to maximize

$$\int V\left(\tau\varphi\left(\frac{\mathcal{S}(\tau)}{\tau}\right) + \mathcal{B}(\tau)\right)\mu(\tau)d\tau \quad (82)$$

by choosing functions $\mathcal{S}(\cdot)$ and $\mathcal{B}(\cdot)$, subject to the physical constraints (71) and (72) and the incentive compatibility constraint (69) for each τ with certainty equivalent consumption given by the menu of portfolios as above.[12]

Structure of solution to planning problem. To solve the planning problem, we first take as given (θ_s, θ_b) and convert its first-order conditions in the solution of two ordinary differential equations subject to two known boundary conditions. The second step is to solve for the values of (θ_s, θ_b), using implied values for these differential equations for two integral equations, namely the feasibility conditions. We first turn to the description of the differential equation system given (θ_s, θ_b). For this we need to use the first-order condition for $x(\tau)$ and solve it as a function of τ and λ. We denote such function as $x = X(\lambda, \tau; \theta_s, \theta_b)$. Then we use the following system for $\tau \in (\tau_L, \tau_H)$:

$$\mathcal{C}'(\tau) = \Phi\left(X\left(\lambda(\tau), \tau; \theta_s, \theta_b\right)\right) \quad (83)$$

$$\lambda'(\tau) = -V'\left(\mathcal{C}(\tau)\right) - \frac{\mu'(\tau)}{\mu(\tau)}\lambda(\tau), \quad (84)$$

[12]The specification of this mechanism design problem is close but not identical to the one in Diamond (1998), which itself is a version of the problem in Mirrlees (1971) with quasilinear utility.

with boundary conditions

$$\lambda(\tau_H) = 0 \text{ and } \lambda(\tau_L) = \frac{\theta_b}{\mu(\tau_L)} .$$

To solve this two-boundary value problem, we implement a shooting algorithm. We evaluate the system of (83)-(84) with initial condition $\lambda(\tau_L) = \theta_b/\mu(\tau_L)$ and some guess for the value for $\mathcal{C}(\tau_L)$. Then we check if the resulting value of $\lambda(\tau_H)$ satisfies the boundary condition, namely, if $\lambda(\tau_H) = 0$. If not, we change the guess for $\mathcal{C}(\tau_L)$ and repeat the procedure. The next lemma ensures that there exists a unique solution to which this procedure converges by studying the properties of the implied mapping between $\mathcal{C}\tau_L$ and $\lambda(\tau_H)$.

Lemma 4. *Fix two arbitrary values of (θ_s, θ_b). Assume that $V'(\cdot)$ is strictly decreasing and that there is a value for C_0 such that $V'(C_0) = \theta_b$. Let $\lambda_H(\mathcal{C}_L) = \lambda(\tau_H)$ be the value of the solution of the system of two ordinary differential equations (83)-(84) where $\lambda(\cdot)$ is evaluated at τ_H, taking as initial conditions $\mathcal{C}(\tau_L) = \mathcal{C}_L$ and $\lambda(\tau_L) = \theta_b/\mu(\tau_L)$. There exists a unique value $\mathcal{C}_L^*$ that solves $\lambda_H(\mathcal{C}_L^*) = 0$. Furthermore, we can find an interval $\left[\underline{\mathcal{C}}_L\,, \bar{\mathcal{C}}_L\right]$ so that $\lambda_H(\underline{\mathcal{C}}_L) > 0 > \lambda_H(\bar{\mathcal{C}}_L)$ and $\frac{\partial}{\partial \mathcal{C}_L}\lambda_H(\mathcal{C}_L) > 0$ for all $\mathcal{C}_L \in [\underline{\mathcal{C}}_L\,, \mathcal{C}_L^*]$.*

Proof. First we note that $\lambda_H(\mathcal{C}_L) = 0$ if and only if $\theta_b = \int_{\tau_L}^{\tau_H} V'(\mathcal{C}(\tau, \mathcal{C}_L))\mu(\tau)d\tau$ where we let $\mathcal{C}(\tau, \mathcal{C}_L)$, and for future reference $\lambda(\tau, \mathcal{C}_L)$, the solution of the system of differential equations with $\mathcal{C}_L = \mathcal{C}(\tau_L)$. Throughout this lemma, we keep the initial condition $\lambda(\tau_L) = \theta_b/\mu(\tau_L)$. We proceed in three steps.

Step 1. We show that $\lambda_H(\underline{\mathcal{C}_L}) < 0$. To see this, we use that the solution to the first-order condition of $x(\tau)$ given λ and τ, which we denote $X(\lambda, \tau; \tau_b\tau_s)$, is bounded from above by $\bar{x}(\tau)$, which in turn is the solution to

$$\frac{\theta_s}{\theta_b} - \varphi'(\bar{x}(\tau)) = -\frac{\Phi'(\bar{x}(\tau))}{\tau\mu(\tau)} \int_{\tau}^{\tau_H} \mu(t)dt.$$

Given this upper bound we can construct an upper bound for $\mathcal{C}(\tau_H, \mathcal{C}_L)$, namely,

$$\mathcal{C}(\tau_H, \mathcal{C}_L) \leq \mathcal{C}_L + (\tau_H - \tau_L)\bar{\Phi} \equiv \mathcal{C}_L + (\tau_H - \tau_L) \max_{\tau} \Phi(\bar{x}(\tau)).$$

Thus by setting $\mathcal{C}_L$ small enough $V'(\mathcal{C}(\tau, \mathcal{C}_L)) > \theta_b$, and hence $\theta_b < \int_{\tau_L}^{\tau_H} V'(\mathcal{C}(\tau, \mathcal{C}_L))\mu(\tau)d\tau$.

Step 2. We show that $\lambda_H(\underline{\mathcal{C}_L}) > 0$. Since $\mathcal{C}(\tau, \mathcal{C}_L)$ is increasing in τ, then for $\mathcal{C}(\tau_L)$ large enough $\theta_b > \int_{\tau_L}^{\tau_H} V'(\mathcal{C}(\tau, \mathcal{C}_L))\mu(\tau)d\tau$.

Step 3. We now show that $\lambda_H(\cdot)$ is strictly increasing whenever $\lambda_H < 0$. To see this, we totally differentiate the system of differential equations with respect to $\mathcal{C}_L$:

$$\frac{\partial}{\partial \mathcal{C}_L}\lambda(\tau, \mathcal{C}_L) = -\frac{1}{\mu(\tau)}\left[\int_{\tau_L} V''(\mathcal{C}(t, \mathcal{C}_L))\frac{\partial}{\partial \mathcal{C}_L}\mathcal{C}(t, \mathcal{C}_L)\mu(t)dt\right]$$
$$\frac{\partial}{\partial \mathcal{C}_L}\mathcal{C}(\tau, \mathcal{C}_L) = 1 + \int_{\tau_L}^{\tau} \frac{\partial \Phi(X(\lambda(t, \mathcal{C}_L), t))}{\partial x}\frac{\partial X(t, \mathcal{C}_L), t)}{\partial \lambda}\frac{\partial}{\partial \mathcal{C}_L}\lambda(t, \mathcal{C}_L)\, dt$$

for all $\tau \in [\tau_L, \tau_H]$. From the first-order conditions, we have

$$\frac{\partial X(\lambda, \tau)}{\partial \lambda} = \frac{\Phi'(x)/(\tau\theta_b)}{\varphi''(x) - \frac{\Phi''(x)}{\tau\mu(\tau)}\Psi(\tau, \lambda)} \text{ where } x = X(\lambda, \tau) \text{ and } \Psi(\tau, \lambda) = \frac{\mu(\tau)\lambda}{\theta_b} - \int_{\tau}^{\tau_H} \mu(t)dt.$$

We can also write

$$\Psi(\tau, \lambda(\tau, \mathcal{C}_L)) = \frac{1}{\theta_b}\int_{\tau_L}^{\tau} [\theta_b - V'(\mathcal{C}(t, \mathcal{C}_L))]\,\mu(t)dt.$$

Thus, if $\theta_b < \int_{\tau_L}^{\tau_H} V'(\mathcal{C}(\tau, \mathcal{C}_L))\mu(\tau)d\tau$ then $\Psi(\tau, \lambda(\tau, \mathcal{C}_L))$ is non-positive and single peaked, then $\partial X/\partial\lambda > 0$ since $\varphi'' < 0$ and $\Phi'' > 0$. Evaluating the derivatives above for each τ, noticing that it is a recursive system, we have that $\frac{\partial}{\partial \mathcal{C}_L}\lambda(\tau, \mathcal{C}_L) > 1$ and thus $\lambda_H(\mathcal{C}_L) > 0$. □

The next lemma shows that in the case of CARA utility function V, given a solution for the ordinary differential equations above —in particular, given the path for $\{x(\tau)\}_{\tau_L}^{\tau_H}$— we can analytically solve for the value of θ_b and $C(\tau_L)$, keeping the ratio θ_s/θ_b constant which ensures that the feasibility constraint for uncontingent claims holds.

Lemma 5. *Assume that V is a CARA utility function, i.e., $V(C) = -\tau_V \exp(-C/\tau_V)$ for some $\tau_V > 0$. Assume that the path $\{x(\tau)\}_{\tau_L}^{\tau_H}$ and ratio θ_s/θ_b solve the first-order conditions with $\mathcal{C}(\tau_L) = 0$, i.e. ,*

$$\frac{\theta_s}{\theta_b} - \varphi'(x(\tau)) = \frac{\Phi'(x(\tau))}{\tau\mu(\tau)}\int_{\tau_L}^{\tau} \mu(t)\left[1 - \frac{V'\left(\int_{\tau_L}^{t} \Phi(x(t))dt\right)}{\int_{\tau_L}^{\tau_H} V'\left(\int_{\tau_L}^{t} \Phi(x(s))ds\right)\mu(\tau(s))ds}\right]dt \qquad (85)$$

for all $\tau \in [\tau_L, \tau_H]$, and that the path $\{x(\tau)\}_{\tau_L}^{\tau_H}$ also satisfies the feasibility constraint for shares. Then the boundary $\mathcal{C}(\tau_L)$ and θ_b

$$\mathcal{C}(\tau_L) = -\int_{\tau_L}^{\tau_H} \Phi(x(\tau))\left[\int_{\tau_L}^{\tau} \mu(t)dt\right]d\tau + \int_{\tau_L}^{\tau_H} \tau\varphi(x(\tau))\mu(\tau)d\tau$$
$$\theta_b = e^{-\mathcal{C}(\tau_L)/\tau_V}\int_{\tau_L}^{\tau_H} V'\left(\int_{\tau_L}^{\tau} \Phi(x(t))dt\right)\mu(\tau)d\tau$$

satisfy the feasibility condition for bonds, and thus give the complete solution to the problem.

Proof. The proof is immediate, since with CARA the first-order condition (85) is independent of $\mathcal{C}(\tau_L)$. □

The next proposition partially characterizes the allocation of aggregate risk in the second-best allocation. We do so in terms of the variable $x(\tau) \equiv \mathcal{S}(\tau)/\tau$ for the second-best allocation and for the equilibrium with equal wealth (or any other conditionally efficient allocation). The variable $x(\tau)$ represents the allocation of shares of aggregate risk to an agent relative to the risk tolerance of the agent. We have shown above that, in any conditionally efficient allocation, $x(\tau)$ is constant at $1/\bar{\tau}$ for all values of τ. As shown in the next proposition, in the second-best allocation, $x(\tau)$ is U-shaped — lower for intermediate values of τ and equal to a common value at both extremes. This proposition, together with the resource constraint for shares of aggregate risk, implies that, in the second-best allocation, the allocation of aggregate risk to agents with extreme (intermediate) values of τ is greater (smaller) than in the equilibrium allocation.

Proposition 16. *Assume that $\mu(\tau) > 0$ for all $\tau \in [\tau_L, \tau_H]$. Denote $x(\tau) \equiv \mathcal{S}(\tau)/\tau$. Let θ_b and θ_s be the Lagrange multipliers of the constraints (72) and (71), respectively. Then $\varphi'(x(\tau))$ at the top and bottom of the support of τ are the same as the ratio of the Lagrange multipliers, but it is higher for intermediate values:*

$$\frac{\theta_s}{\theta_b} = \varphi'(x(\tau_H)) = \varphi'(x(\tau_L)) < \varphi'(x(\tau)) \text{ for all } \tau \in (\tau_L, \tau_H). \tag{86}$$

Proof. Rearranging the first order condition with respect to $x(\tau)$,

$$\frac{\theta_s}{\theta_b} - \varphi'(x(\tau)) = \frac{\Phi'(x(\tau))}{\tau\mu(\tau)} \left[\frac{\mu(\tau)\lambda(\tau)}{\theta_b} - \int_{\tau}^{\tau_H} \mu(t)dt \right]. \tag{87}$$

The left-hand side gives the different shadow prices, or implicit tax rates faced by agents. The right-hand side determines the sign. It is proportional to the difference between two functions, namely, $\mu(\tau)\lambda(\tau)/\theta_b$ and $\int_{\tau}^{\tau_H} \mu(t)dt$. Both functions start at the value of one at τ_L and decrease to zero as τ increases to τ_H.

Evaluating the first-order condition for x at τ_H and τ_L, and assuming that $\mu(\tau_L) > 0$ and $\mu(\tau_H) > 0$, we obtain

$$\frac{\theta_s}{\theta_b} - \varphi'(x(\tau_H)) = \frac{\theta_s}{\theta_b} - \varphi'(x(\tau_L)). \tag{88}$$

Then differentiating the first-order condition of x with respect to τ,

$$\theta_s\,[\mu(\tau)+\tau\mu'(\tau)]+\theta_b\,\Phi''(x(\tau))x'(\tau)\left[\int_\tau^{\tau_H}\mu(t)dt\right]-\theta_b\,\Phi'(x(\tau))\mu(\tau)$$
$$=\theta_b\,\varphi'(x(\tau))\,[\mu(\tau)+\tau\mu'(\tau)]+\theta_b\,\tau\,\varphi''(x(\tau))x'(\tau)\,\mu(\tau)$$
$$+\Phi''(x(\tau))x'(\tau)\mu(\tau)\lambda(\tau)-\Phi'(x(\tau))V'\left(\mathcal{C}(\tau)\right)\mu(\tau).$$

Rearranging, we obtain

$$x'(\tau)=\frac{\left[\varphi'(x(\tau))-\frac{\theta_s}{\theta_b}\right][\mu(\tau)+\tau\mu'(\tau)]+\Phi'(x(\tau))\mu(\tau)\left[1-\frac{V'(\mathcal{C}(\tau))}{\theta_b}\right]}{\Phi''(x(\tau))\left[\int_\tau^{\tau_H}\mu(t)dt-\frac{\mu(\tau)\lambda(\tau)}{\theta_b}\right]-\tau\,\varphi''(x(\tau))\,\mu(\tau)}.$$

Evaluating this at the extremes, using the values of $\lambda(\tau)\mu(\tau)$, and that $\Phi'(x)=-\varphi''(x)x$, we obtain

$$x'(\tau_L)=\frac{x(\tau_L)\left[\theta_b-V'\left(\mathcal{C}(\tau_L)\right)\right]}{\theta_b\tau_L}\text{ and }x'(\tau_H)=\frac{x(\tau_H)\left[\theta_b-V'\left(\mathcal{C}(\tau_H)\right)\right]}{\theta_b\tau_H}.$$

The equality

$$\theta_b=\int_{\tau_L}^{\tau_H}V'\left(\mathcal{C}(\tau)\right)\mu(\tau)d\tau$$

follows by integrating with respect to τ both sides of the first-order condition with respect to $\mathcal{C})(\tau)$ at $\tau\in(\tau_L,\tau_H)$, obtaining

$$-\int_{\tau_L}^{\tau_H}V'\left(\mathcal{C}\left(\tau\right)\right)\mu\left(\tau\right)d\tau=\int_{\tau_L}^{\tau_H}\left[\lambda'(\tau)\mu(\tau)+\mu'(\tau)\right]d\tau=\lambda(\tau_H)\mu(\tau_H)-\lambda(\tau_L)\mu(\tau_L),$$

and evaluating the right hand side using the first-order conditions with respect to $\mathcal{C}(\tau)$ at the two extremes values, i.e., $\tau=\tau_L$ and $\tau=\tau_H$. Using that $\mathcal{C}(\tau)$ is increasing in τ,

$$V'\left(\mathcal{C}(\tau_H)\right)<\theta_b<V'\left(\mathcal{C}(\tau_L)\right).$$

Hence, $x'(\tau_H)>0>x'(\tau_L)$. To show that $\theta_s/\theta_b-\varphi'(x(\tau))<0$ in the interior we analyze the function

$$\Psi(\tau)\equiv\frac{\mu(\tau)\lambda(\tau)}{\theta_b}-\int_\tau^{\tau_H}\mu(t)dt,$$

since we can write $\theta_s/\theta_b-\varphi'(x(\tau))=\Phi'(x(\tau)/[\tau\mu(\tau)]\Psi(\tau)$. Thus, we have

$$\Psi(\tau)=\int_{\tau_L}^{\tau}\mu(t)\left[1-\frac{V'\left(\mathcal{C}(t)\right)}{\int_{\tau_L}^{\tau_H}V'\left(\mathcal{C}(s)\right)\mu(\tau(s))ds}\right]dt. \tag{89}$$

Note that

$$\Psi(\tau_L)=\Psi(\tau_H)=0\text{ and }\Psi'(\tau)=\mu(\tau)\left[1-\frac{V'\left(\mathcal{C}(\tau)\right)}{\theta_b}\right].$$

Using that $\mathcal{C}'(\tau) > 0$, and that $V'(\mathcal{C}(\tau_H)) < \theta_b < V'(\mathcal{C}(\tau_L))$, so that $\Psi'(\tau_L) < 0$, $\Psi'(\tau_H) > 0$ and $\Psi'(\tau^*) = 0$ at a unique value of τ for which $V'(\mathcal{C}(\tau^*)) = \theta_b$. Hence, it has a unique minimum, and thus $\Psi(\tau) < 0$ for all $\tau \in (\tau_L, \tau_H)$. Since $\Phi'(x)/[\mu(\tau)\tau] > 0$, this gives the result that $\theta_s/\theta_b - \varphi'(x(\tau)) < 0$ in the interior.

Finally we show that $x(\tau)$ is single peaked in τ. For this we show that at τ^* for which $x'(\tau^*) = 0$, we have that $x''(\tau^*) > 0$. To do so, we write

$$x'(\tau) = \frac{f(\tau)}{g(\tau)} \text{ so that } x''(\tau^*) = \frac{f'(\tau^*)}{g(\tau^*)} \text{ where}$$
$$f(\tau) = \left[\varphi'(x(\tau)) - \frac{\theta_s}{\theta_b}\right]\left[1 + \frac{\tau\mu'(\tau)}{\mu(\tau)}\right] + \Phi'(x(\tau))\left[1 - \frac{V'(\mathcal{C}(\tau))}{\theta_b}\right]$$
$$g(\tau) = \frac{\Phi''(x(\tau))}{\mu(\tau)}\left[\int_\tau^{\tau_H}\mu(t)dt - \frac{\mu(\tau)\lambda(\tau)}{\theta_b}\right] - \tau\,\varphi''(x(\tau)),$$

where we have used that $x'(\tau^*) = 0$. Since $\int_\tau^{\tau_H}\mu(t)dt - \frac{\mu(\tau)\lambda(\tau)}{\theta_b} > 0$, $\Phi''(x(\tau)) > 0$ and $\varphi''(x(\tau)) < 0$, we have that $sign\,(x''(\tau^*)) = sign\,(f'(\tau^*))$. Direct computation gives

$$f'(\tau^*) = \left[\varphi'(x(\tau^*)) - \frac{\theta_s}{\theta_b}\right]\left[\frac{\partial}{\partial\tau}\frac{\tau\mu'(\tau)}{\mu(\tau)}\right]_{\tau=\tau^*} - \Phi'(x(\tau^*))\frac{V''(\mathcal{C}(\tau^*))}{\theta_b}\Phi(\tau^*),$$

where we have repeatedly used that $x'(\tau^*) = 0$. Since we have shown that $\varphi'(x(\tau^*)) - \frac{\theta_s}{\theta_b} > 0$, and we have that $\Phi(\tau^*) > 0$, $\Phi'(\tau^*) > 0$, and $V''(\mathcal{C}(\tau^*)) < 0$, then the assumption that $\frac{\partial}{\partial\tau}\frac{\tau\mu'(\tau)}{\mu(\tau)}|_{\tau=\tau^*} > 0$ implies that $f'(\tau^*) > 0$, hence x achieves a minimum at τ^*, and thus it is single peaked.

□

This result is not surprising — it is the famous result that there are no distortions at the bottom and the top in the Mirrlees model. This result, as shown by Seade (1977), requires bounded support for the types, continuous type density, and interior allocations, which are all conditions satisfied in our setup.

Figure 3 illustrates Proposition 16 by plotting $x(\tau) = \mathcal{S}(\tau)/\tau$ for the second-best case and for the equilibrium with equal wealth (or any other conditionally efficient allocation). As stated in the proposition, x is U-shaped with the same values at both extremes. This is done for a particular numerical example where y is normally distributed, where τ is uniformly distributed, and where the utility function V is also CARA; see the parameter values indicated in the notes to the figure.

We next consider the marginal rates of substitution between shares and bonds for agents under the second-best and equilibrium allocations. We show that this marginal rate of substitution is equal to $\varphi'(x(\tau))$. Accordingly, we refer to $\varphi'(x(\tau))$ as the *shadow value of risk*. This shadow value of risk is the rate at which an agent perceives that he or she can trade bonds and shares at the margin under the optimal non-linear tax/subsidy scheme.

We let $\mathcal{M} = \{(S, B)\}$ be the menu of contracts offered to investors. Each point on the frontier of this set corresponds to the values of $B = \mathcal{B}(\tau)$ and $S = \mathcal{S}(\tau)$ for some $\tau \in [\tau_L, \tau_H]$, where the functions $\mathcal{S}(\cdot), \mathcal{B}(\cdot)$ are the solution to the mechanism design problem.

We compare the slope of the frontier of $\mathcal{M}$ with the slope of the budget line in the equal wealth equilibrium and in the optimal allocation. In the equal wealth equilibrium and in the optimal allocation, agents' marginal rate of substitution is constant across τ. In the equal wealth equilibrium, it is $dB^e/dS = -\bar{D}_1$. In the optimal allocation, it is $dB^*/dS = -\bar{C}_1$. Now consider any allocation defined by functions $\hat{x} : [\tau_L, \tau_H] \to \mathbb{R}$ and $\hat{\mathcal{B}} : [\tau_L, \tau_H] \to \mathbb{R}$ that are incentive compatible (IC) and feasible — in the sense that (72) and (71) hold. Define $\hat{\mathcal{M}}$ as the menu of contracts that decentralize the allocation $\hat{x}, \hat{\mathcal{B}}$. Note that $\hat{\mathcal{S}}(\tau) \equiv \hat{x}(\tau)\tau$. We have the following simple result that we use to evaluate agents' marginal rate of substitution between shares and bonds.

Lemma 6. *The slope of the frontier for $\hat{\mathcal{M}}$ is given by*

$$-\frac{d\mathcal{B}(\tau)}{d\mathcal{S}(\tau)} = \varphi'(\hat{x}(\tau)) \text{ for all } \tau \in [\tau_L, \tau_H] \tag{90}$$

$$-\frac{dB^e}{dS} = \varphi'(1/\bar{\tau}) < -\frac{dB^*}{dS} = \bar{\tau}\varphi(1/\bar{\tau}). \tag{91}$$

Proof. We now compute $d\bar{B}(\mathcal{S}(\tau))/dS = \mathcal{B}'(\tau)/\mathcal{S}'(\tau)$. We have $\mathcal{S}(\tau) = \tau x(\tau)$ so $\mathcal{S}'(\tau) = x(\tau) + \tau x'(\tau)$. Likewise, we have $\mathcal{B}(\tau) = \mathcal{C}'(\tau) - \varphi(\tau) - \tau\varphi(x'(\tau))x'(\tau)$. From the incentive constraints we have $\mathcal{C}'(\tau) = \varphi(x(\tau)) - x(\tau)\varphi'(x(\tau))$. Thus, combining them, we have

$$\frac{d\bar{B}(\mathcal{S}(\tau))}{dS} = \frac{\mathcal{B}'(\tau)}{\mathcal{S}'(\tau)} = \frac{-x(\tau)\varphi'(x(\tau)) - \tau\varphi(x'(\tau))x'(\tau)}{x(\tau) + \tau x'(\tau)}. = -\varphi'(x(\tau)).$$

For the equal wealth incomplete markets equilibrium, we have

$$S^e = 1 + \frac{\tau - \bar{\tau}}{\bar{\tau} + \bar{D}_1/\gamma} \text{ and } B^e = -[\tau - \bar{\tau}]\frac{\bar{D}_1}{\bar{\tau} + \bar{D}_1/\gamma} \text{ so } \frac{dB^e}{dS} = -\bar{D}_1 = -\varphi(1/\bar{\tau}).$$

For the complete markets equilibrium with the first-best allocation, we have

$$S^* = 1 + \frac{\tau - \bar{\tau}}{\bar{\tau} + \bar{C}_1/\gamma} \text{ and } B^* = -[\tau - \bar{\tau}]\frac{\bar{C}_1}{\bar{\tau} + \bar{C}_1/\gamma} \text{ so } \frac{dB^*}{dS} = -\bar{C}_1 = -\tau\varphi(1/\bar{\tau}),$$

where we use that in both the equal wealth and complete market equilibrium, x is constant, and hence it must be equal to $1/\bar{\tau}$. Also since the complete market allocation is conditionally efficient, then $x^*(\tau) = 1/\bar{\tau}$ for all τ. In the complete markets equilibrium allocation we have that $\bar{C}_1(\tau) = \tau\varphi(1/\bar{\tau}) + \bar{B}(\tau)$ is the same for all τ. Multiplying by $\mu(\tau)$, integrating it across τ, and using that uncontingent transfers have zero expected value across τ's, we have $\bar{C}_1 = \bar{\tau}\varphi(1/\bar{\tau})$. Finally, since φ is concave, and $\varphi(0) = 0$, then $\varphi(1/\bar{\tau}) > (1/\bar{\tau})\varphi'(1/\bar{\tau})$ or $\bar{C}_1 = \bar{\tau}\varphi(1/\bar{\tau}) > \varphi'(1/\bar{\tau}) = \bar{D}_1$. $\square$

Normal y case. In the case in which $y \sim N(\mu_y, \sigma_y^2)$, then $\varphi'(x) = \mu_y - \sigma_y^2 x$ is linear in x. In this case we have that a feasible and incentive compatible allocation $\hat{x}(\tau)$ must satisfy

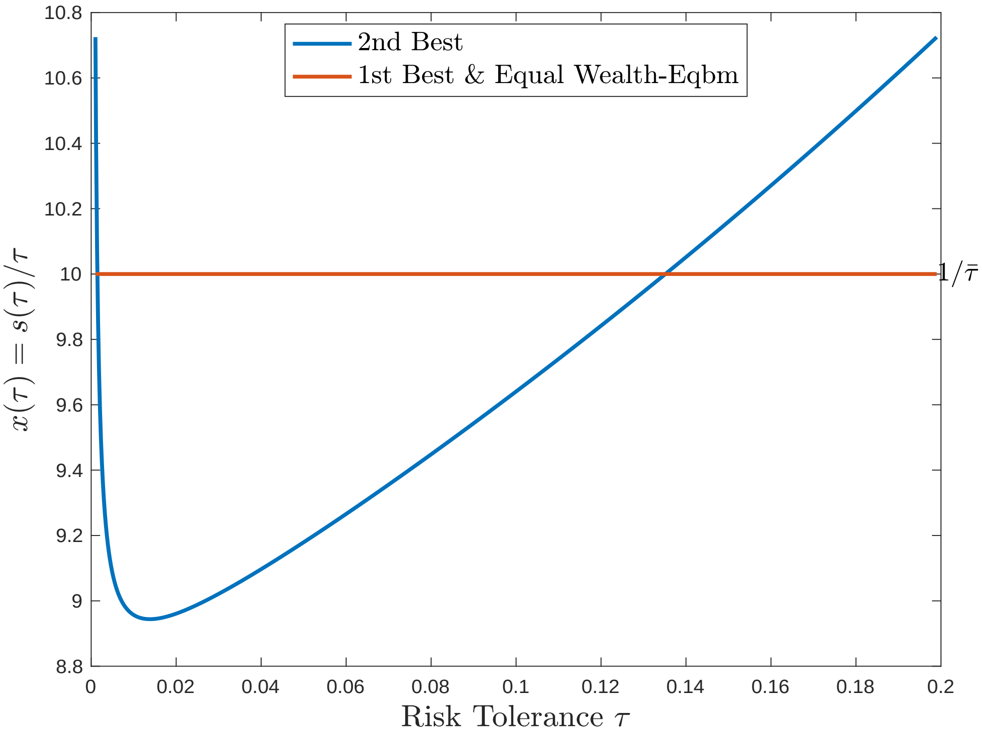
$$\int_{\tau_L}^{\tau_H} \frac{d\bar{B}(\tau\hat{x}(\tau))}{dS}\tilde{\mu}(\tau)d\tau = -\bar{D}_1 \text{ where } \tilde{\mu}(\tau) \equiv \frac{\tau\mu(\tau)}{\int_{\tau_L}^{\tau_H}\tau\mu(\tau)d\tau} \text{ for all } \tau \in [\tau_L, \tau_H]\,. \tag{92}$$

So, in the CARA-normal case, IC and feasibility imply that the *weighted average of the shadow value of risk* should be the same as the one in the equal wealth incomplete market equilibrium.

We illustrate the nature of the second-best allocation as well as its differences with a Tobin tax/subsidy in Figure 4. We do so for a CARA-normal example. The function V is also assumed to have the CARA form. The parameter values are listed in the notes to the figure. In this figure, we plot the value of the shadow value of risk $\varphi'(\hat{x}(\tau))$ for different allocations. In the case of the equal wealth incomplete market equilibrium allocation, this shadow value is constant for all τ and equal to $\varphi'(1/\bar{\tau})$. In the case of a Tobin subsidy to trade financed with an entry fee paid in terms of bonds, this shadow value is given by the declining step function shown in the figure. In that allocation, sellers of shares (with $\mathcal{S} < 1$) receive a selling price that is higher than the price paid by buyers of shares (with $\mathcal{S} > 1$) with the gap equal to the subsidy to trade ω. For the second-best allocation, we plot the shadow value of risk $\varphi'\left(\hat{\mathcal{S}}\right)$, as well as the average of this shadow value, by integrating the marginal shadow value from the value of τ_1 such that $\mathcal{S}(\tau_1) = 1$ to τ. Note that the pattern for the average prices in the second best "resemble" the prices for a Tobin subsidy in the sense that they are higher for "sellers" of shares (with $\mathcal{S}(\tau) < 1$) than for "buyers" of shares (with $\mathcal{S}(\tau) > 1$).

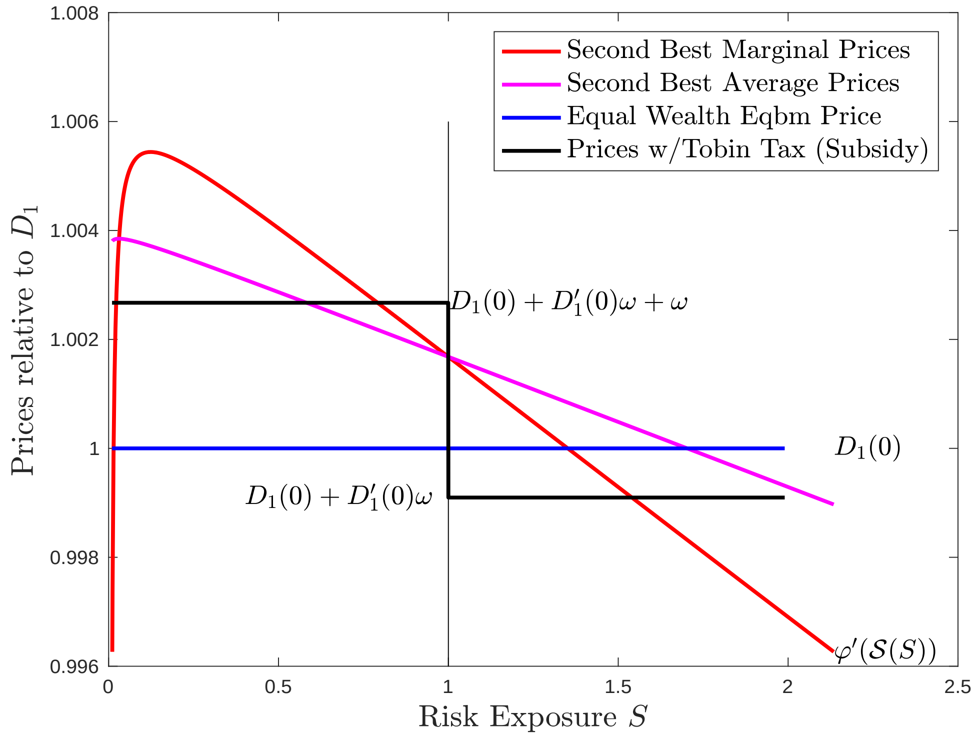
Finally, we consider the implications of our model for trade volumes measured as $\mathbb{E}|\mathcal{S}(\tau) - 1|$. Figure 5 plots the ratio of this statistic computed for the second-best allocation relative to the value of this statistic in the equal wealth equilibrium. The ratios are computed for several values of σ_y and the ratio τ_L/τ_H for the same specification as in the previous plots. As can be seen, trade volume in the second best is between 1% and 12% higher than in the equal wealth equilibrium as we compute both allocations for different parameter values.

Figure 3: Risk exposure, normalized relative to risk tolerance



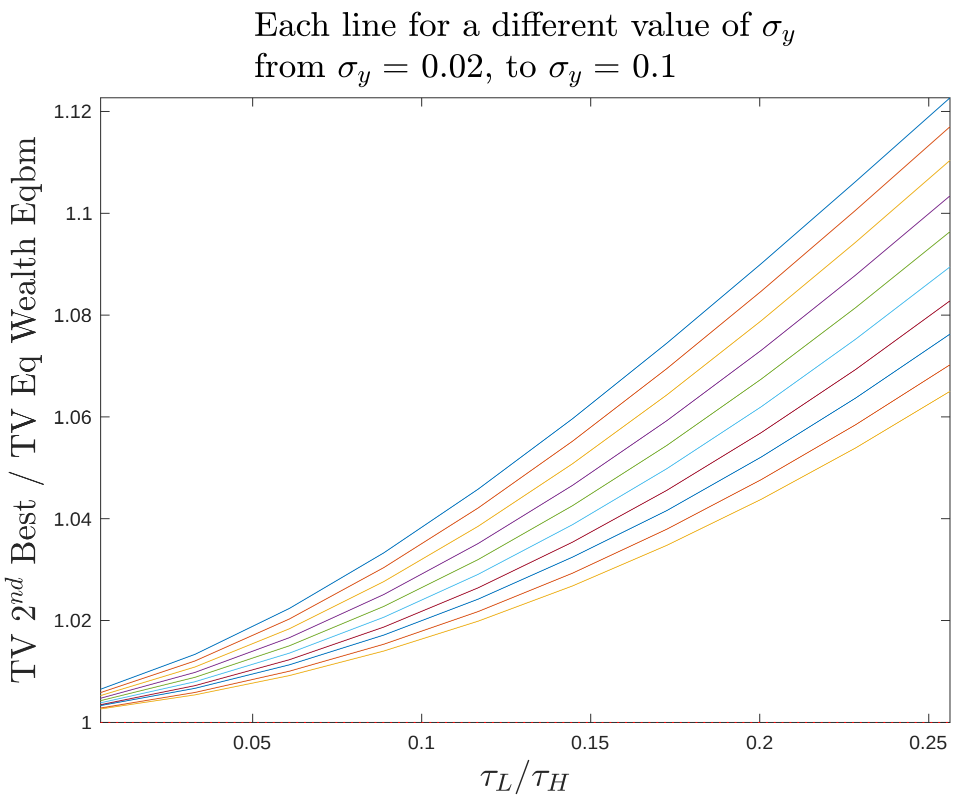
Notes: CARA-CARA-normal-uniform case.
Parameters $\mu_y = 1, \sigma = 0.07, \bar{\tau} = 1/10$, $\tau_L = 0.001, \tau_H = 0.199, \tau_V = \bar{\tau}$.

Figure 4: Marginal & Avg. Prices in the 2^{nd} Best, and Prices in Equal Wealth Eqbm



Notes: CARA-CARA-normal-uniform case.
Parameters $\mu_y = 1, \sigma = 0.07, \bar{\tau} = 1/10, \tau_L = 0.001, \tau_H = 0.199$,
$\tau_V = \bar{\tau}$, $(\omega/D_1(0)) \times 100 = -0.358(36 basispoints)$,
$TV^e = 0.2475$, $TV^{2^{nd}\ Best} = 0.2645$.

Figure 5: Trade Volume in the 2^{nd} Best relative to Trade Volume in Equal Wealth Equilibrium



Notes: CARA-CARA-normal-uniform case.
Parameters $\mu_y = 1, \sigma_y \in [0.2, 0.1], \bar{\tau}_H = 1/10$, $\tau_L \in [0.001, 0.51]$, $\tau_v = \bar{\tau}$.