Liquidity Discovery and Asset Pricing*

Michael Gallmeyer
Carnegie Mellon University

Burton Hollifield
Carnegie Mellon University

Duane Seppi
Carnegie Mellon University

First Draft: April 2003
Current Draft: June 3, 2004

*We thank Robert Engle, Paolo Fulghieri, Rick Green, Eric Hughson, Alfred Lehar, Spencer Martin, Ernst Maug, Maureen O'Hara, Bryan Routledge, Gideon Saar, Alex Stomper, Allan Timmermann, Stathis Tompaidis, Tan Wang, Robert Whitelaw, Christopher Yung, and seminar participants at Arizona State University, Carnegie Mellon University, the Commodity Futures Trading Commission, Humboldt University, the Instituto Tecnológico Autónomo de México, New York University, the Norwegian School of Management, the Sante Fe Institute, the Stockholm School of Economics, the University of Colorado at Boulder, the University of North Carolina, the University of Vienna, and the University of British Columbia Summer Finance Conference for helpful comments.

Abstract

Most investors purchase securities knowing they will resell those securities in the future. Uncertainty about the preferences of future trading counter-parties causes randomness in future resale prices that we call *liquidity risk*. It is natural to suppose that investors are asymmetrically informed about liquidity risk. Through a process of *liquidity discovery*, trading volumes and prices reveal private information about future counter-party preferences. The liquidity discovery process leads to endogenous joint dynamics for prices, trading volume, volatility, and expected returns. In particular, market liquidity is a forward-looking predictor of future liquidity risk and, as such, is priced. Liquidity discovery provides an alternative explanation to transaction costs for the relationships between current market liquidity measures and future returns.

Stocks and bonds are claims on streams of cash flows that continue long after the typical holding horizons of most investors. The large volume of trading observed empirically — with annual turnover rates of roughly 100 percent on the NYSE[1] — suggests that *marginal* investors hold long-dated securities expecting to re-trade in the future. When investors trade dynamically over time, asset prices depend, not only on the underlying future cash flows, but also on investors' perceptions of the likely future demands of trading counter-parties for those cash flows.

Uncertainty about the preferences of future counter-parties leads to randomness in future prices that we call *liquidity risk*. This is in addition to *cash flow risk*. We call preference-induced price randomness "liquidity" risk to highlight the fact that investors are uncertain about the terms-of-trade they will face — that is to say, the prices at which liquidity will be available — when re-trading cash flows in the future. Possible sources of randomness in preferences include endowment shocks (Constantinides and Duffie [1996]), exogenous habits (Campbell and Cochrane [1999]), stochastic risk aversion (Huang, Hughson, and Leach [2003]), future marginal tax rates, and uncertain subjective discount rates.

Personal preferences — and, hence, investors' future state-contingent asset demands — are unlikely to be common knowledge. If preferences change randomly over time, then investors are likely to learn their own preferences before these become known to others. Alternatively, even if investors know their own future preferences, this information might not be known in full by their counter-parties. In other words, investors' individual preferences may appear random to others, even if not to themselves. Both intuitions are consistent with the premise that investors know more about their own preferences than they do about other investors' preferences. In contrast to previous models of asset pricing with stochastic preferences, we study investors who are asymmetrically informed about each others' preferences.

The goal of this paper is to model the endogenous dynamics of liquidity risk when preference information is revealed through the trading process. Trading is a window into investor heterogeneity. Trades occur when low valuation investors transfer ownership of future cash flows to other higher valuation investors. A large volume of trading is prima facie evidence of time-varying investor heterogeneity and, hence, of liquidity risk.

[1]See the NYSE Fact Book Online at http://www.nysedata.com/factbook/main.asp.

Public trading plays an central learning role in revealing asymmetric information about future liquidity risk. Investors learn about their counter-parties' preferences by observing their trading decisions over time. Rational security demands today are linked, via optimal dynamic plans, to contingent demands for securities in the future. Hence, investors can use current prices and trades to update their beliefs about the prices at which they will find liquidity from willing counter-parties in the future. We call the process of learning about counter-party preferences *liquidity discovery.* Liquidity discovery reduces liquidity risk. Unlike learning about exogenous cash flows, however, the distribution of future prices is endogenous. There is a feedback effect. The demand for securities in the future depends, not only on investors' innate psychological predispositions, but also on the future wealth effects of their current portfolio decisions.

A key precondition for liquidity discovery is that investors are asymmetrically informed about systematic components in preferences. Idiosyncratic components of individual investors' preferences average out and do not affect asset prices in competitive markets. Asymmetric information about the preferences of heterogeneous investor clienteles is not implausible. Casual empiricism suggests there is often considerable uncertainty about the preferences of retail investors, financial institutions, and overseas investors. Alternatively, liquidity discovery could involve the preferences of strategic investors, such as central banks or large hedge funds.

We illustrate liquidity risk and liquidity discovery in a simple competitive model. Investors in our model differ in their holding periods. Short-horizon investors trade default-free long-dated bonds with long-horizon investors, but are uncertain about the long-horizon investors' future time preferences. Consequently, the short-horizon investors are uncertain about the price at which they will be able to trade bonds in the future. Our main results are

- Trading volume and the price impact of order flow are forward-looking predictors of future preferences and the distribution of future prices. Future preferences are fully or partially revealed via prior trading implying that the level of prevailing liquidity risk is endogenous.

- Correlation with future prices causes current trading to be correlated with interest rates and risk premia. In particular, trading and liquidity variables are priced.

- We give sufficient conditions for monotone and also non-monotone relations between prices, volumes, liquidity risk, and the liquidity risk premium. An appropriate choice of model parameters leads to co-movements of volume and bond/bill spreads resembling the "flight to quality" seen in the aftermath of the 1998 Russian bond default.

- Liquidity discovery can lead to endogenous price supports. Small surprises in trading volume can cause abrupt, even discontinuous crashes from one price support level to another if they lead to large changes in investors' beliefs about future security demands and prices.

- Unverifiable self-reports about preferences are not incentive compatible in competitive equilibria and, hence, neither reduce liquidity risk nor Pareto improve investor welfare.

Our model — with its focus on the interaction of order flows, prices, learning, and risk — is at the intersection of general equilibrium theory and market microstructure. In particular, liquidity discovery offers a new perspective on the role of the trading process in asset pricing as called for in O'Hara [2003]. In Duffie and Huang [1985], trading goes on continuously but reveals no information. Trading simply "digests" news that arrives exogenously from other sources by reallocating securities across investors. In our model, the trading process is both a mechanism for learning about non-public investor preferences as well as for reallocating ownership rights. Learning is central in the microstructure approach of Kyle [1985] and Glosten and Milgrom [1985] and also in the rational expectations model of Grossman and Stiglitz [1980]. However, uninformed investors only learn about cash flows; not about their informed counter-parties' preferences. Indeed, the informed investors' preferences play no direct role in asset pricing once the signal extraction problem is solved in Kyle [1985] and Glosten and Milgrom [1985] or in the uninformed investors' security demands once prices are set in Grossman and Stiglitz (1980).

Our analysis builds on seminal work by Grossman [1988] and Kraus and Smith [1989]. Grossman [1988] introduces the idea that investors' future trading plans are not common knowledge and that prices change, sometimes dramatically, as the flow of orders into the market reveals the latent strategies investors are following. Kraus and Smith [1989] presents the first formal model of liquidity discovery. Their model has multiple sunspot equilibria with differing levels of liquidity risk. We show that liquidity risk and liquidity discovery are not limited to situations with multiple equilibria. Indeed, liquidity risk arises as the unique equilibrium of our model. We provide a detailed description of fully and partially revealing equilibrium outcomes and derive a market risk premium for liquidity risk.

Other related work includes Kraus and Smith [1996, 1998] which use counter-party uncertainty to endogenize noise trading in a rational equilibrium model with asymmetric cash flow information. Smith [1993] develops an overlapping generations model with liquidity risk due to uncertainty about how many traders will arrive in the future but with symmetric information. Leach and Madhavan [1992] and Saar [2001] model dynamic learning by market makers about the distribution from which a sequence of investors are drawn. In contrast, we model learning about the same investors with whom one interacts repeatedly over time. Grundy and McNichols [1989] obtain multiple equilibria in a normal/exponential model of which one has dynamic trading with partial revelation of information. Vayanos [1999, 2001] models learning about a strategic uninformed investor who trades dynamically given a series of private endowment shocks. The endowment shocks change both the investors' preferences and also the aggregate risk in the economy. Our model, in contrast, focuses on pure preference learning effects. Finally, Detemple [2002] presents a partially-revealing dynamic equilibrium with private cash flow information and state-dependent preferences.

The market crash of 1987 prompted interest in how small variations in order flow can cause large changes in prices. Regions of high price/order flow sensitivity resemble transitions between adjacent price support levels. In Gennotte and Leland [1990] and Madrigal [1996], prices are discontinuous because of confusion about whether traders are informed about future payoffs or about supply shocks/noise trading. In Madrigal and Scheinkman [1997], prices can be discontinuous due to learning with heterogeneously informed agents. In all three papers, small changes in trading volume sometimes produce large revisions in expectations about future cash flows. In contrast, abrupt price sensitivities to order flow in our model are due to uncertainty about the endogenous prices at which current counter-parties will be willing to re-trade in the future.

Our paper is organized as follows. Section 1 presents a specific model of liquidity discovery given uncertainty about the time preferences of long-lived investors. Sections 2 and 3 discuss generalizations of liquidity risk and liquidity discovery and their empirical content. Section 4 concludes. All proofs are in the Appendix.

1 Liquidity discovery and time preferences

Consider a pure exchange economy with three trading dates $t = 1, 2, 3$. The sequence of events is in Figure 1. The only traded security is a two-period discount bond which is normalized to pay one unit of consumption at time 3. The bond has no cash flow risk because it is default-free. Let P_1 and P_2 be the prices of the bond at times 1 and 2. The bond proxies for the entire long-term bond market since our intent is to model systematic risk premia.

The motive for trading in this economy is intertemporal consumption smoothing between dates 1 and 2. Two groups of competitive investors trade with each other. The first is a continuum of identical long-horizon investors, denoted by the subscript L, with three-period preferences

$$u(c_{L1}) + \delta_1\, u(c_{L2}) + \delta_1 \delta_2\, u(c_{L3}) \tag{1}$$

where u is increasing, concave, differentiable, and satisfies the Inada conditions. The two subperiod time-preference parameters δ_1 and δ_2, for discounting between dates 1 and 2 and between 2 and 3 respectively, are known ex ante only to the long-horizon investors.[2] In equilibrium, the long-horizon investors face no price randomness. They know their time-preferences and there is no cash flow risk, so no expectations are necessary in (1).

The long-horizon investors have individual endowments of $e_{L1} \geq 0$ of the consumption good at date 1, $e_{L2} > 0$ units of consumption at date 2, and start out holding $\theta_{L0} > 0$ of the bond. The endowment e_{L2} cannot be traded directly at date 1. Markets for short-dated cash flows may be absent due to moral hazard or ex ante unverifiability of endowment ownership. The endowment structure and market incompleteness force investors to trade the long-dated bonds to shift consumption between dates 1 and 2. Let θ_{L1} denote the number of bonds the long-horizon investors hold per capita at date 1 and let θ_{L2} be their bond holdings at date 2. The long-horizon investors trade $\theta_{L0} - \theta_{L1}$ bonds to buy or sell additional consumption at date 1 and then trade $\theta_{L1} - \theta_{L2}$ bonds at date 2 to buy or sell consumption at date 2. Substituting the budget constraints in (1)

[2] Although the long-horizon investors' time preferences are not constant across time periods, their optimization problem is still time-consistent in contrast to Strotz (1956). Time consistency is preserved because our time preferences only change with the *calendar date*. The long-horizon investors' optimization problem one period ahead has the same relative time preference trade-off anticipated in their optimal plans one period before.

gives the portfolio problem for a generic long-horizon investor

$$\max_{\theta_{L1},\theta_{L2}} u(e_{L1} + P_1(\theta_{L0} - \theta_{L1})) + \delta_1\, u(e_{L2} + P_2(\theta_{L1} - \theta_{L2})) + \delta_1\, \delta_2\, u(\theta_{L2}). \tag{2}$$

The second group of investors is a continuum of identical short-horizon investors, denoted by the subscript S, who have expected utility preferences over consumption at dates 1 and 2

$$v(c_{S1}) + \beta\, \mathrm{E}_{S1}[v(\tilde{c}_{S2})] \tag{3}$$

where v is increasing, concave, differentiable, and satisfies the Inada conditions and E_{S1} denotes the short-horizon investors' expectations given the information available to them at date 1. The short-horizon investor's preferences are common knowledge. They have initial per capita endowments of $e_{S1} > 0$ units of consumption at time 1, $e_{S2} \geq 0$ of consumption at date 2 and $\theta_{S0} = 1 - \theta_{L0} \geq 0$ bonds. At date 1 they trade $\theta_{S1} - \theta_{S0}$ bonds to bring their total holdings to θ_{S1}. At date 2, since the short-horizon investors do not value consumption at date 3, they inelastically close out their bond position so that $\theta_{S2} = 0$. Substituting their budget constraints into (3) gives the portfolio problem for a generic short-horizon investor

$$\max_{\theta_{S1}}\; v(e_{S1} + P_1(\theta_{S0} - \theta_{S1})) + \beta\, E_{S1}\left[v\left(e_{S2} + \tilde{P}_2\; \theta_{S1}\right)\right]. \tag{4}$$

The price $\tilde{P}_2$ in (4) is potentially random from the short-horizon investors' perspective because of *liquidity risk*. As we show below, the equilibrium price P_2 depends on the long-horizon investors' second subperiod time-preference δ_2. A priori the short-horizon investors do not know δ_1 or δ_2. Rather they have priors over (δ_1, δ_2) given by a joint distribution with a bounded positive support

$$\mathrm{Prob}\,(\delta_1 \leq x, \delta_2 \leq y) \equiv F(x, y). \tag{5}$$

Our assumptions are consistent with the long-horizon investor having advance knowledge of her future stochastic preference or with deterministic preferences that are not common knowledge.

Short-horizon investors use market conditions at date 1 — prices P_1 and holdings θ_{L1} — to learn about the long-horizon investors' time-preferences and, hence, about P_2. In a rational expectations equilibrium, P_1 and θ_{L1} must satisfy first-order conditions calculated using the long-horizon investors' preferences. Thus, the updated posterior probability is zero for all preferences (δ_1, δ_2) for which the observed P_1 and θ_{L1} do not satisfy the associated first-order conditions. Let π be the short-horizon investors' probability distribution for P_2 given their updated beliefs about the long-horizon investors' preferences conditional on P_1 and θ_{L1}. We call this learning process *liquidity discovery*. The short-horizon investors are learning at date 1 about the future terms-of-trade they will face when trading bonds with the long-horizon investors at date 2. If trading at date 1 fully reveals δ_2, then the bond is riskless for the short-horizon investors between dates 1 and 2. If δ_2 cannot be inferred at date 1, then P_2 is uncertain and holding any non-zero bond position θ_{S1} exposes the short-horizon investors to liquidity risk. Thus, the prevailing level of liquidity risk is endogenous given liquidity discovery.

Long-horizon investors face no liquidity risk. From their perspective, the long-dated bond is riskless between dates 1 and 2 as well as between dates 2 and 3. In a sense, they are like the "informed" investors in Grossman and Stiglitz [1980] but with one significant difference. The component of future prices about which they are informed arises, not from private information about exogenous cash flows, but rather endogenously from advance knowledge about their own future behavior (i.e., their aggregate net demand at date 2 given their preferences). Trading at date 1 affects not only what "uninformed" short-horizon traders learn about P_2, but also the future price P_2 itself via the impact of the long-horizon investors' date 1 portfolio holdings on their date 2 trades.

1.1 Equilibrium

There are obvious potential welfare gains from dynamic trading in this environment. Different investors optimally hold different amounts of bonds over time as a vehicle to smooth any lumpiness in endowments. If short-horizon investors cannot perfectly anticipate the long-horizon investors' future bond demands, then liquidity risk distorts consumption smoothing. Our interest is in understanding how this friction affects investor welfare and the equilibrium dynamics of order flows, bond prices, and risk premia. A symmetric rational expectations equilibrium consists of:

- Posterior beliefs $\pi(P_2|\theta_{L1}, P_1)$ for the short-horizon investors about the distribution of the date 2 bond price P_2 conditional on trading at date 1

- Bond demand schedules $\theta_{L1}(P_1|\delta_1, \delta_2)$ and $\theta_{L2}(P_2|\theta_{L1}, \delta_2)$ for the long-horizon investors and $\theta_{S1}(P_1|\pi)$ for the short horizon investor given their respective information sets

- Bond prices $P_1(\delta_1, \delta_2)$ and $P_2(\theta_{L1}, \delta_2)$ given the long-horizon investors' realized preferences

that satisfy:

- Optimality: Both the long-horizon and short-horizon investors' portfolios are optimal given their information sets.

- Rational expectations: The short-horizon investors' beliefs about P_2 satisfy rational expectations.

- Walrasian market clearing: Prices equate the supply and demand for bonds at each date.

The equilibrium is most easily understood in terms of supply and demand. At date 1 the bond demand of the short-horizon investors is given by the first-order condition from (4)

$$v_c(e_{S1} + P_1(\theta_{S0} - \theta_{S1}))\, P_1 = \beta\, E_{S1}\left[v_c(e_{S2} + \tilde{P}_2\theta_{S1})\tilde{P}_2\right]. \tag{6}$$

At date 2, the short-horizon investors exit the market. They sell any bonds they are long to finance consumption or buy bonds to close out any outstanding short position.

The first-order conditions from (2) for the long-horizon investors

$$\begin{aligned} u_c(e_{L1} + P_1(\theta_{L0} - \theta_{L1}))\, P_1 &= \delta_1\, u_c(e_{L2} + P_2(\theta_{L1} - \theta_{L2}))\, P_2 && (7) \\ u_c(e_{L2} + P_2(\theta_{L1} - \theta_{L2}))\, P_2 &= \delta_2\, u_c(\theta_{L2}) && (8) \end{aligned}$$

pin down their net bond demands at dates 1 and 2. Market clearing requires that the long-horizon investors absorb the bonds traded by the short-horizon investors at date 2. Imposing market-clearing, $\theta_{L2} = 1$, and substituting (8) into (7), gives:

$$\begin{aligned} u_c(e_{L1} + P_1(\theta_{L0} - \theta_{L1}))\, P_1 &= \delta_1\delta_2\, u_c(1) && (9) \\ u_c(e_{L2} + P_2(\theta_{L1} - 1))P_2 &= \delta_2\, u_c(1). && (10) \end{aligned}$$

From equation (10), the price P_2 only depends on the net trade $\theta_{L1} - 1$ at date 2 and on the long-horizon investors' subperiod time-preference δ_2. Since the equilibrium net trade at date 2 is perfectly predictable given investors' positions at date 1, the only reason P_2 is random is because of uncertainty about δ_2. If δ_2 is large, then P_2 will be high. On the other hand, very low δ_2 realizations lead to "liquidity crises" in which liquidity is only available at very low prices P_2. This confirms our earlier claim about δ_2 as the source of liquidity risk in the model. The bond is risky for the short-horizon investors unless they can infer δ_2 from trading at date 1.

The key intuition in equation (9) is that P_2 does not enter the long-horizon investors' decision directly at date 1. The long-horizon investors know that, in equilibrium, P_2 will be set at date 2 so that the solution to (9) also solves (7) given their realized δ_2. Hence, the long-horizon investors' demand at date 1 only depends on their *cumulative* time-preferences $\delta_1\delta_2$ between dates 1 and 3, but not on P_2 or on δ_1 and δ_2 separately. This is the source of the difficulty for the short-horizon investors in learning P_2 from the long-horizon investors' portfolio choice at date 1.

Substituting the market-clearing price P_1 and volume $\theta_{L0} - \theta_{L1}$ into the date 1 first-order condition for the long-horizon investors (9) lets the short-horizon investors compute a summary statistic for the long-horizon investor's cumulative time-preference

$$z \equiv \frac{u_c(e_{L0} + P_1(\theta_{L0} - \theta_{L1}))\, P_1}{u_c(1)} = \delta_1\delta_2. \tag{11}$$

Liquidity discovery is based on the statistic z. The subperiod time-preference δ_2 is fully revealed if just one single δ_2 is possible given the observed z and the joint priors $F(\delta_1, \delta_2)$. Otherwise, δ_2 is not fully revealed by trading at date 1.

Lemma 1 *The short-horizon investors' equilibrium beliefs about the long-horizon investors' second-period time-preferences are* $Prob\,(\,\delta_2 \leq x \,|\, \theta_{L1}, P_1) = Prob\,(\,\delta_2 \leq x \,|\, \delta_1\delta_2 = z)$.

The short-horizon investors' beliefs π about P_2 follow from Lemma 1 and the second-period market clearing price function in equation (10). If either δ_1 or δ_2 have continuous distributions, then, as in Radner [1972], the equilibrium is not generically fully revealing. For technical reasons in proving existence, we assume the conditional distribution for δ_2 given $\delta_1\delta_2$ is discrete, but, otherwise, the distribution F can be any joint distribution with bounded support.

A useful distinction here is between *local preferences* and *global preferences.* Market data only reveal information about aspects of investor preferences that are relevant for the marginal demand and supply of securities under current market conditions. In general, prices and volumes at a single date need not reveal enough information to infer investor preferences under *all* possible future conditions. In our model, the statistic $z = \delta_1 \delta_2$ summarizes the long-horizon investors' local preferences as they matter for market clearing at date 1, but it is not always possible to separately identify δ_1 and δ_2, the parameters that determine the long-horizon investors global preferences.

Lemma 2 *The set of long-horizon investor types that pool in equilibrium is independent of the other parameters of the economy:* (δ_1', δ_2') *pools with* (δ_1, δ_2) *if and only if* $\delta_1' \delta_2' = \delta_1 \delta_2$.

Existence of equilibrium is established by showing that, for any possible realization of the long-horizon investors' time-preferences, the market-clearing conditions can be solved.

Proposition 1 *A symmetric rational expectations equilibrium always exists under the stated assumptions concerning preferences and endowments. If the long-horizon investors have an elasticity of intertemporal substitution* $-\frac{u_c(x)}{x\, u_{cc}(x)}$ *strictly greater than one, the equilibrium is unique.*

The restriction on the elasticity of intertemporal substitution (IES) ensures that consumption at date 2 is a normal good for the long-horizon investors. In this case, the long-horizon investors' bond supply curve slopes up in P_1 and can cross the short-horizon investor's demand curve only once.

Analytic example. The equilibrium can be computed in closed-form given logarithmic preferences $u(c) \equiv v(c) \equiv \ln(c)$.[3] This lets us illustrate the workings of the model concretely. The intuitions are then formalized in Section 1.2. In addition, we assume endowments $e_{L1} = 0$ and $\theta_{L0} = 1$. In this special case, the long-horizon investors always sell bonds at date 1 and, moreover, their net trade does not depend on the bond price P_1. From equations (9) and (10), the date 1 net trade is

$$1 - \theta_{L1} = \frac{1}{\delta_1 \delta_2} \tag{12}$$

[3] The log preference equilibrium is unique even though the intertemporal elasticity of substitution for log preferences $-\frac{u_c(x)}{x\, u_{cc}(x)} = 1$. The strict inequality condition in Proposition 1 is sufficient but not necessary for uniqueness.

and the date 2 market-clearing bond price is

$$P_2 = \frac{\delta_1 \delta_2 \, e_{L2}}{1+\delta_1}. \tag{13}$$

Given (12), the short-horizon investors can infer $\delta_1\delta_2$ from the date 1 trading volume alone without reference to P_1. Letting $E[\cdot|\delta_1\delta_2]$ denote the short-term investor's expectations conditional on $\delta_1\delta_2$ using Lemma 1, the equilibrium bond price at date 1 is

$$P_1 = \frac{e_{S1} e_{L2} \, \delta_1 \delta_2 \, E\left[(e_{S2}(1+\delta_1) + e_{L2})^{-1} \middle| \, \delta_1 \delta_2 \right]}{\frac{1+\beta}{\beta} - e_{S2} \, E\left[\left(e_{S2} + \frac{e_{L2}}{1+\delta_1} \right)^{-1} \middle| \, \delta_1 \delta_2 \right]}. \tag{14}$$

In order to compute a risk premium for liquidity risk, we also calculate the short-horizon investor's shadow price for a one-period bill paying one unit of risk-free consumption at date 2

$$b_1 = \frac{e_{S1} \, E\left[(e_{S2} + \frac{e_{L2}}{1+\delta_1})^{-1} \middle| \, \delta_1 \delta_2 \right]}{\frac{1+\beta}{\beta} - e_{S2} \, E\left[\left(e_{S2} + \frac{e_{L2}}{1+\delta_1} \right)^{-1} \middle| \, \delta_1 \delta_2 \right]}. \tag{15}$$

Including a *tradable* one-period bill along with the long-dated bond would eliminate all liquidity risk in this simple setting. With two securities and only two sources of preference uncertainty, the short-lived investors could generically recover both δ_1 and δ_2. Liquidity risk with multiple securities requires multiple dimensions of preference uncertainty.[4] An attraction of our simple model is that it leads to a particularly tractable illustration of liquidity risk.

With log preferences, it is more intuitive to work with trading volume $1 - \theta_{L1} = 1/z$ rather than directly with z. Indeed, the model has a natural market microstructure interpretation. The inelastic bond supply from the long-horizon investors at time 1 can be viewed as a market order. Similarly, the short-horizon traders' first-order condition at date 1 characterizes a liquidity supply schedule giving the prices P_1 at which they are willing to absorb different quantities $1/z$ of bonds.

Figures 2a and 2b illustrate a partially revealing equilibrium outcome. Suppose that, given the realized date 1 volume $1/z$, only two (δ_1, δ_2) pairs are possible in the sense that $\text{Prob}(\delta_1\delta_2 = z) > 0$.

[4] Cash flow risk or some other type of randomness is also needed to avoid arbitrage by the better-informed long-horizon investors between the otherwise riskless bills and bonds.

Denote these as (δ_1^I, δ_2^I) and (δ_1^P, δ_2^P) where $\delta_2^I < \delta_2^P$ and $\delta_1^I > \delta_1^P$. We say the long-lived investors are *impatient* at date 2 if δ_2 is δ_2^I and that they are *patient* if δ_2 is δ_2^P. Figure 2b shows supply and demand at time 2. Since the short-horizon investors inelastically sell all of their bonds, the supply curve is vertical. The two downward-sloping curves are the long-horizon investor's demands conditional on being patient δ_2^P or impatient δ_2^I. Hence, two prices are possible at date 2.

Figure 2a shows supply and demand at time 1. The solid vertical line is the supply curve of the long-horizon investors. It is inelastic due to logarithmic utility. Since the volume $1/z = 1/\delta_1^I\delta_2^I = 1/\delta_1^P\delta_2^P$ does not fully reveal the long-horizon investors' type, the short-horizon investors cannot tell whether they are trading with a patient investor (who will pay a high price for bonds at date 2) or an impatient investor (who will pay a low price). The solid downward sloping line is the short-horizon investors' demand given their uncertainty about the long-horizon investors' type. For comparison, the upper dashed (- - -) curve shows the short-horizon investors' hypothetical demand if date 1 trading instead fully reveals that the long horizon investors will be patient. In this case the short-horizon investors know with certainty they will receive the higher price $P(\delta_2^P, 1/z)$ when reselling the bonds at date 2. The lower dash-dot-dot (- · ·) curve is their demand if they know they are trading with impatient δ_2^I investors. Prices on the partially revealing (solid) curve are not the expectation of the fully revealing prices. There is a risk premium due to the short-horizon investors' risk aversion. Hence, there is an expected welfare loss due to liquidity risk and the resulting less efficient consumption smoothing.

Example 1 Suppose that δ_2 is binomial with a state space $\{\delta_2^I, \delta_2^P\}$ where patient and impatient investors are equally likely, $\text{Prob}(\delta_2^I) = \text{Prob}(\delta_2^P) = 0.5$. For the patient investors, $\delta_2^P = 1.5$ and δ_1^P is uniformly distributed on the interval [0.95, 1.4]. For the impatient investors, $\delta_2^I = 1.1$ and δ_1^I is uniformly distributed on [1.3, 1.55]. The short-horizon investors' time-preference is $\beta = 1.2$.[5] The short-horizon investors have per capita endowments of $e_{S1} = 1.4$ and $e_{S2} = 0.6$ and the long-horizon investors' date 2 endowments are $e_{L2} = 1.0$.

Given these assumptions, the support [1.43, 1.71] for $z = \delta_1\delta_2$ conditional on δ_2^I is nested and "left justified" inside the support [1.43, 2.1] for z conditional on δ_2^P. This implies there is a critical volume $1/z' = 1/1.71$ such that the equilibrium outcome is fully revealing when the realized volume

[5] "Discount" factors need not be less than 1 in a finite horizon setting.

$1/\delta_1\delta_2 < 1/z'$ and only partially revealing when $1/\delta_1\delta_2 > 1/z'$. In particular, there are δ_1^Is and δ_1^Ps such that both the patient and impatient types sell more than $1/z'$ bonds, but only patient δ_2^P types sell less than $1/z'$. The probability of a non-fully revealing outcome here is 60 percent.

The top left plot in Figure 3 plots the bond price P_1 and expected price $\mathrm{E}_{S1}[P_2]$ versus the trading volume $1/z$ at date 1. Small trades $1/z < 1/z'$ fully reveal that the long-horizon investors are the patient δ_2^P type. In this case, short-horizon investors know, from (13), that P_2 will be $\frac{e_{L2}}{1/z+1/\delta_2^P}$ with certainty and, hence, are willing to pay a high price P_1 for the bond at time 1. Prices are decreasing in volume because more selling at date 1 means the long-horizon investors must buy back more bonds at date 2 which depresses P_2. Since the date 2 net trade $1-\theta_{L1}$ is predictable given public information at time 1, the "market overhang" effect at date 2 is fully anticipated when P_1 is set. At the critical volume $1/z'$, both P_1 and $\mathrm{E}_{S1}[P_2]$ have discrete jumps downward because volumes $1/z > 1/z'$ are consistent with δ_2^P or δ_2^I. Now there is liquidity risk.

The interplay between trading and liquidity discovery is particularly stark in this example due to the inelasticity of the logarithmic long-horizon investors' net demand at date 1 and the uniform/binomial distribution over the subperiod time-preferences (δ_1, δ_2). Later examples relax these restrictive features.

Turning to the pricing of liquidity risk, the top right plot in Figure 3 shows that the standard deviation of the future bond return is roughly 10 percent in non-fully revealing states. The bottom left plot in Figure 3 contrasts the shadow risk-free return on one-period bills and the expected return on two-period bonds. The two are equal when the equilibrium outcome is fully revealing, but there is a small but non-trivial liquidity premium when the outcome is only partially revealing. The bottom right plot in Figure 3 plots the liquidity premium against the volume of trade. The liquidity premium is roughly 30 basis points in the partially revealing equilibrium outcomes.

The large swing in the shadow one-period interest rate in Figure 4 follows from the changing consumption allocations associated with different levels of trading at date 1. Low volumes $1/z$ fully reveal the long-horizon investors are patient. As long-horizon investor sell more bonds, short-horizon investors forgo more date 1 consumption and have more date 2 consumption. This raises both the one-period shadow rate and the bond return. At the critical volume $1/z'$ the likelihood that the long-horizon investors are impatient jumps up. Consequently, the date 1 consumption of the short-horizon investors increases discontinuously — since they now pay less for essentially the same number of bonds — and their expected date 2 consumption falls — since impatient investors, on average, pay less to buy back the bonds — thereby causing the one-period shadow rate to fall. Indeed, the interest rate effect in this example is so strong that the total expected bond return actually falls despite the now positive risk premium. ∎

The patterns of prices, returns, and volumes in Example 1 offer a stylized explanation for "flights to quality" such as followed the 1998 Russian default.[6] Consider the market for non-Russian long-dated debt. We interpret our short-horizon investors as short-term liquidity providers such as investment bank trading desks who do not initially hold long-dated bonds. Our long-horizon investors are pension plans and mutual funds who do hold bonds. The trading desks know the pensions and mutual funds will react to major bond defaults by initially scaling back their bond portfolios due to some combination of short-term liquidity needs and updated longer-term views about bond fundamentals. In particular, long-horizon investors may emerge fundamentally more bearish about long-dated debt (i.e., impatient at date 2 in our model) leading to low future demand for bonds. Alternatively, they might stay fundamentally bullish (i.e., patient) in which case their future bond demand will bounce back once the Russian crisis passes.

A small bond sell-off fully reveals the long-horizon investors are fundamentally bullish and that the sell-off is solely due to transitory liquidity needs. However if the sell-off is large, then short-horizon investors cannot tell whether the long-horizon investors are bullish with large short-term liquidity needs or bearish with small short-term liquidity needs. Short-term bill prices rise because of the shift between current and future expected consumption. Spreads between short-term treasuries and long-dated corporate bonds rise, not due to new information about future bond cash flows, but rather due to the liquidity risk at time 2 given the uncertainty about the future bond preferences of the long-horizon investors.

1.2 General Properties

The specific co-movements of prices, returns, volatility, and risk-premia with volume follow largely from how the supports for the cumulative time-preferences $\delta_1\delta_2$ overlap and are ordered conditional on different values of δ_2. The roughly monotone pattern in Example 1 — up to the volume

[6] A similar story can be told with cash flow risk and uncertainty about investors' willingness to bear credit risk.

"overhang" effect — is an immediate consequence of the particular nesting assumption. It is easy to construct priors $F(\delta_1, \delta_2)$ for which the predicted relationships are non-monotone or in which the correlations are reversed. In addition, price supports in the volume/price relation can still arise with continuous priors rather than the discontinuous uniform priors.

Proposition 2 *If, given the priors $F(\delta_1, \delta_2)$, the sets of $\delta_1\delta_2$ products associated with multiple possible values of δ_2 strictly nest (or are nested by) the sets of $\delta_1\delta_2$ products uniquely associated with single values of δ_2, then the resulting prices P_1, expected prices $\mathrm{E}_{S1}[P_2]$, expected bond returns, return volatilities, and liquidity risk premia can be non-monotone in z.*

Proposition 3 *Abrupt changes in the conditional density $f(\delta_1|\delta_2)$ for some possible values of δ_2 lead to abrupt changes in prices, spreads, volatility, and risk premia relative to z.*

Example 2 As an illustration of Proposition 3, we keep log preferences but change the short-term investors' priors over δ_1 to a beta (2,2) distribution scaled to cover the same conditional supports for δ_1^P and δ_1^I as in Example 1. Otherwise, the parameters are unchanged. Figure 4 shows the two conditional beta densities and the resulting likelihood of the patient type given different realizations of the summary statistic z. Figure 5 has the same layout as Figure 3. The market at date 1 still has an endogenous "price support" level as can be seen in the top left plot in Figure 5. The only difference is that, with beta priors, the transition between the fully revealing and the pooling outcomes is continuous; unlike the discontinuous jump with the uniform priors in Figure 3. ■

Volume alone is not always sufficient to compute the summary statistic z as in our log examples. From (11), prices are, in general, also needed for liquidity discovery.

Example 3 Consider long-horizon and short-horizon investors with constant relative risk aversion utility $u(c) = c^{1-\gamma}/(1-\gamma)$. The long-horizon investors have square root utility with a coefficient of relative risk aversion $\gamma = 0.5$. The short-horizon investors have power utility with $\gamma = 3$. The other parameters are given in Figure 6. We see that prices are needed for liquidity discovery in the top left plot of Figure 6 where there are multiple prices for some volumes.[7] From the long-horizon investor's

[7] It is also possible to construct environments in which both P_1 and θ_{L1} are needed to compute z. This happens with square root preferences when there are fully revealing (δ_1^P, δ_2^P) pairs such that $\delta_1^P\delta_2^P < \delta_1^I\delta_2^I$ for some second period preferences $\delta_2^P > \delta_2^I$.

first-order condition (9), the summary statistic $z = \delta_1\delta_2 = \sqrt{P_1/(1-\theta_{L1})}$. Unlike the inelastic logarithmic bond supply, the bond supply at date 1 is now increasing in P_1. Rather than a critical volume $1/z'$, there is a minimal (left most) positively-sloped bond supply schedule for impatient investors. Long-horizon investors who trade price/volume pairs $(P_1, 1-\theta_{L1})$ on the upper part of the short-horizon investors demand schedule to the left of the minimal impatient bond supply curve are fully revealed to be the patient type. In the top left plot of Figure 6, the right-most point on the upper demand curve at date 1 (used when trading with fully revealed patient investors) and the left most point on the lower demand curve (when the long-horizon investor's type is not fully revealed) are connected by this minimal bond supply schedule. ■

Investors' preferences affect prices and risk premia through two channels. The first is a direct effect holding the trading volume at date 2 fixed. Perturbations of the long-horizon investors' future time-preferences δ_2 raise or lower their future demand schedules and, hence, change bond prices at date 2. This, in turn, affects the supply and demand for bonds at date 1 and, hence, the bond price P_1. The second channel is an indirect feedback effect. Perturbations in the prices at date 2 can alter the short-horizon investors' equilibrium bond holdings at date 1. Consequently, when the short-horizon investors liquidate their perturbed bond position at date 2, the market clears at a different volume *along* the long-horizon investors' date 2 demand curve. Simply knowing that the entire demand curve is higher does not guarantee, for example, the market-clearing price P_2 is higher. If the perturbed prices at date 2 cause the short-horizon investors to increase their bond holdings at date 1, then P_2 is set further down a higher demand curve. Hence, the net effect can be ambiguous depending on the size and direction of the direct and indirect effects. This, in turn, depends on substitution and wealth effects in investors' choices for date 1 and 2 consumption and their net effect on bond demand.

We consider several preference perturbations and their impact on the distribution of prices and trading volume. The first two are comparative statics for the mean and volatility of the distribution of the long-horizon investors' time-preference δ_2. We use perturbations of δ_1 and δ_2 such that the long-horizon investors' cumulative time-preference $\delta_1\delta_2$ is unchanged. We call such perturbations *local demand invariant* (LDI) since, from Lemma 2, they leave the long-horizon investor's bond supply curve unchanged at date 1. In this discussion we restrict ourselves to cases in which the short-horizon investor is *long* bonds at date 1, implying a positive premium for liquidity risk. We also assume that the long-horizon investors' preferences satisfy the intertemporal elasticity of substitution (IES) restriction $-\frac{u_c(x)}{x\,u_{cc}(x)} > 1$ so that their date 1 bond demand is monotone.

Proposition 4 *Consider a LDI perturbation which increases each δ_2 to a new $\delta_2' > \delta_2$. Assume the IES restriction $-\frac{u_c(x)}{x\,u_{cc}(x)} > 1$ is satisfied for the long-horizon investors and that the short-horizon investors are long bonds, $\theta_{S1} > 0$, in the unperturbed equilibrium. If date 2 consumption is not a Giffen good for the short-horizon investor, then the short-horizon investors' perturbed bond holdings increase, $\theta_{S1}' > \theta_{S1}$, and the perturbed date 1 price increases, $P_1' > P_1$.*[8]

- *If the increase in θ_{S1}' is sufficiently small, then for each δ_2' the perturbed prices P_2' are pointwise higher than the corresponding unperturbed prices P_2.*

- *If the increase in θ_{S1}' is sufficiently large, then the perturbed prices P_2' are pointwise lower then the unperturbed prices P_2. In addition, given $P_1' > P_1$, the expected return falls.*

Figure 7 illustrates the intuition. Although the perturbed demand curve is strictly higher, the perturbed market-clearing price P_2' is above or below the unperturbed price depending on the size of the short-horizon investors' optimal bond position. Thus, learning about endogenous future prices — which are affected by investors' actions at date 1 — has subtleties not present when learning about purely exogenous future cash flows. Next, consider perturbing the volatility of future preferences.

Proposition 5 *Consider a LDI volatility increasing spread in the distribution of δ_2 such that the long-horizon investors' expected willingness-to-pay at date 2 is unchanged at the unperturbed volume. Assume that the IES restriction $-\frac{u_c(x)}{x\,u_{cc}(x)} > 1$ is satisfied for the long-horizon investors and that the short-horizon investors are long bonds in both the original and the perturbed equilibria, $\theta_{S1} > 0$ and $\theta_{S1}' > 0$.*

- *If the short-horizon investors' precautionary savings motive is sufficiently weak, then the*

[8] If date 2 consumption is a Giffen good for the short-horizon investors, then the arguments in the proof of Proposition 4 can be used to show that the short-horizon investors' bond holdings decrease $\theta_{S1}' < \theta_{S1}$ at date 1. Consequently, each perturbed P_2' price is pointwise higher than the corresponding unperturbed P_2 price. In addition, the perturbed P_1' price is lower than the unperturbed P_1 price so that the expected return increases unambiguously.

short-horizon investors' bond holdings θ'_{S1} are lower which, in turn, increases the expected price $\mathrm{E}_{S1}[P'_2]$, lowers the initial price P'_1, and increases the expected return.

- *If the short-horizon investors' precautionary savings motive is sufficiently strong, then θ'_{S1} is higher which, consequently, reduces $\mathrm{E}_{S1}[P'_2]$, raises P'_1, and reduces the expected return.*

The short-horizon investors' preferences also influence expected returns. The relation between their time preferences β and asset pricing is intuitive.

Proposition 6 *Consider short-horizon investors who become more patient in that $\beta' > \beta$. Assume the IES restriction $-\frac{u_c(x)}{x\, u_{cc}(x)} > 1$ holds for the long-horizon investors. The date 1 price increases, $P'_1 > P_1$, and the expected date 2 price decreases, $\mathrm{E}_{S1}[P'_2] < \mathrm{E}_{S1}[P_2]$. Hence, the expected bond return falls.*

The relation between short-horizon investor risk aversion and asset pricing is more complicated. Intuitively, short-horizon investors with greater risk aversion might be expected to require higher risk premia — the expected bond return less the return on the shadow one-period bill — for markets to clear. In general, however, changing the short-horizon investors' risk aversion leads to income and substitution effects for both the long-horizon and the short-horizon investors. In the special case of log preferences and a zero initial long-horizon endowment $e_{L1} = 0$, the liquidity premium can be shown analytically to be monotonically increasing in the risk aversion parameter γ of short-horizon investors with CRRA utility. This is because in this special case θ_{L1} and the distribution of date 2 bond prices are independent of the short-horizon investor's preferences. With square-root long-horizon preferences it is possible to construct examples in which this relation is non-monotone. This is illustrated in Figure 8 which varies the short-horizon investors' risk aversion and plots the corresponding equilibrium liquidity premium holding z and all other parameters fixed.

Liquidity risk prevents investors from achieving first-best consumption smoothing. It is natural to wonder whether the long-horizon investors can improve their welfare by simply announcing their type before the first round of trade.

Proposition 7 *Assume the IES restriction* $-\frac{u_c(x)}{x\, u_{cc}(x)} > 1$ *holds for the long-horizon investors. Unverifiable statements by long-horizon investors about their preferences are generically not credible in a competitive equilibrium.*

Non-verifiable statements generically do not lead to a separating equilibrium. In a competitive market, long-lived sellers who trade at low prices P_1 if their type is revealed at date 1 would benefit by being confused with investors who trade at higher bond prices at date 1. The reverse is true of long-lived buyers. As a consequence, long-horizon investors cannot avoid liquidity risk by simply announcing their type, unless their announcements are verifiable.

2 Generalizations

The essential building blocks of liquidity discovery are randomness in future preferences, investor heterogeneity, and asymmetric preference information. To understand the role of each in a general theory of liquidity risk and liquidity discovery, write the market price of a generic security j at time t as $p_{jt} = \mathrm{E}_t[m_{t+1}(c_{jt+1} + p_{jt+1})]$ where the pricing kernel m_{t+1} reflects the preferences of the current marginal investor for next period's dividend (or coupon) c_{jt+1} and p_{jt+1} is the market price at date $t+1$. Using this identity recursively, the price at date t can be expressed in terms of a sequence one-period pricing kernels $m_{t+1}, m_{t+2}, \ldots$ and cash flows $c_{jt+1}, c_{jt+2}, \ldots$

$$p_{jt} = \mathrm{E}_t\left[m_{t+1}c_{jt+1} + m_{t+1}m_{t+2}c_{jt+2} + \ldots + \left(\prod_{s=t+1}^{T} m_s \right) c_{jt+2} + \ldots \right]. \tag{16}$$

The future pricing kernels are random, not only due to random aggregate consumption, but also because of random preference factors affecting investors' future marginal utilities. This is liquidity risk.

Investor heterogeneity means that the future realizations of investors' different preferences result in trading as well as price changes. Investors with high valuations buy securities from investors with low valuations.[9] Heterogeneity is likely to exacerbate liquidity risk. While next period's cash flows are valued using the preferences of the current marginal investors, as impounded in m_{t+1},

[9] Detemple and Selden [1991] and Vanden [2004] study investor heterogeneity through the prices of zero net supply derivatives. Jointly modelling prices and trading of positive net supply securities is an alternative approach. Our approach also differs from their's in that we have asymmetric information.

more distant cash flows are valued using the anticipated preferences of *future marginal* investors, as impounded in m_{t+2}, $m_{t+3}, \ldots$, about which the current marginal investors are less informed.

Liquidity discovery occurs when investors are asymmetrically informed about each others' future preferences. Trading is then potentially informative about the pricing kernels of future marginal investors as investors rebalance their portfolios over time. With liquidity discovery, the liquidity risk in the market changes endogenously with the amount of preference information revealed by the trading process. As a consequence, prices, volumes, future price volatility, and expected returns co-move.

Liquidity risk and liquidity discovery are robust in more complex economic environments. Anytime investors are asymmetrically informed about each others' preferences, then trading can be informative about counter-parties' preferences and, hence, about prices in the future. We discuss several extensions below.

Market incompleteness. A necessary condition for liquidity risk is that markets are statically incomplete. This forces investors to use dynamic trading strategies. Adding *tradable* one-period bills in our model, for example, fully reveals P_2 and, thereby, eliminates liquidity risk at date 1. Liquidity risk can still occur with multiple traded securities provided the dimensionality of the future preference uncertainty is large enough. Our model only considers two-dimensional uncertainty (concerning δ_1 and δ_2) about a single homogenous group of investors. In practice, the preferences of the marginal investor are likely to be complex multi-dimensional objects including risk aversion, expected holding periods, tax rates, labor income or future endowments (as in Vayanos [1999, 2001]), changing family circumstances, heterogeneous beliefs about cash flows and, indeed, the cross-sectional distribution of different heterogeneous investor types in the economy.

Trading constraints. Short sale constraints and other trading restrictions are natural responses to moral hazard and unverifiable consumption endowments, but they can complicate liquidity discovery by causing the trades of more investors to pool. In our model a short-sale constraint can lead to pooling, not just by investors with the same cumulative time preferences $\delta_1\delta_2$, but also with different $\delta_1\delta_2$s. This is easily illustrated in Example 1 by changing the supports for δ_1^I and δ_1^P so that the conditional support for $\delta_1^P\delta_2^P$ is [0.9, 2.1] and the support for $\delta_1^I\delta_2^I$ is [0.6, 0.85]. In the absence of short sale restrictions the equilibrium is fully revealing. The support of the unrestricted patient volumes [1/2.1, 1/0.9] is strictly below the support for impatient volumes [1/0.85, 1/0.6]. With a short sale constraint $\theta_{L1} \geq 0$, however, some patient investors with unrestricted volumes [1, 1/0.9] are now forced to pool with all of the impatient volumes at $1 - \theta_{L1} = 1$. As a consequence, liquidity discovery is impeded. Now there is liquidity risk.

Strategic counter-parties. Liquidity risk is even more natural in markets with strategic counter-parties. For example, the general public may be uncertain about the policy preferences of large central banks in currency and bond markets or about the reservation prices of major shareholders in particular stocks.

Multiple rounds of trade. Investors learn about each other as they interact repeatedly over multiple rounds of trade. The linearity of our time-preference learning problem keeps the analysis tractable. An immediate extension of our model has T periods with a sequence of overlapping generations of short-lived investor cohorts arriving at dates $t = 1, \dots T-2$ and one group of long-lived investors. At each date $t < T-1$ there are two cohorts of short-lived investors, one young (labelled with a subscript t) and one old (labelled as $t-1$). At date $T-1$ the last cohort leaves, but no new cohort arrives. Each cohort t is endowed with $e_{t,t} > 0$ of consumption at date t when they are young and $e_{t,t+1} \geq 0$ of consumption at $t+1$ when they are old. When the short-lived investors in cohort t arrive, the bond is held by the long-lived investors and the departing investors of cohort $t-1$. The new short-lived cohort choose their bond holdings $\theta_{t,t}$ at date t to solve

$$\max_{\theta_{t,t}} \; v(e_{t,t} - P_t \theta_{t,t}) + \beta \, E_t \left[v \left(e_{t,t+1} + \tilde{P}_{t+1} \theta_{t,t} \right) \right]. \tag{17}$$

The long-lived investors start with endowments of consumption $e_{L1}, \dots, e_{LT}$ and $\theta_{L0} = 1$ of the bond and choose bond holdings $\theta_{L1}, \dots, \theta_{LT-1}$ over time to solve

$$\max_{\theta_{L1},\dots,\theta_{LT-1}} u(e_{L1} + P_1(1 - \theta_{L1})) + \dots + \prod_{j=1}^{t} \delta_j \, u(e_{Lt} + P_t(\theta_{Lt-1} - \theta_{Lt})) + \dots + \prod_{j=1}^{T} \delta_j \, u(e_{LT} + \theta_{LT-1}). \tag{18}$$

As in the three-period model, the long-horizon investors' endowments and the short-horizon investors' preferences and endowments are common knowledge. Long-horizon investors know the sequence of their time preferences $\delta_1, \dots, \delta_{T-1}$, but the short-horizon investors only have priors over the long-horizon investors' preferences $F(\delta_1, \ldots, \delta_{T-1})$.

The critical variable here is the long-lived investors' cumulative time-preference. The first-order conditions to the long-horizon investor's problem

$$\begin{aligned} P_1\, u_c(e_{L1} + P_1(1-\theta_{L1})) &= \delta_1\, P_2\, u_c(e_{L2} + P_2(\theta_{L1} - \theta_{L2})) && (19) \\ P_t\, u_c(e_{Lt} + P_t(\theta_{Lt-1} - \theta_{Lt})) &= \delta_t\, P_{t+1}\, u_c(e_{Lt+1} + P_{t+1}(\theta_{Lt} - \theta_{Lt+1})) \quad 2 \le t \le T-2 \\ P_{T-1}\, u_c(e_{LT-1} + P_{T-1}(\theta_{LT-2} - 1)) &= \delta_{T-1}\, u_c(e_{LT} + 1) \end{aligned}$$

can be rearranged by recursive substitution to obtain

$$\begin{aligned} P_1\, u_c(e_{L1} + P_1(1-\theta_{L1})) &= \left(\prod_{j=1}^{T-1} \delta_j\right) u_c(e_{LT} + 1) && (20) \\ P_t\, u_c(e_{Lt} + P_t(\theta_{Lt-1} - \theta_{Lt})) &= \left(\prod_{j=t}^{T-1} \delta_j\right) u_c(e_{LT} + 1) \quad 2 \le t \le T-2 && (21) \\ P_{T-1}\, u_c(e_{LT-1} + P_{T-1}(\theta_{LT-2} - 1)) &= \delta_T\, u_c(e_{LT} + 1). && (22) \end{aligned}$$

Two insights follow immediately. First, since current and past trades are observable, the young short-lived investors can use P_t, θ_{Lt-1}, and θ_{Lt} at each date t to compute the statistic

$$z_t \equiv \frac{P_t\, u_c(e_{Lt} + P_t(\theta_{Lt-1} - \theta_{Lt}))}{u_c(e_{LT} + 1)} = \prod_{j=t}^{T-1} \delta_j. \tag{23}$$

Lemma 3 *Short-horizon investors' equilibrium beliefs about the long-horizon investors' next-period cumulative time-preferences are* $Prob\,(\, z_{t+1} \le x \,|\, (\theta_{L1}, P_1), \ldots, (\theta_{Lt}, P_t)) = Prob\,(\, z_{t+1} \le x \,|\, z_1, \ldots, z_t)$.

Second, young short-lived investors only care about the distribution of bond prices one period ahead. From (22), the price at time $T-1$ is a function $P_{T-1}(z_{T-1}, \theta_{LT-2})$ which means that, from the perspective of the last young cohort at time $T-2$, the only reason P_{T-1} is random is because of z_{T-1}. A recursive argument establishes an analogous result at each earlier date. In particular, from the perspective of the $t-1$ cohort at time $t-1$, the price P_t is random only because of z_t.

Proposition 8 *Given the IES restriction* $-\frac{u_c(x)}{x\, u_{cc}(x)} > 1$ *holds for the long-horizon investors, a symmetric rational expectations equilibrium exists in which bond prices are functions* $P_t(\theta_{Lt-1}, z_t, \ldots, z_1)$.

Multi-period liquidity discovery causes order flows, prices, volatility, and risk premia to co-move as in our three-date model. Over time, short-horizon investors use the trading process to learn about the sequence of the long-lived investors' preferences $\delta_1, \ldots, \delta_{T-1}$ and, hence, about the distribution of future preferences and prices. Since preference uncertainty is multi-dimensional here, early trading is not necessarily fully revealing. The short-horizon investors' information about the summary statistics $z_t = \prod_{j=t}^{T-1} \delta_j$ becomes finer through time, but liquidity risk need not decrease monotonely through time. Our next example illustrates how the market can switch randomly between regimes of high and low liquidity risk.

Example 4 Consider an economy with four dates where the sequence $(\delta_1, \delta_2, \delta_3)$ of subperiod time-preferences for the long-horizon investor is drawn from the set $\{(L, H, H), (H, L, H), (H, H, L)\}$ with $H > L$. Suppose at time 1 the short-lived investors initially believe the sequence (L, H, H) is highly probable and that the two alternative sequences (H, L, H) and (H, H, L) have equal, but low probabilities. Given these beliefs, the market starts in a regime of low liquidity risk since at date 1 the short-horizon investors expect high resale prices P_2 with $z_2 = \delta_2\delta_3 = HH$. However, if the trade at date 2 unexpectedly reveals that δ_1 was actually H and not L, then the market switches to a regime of high liquidity risk given the uncertainty about whether the true sequence of long-lived time preferences is (H, L, H) or (H, H, L). While simple, this example illustrates that the level of liquidity risk can evolve non-monotonically over time and that innovations in trading activity can trigger shifts between different liquidity risk regimes. ■

3 Empirical implications

Liquidity discovery offers a promising explanation for a growing body of empirical evidence that microstructure variables — bid-ask spreads, order flow/price impact coefficients, and the probability of informed trading — are priced in the sense that they explain expected returns. Liquidity premia are documented in Amihud and Mendelson [1986] and more recently in Brennan and Subrahmanyam [1996], Easley, Hvidkjaer, and O'Hara [2002], Stambaugh and Pastor [2003], and Hasbrouck [2004]. The theoretical interpretation of this evidence is, however, still unresolved. The debate has largely defined liquidity in terms of transaction costs. Constantinides [1986] and Vayanos [1998] show that high transaction costs cause investors to trade less, but have little impact on equilibrium expected returns. Huang [2003] finds that transaction cost do lead to higher expected returns when random shocks and borrowing constraints force early liquidation. Acharya and Pedersen [2003] explain the Stambaugh and Pastor [2003] results in a model with cross-sectionally correlated transaction costs and short-holding periods.

O'Hara [2003] argues that a fundamental rethinking of the linkage between micro trading decisions and macro asset pricing is needed to account for the empirical evidence. The idea of liquidity discovery does this. Our analysis links liquidity premia to uncertainty about the aggregate future demand for securities rather than to transactional frictions requiring compensation. Investors care about liquidity, not just as an incremental cost of personal trading relative to a baseline price, but because the current price/order flow sensitivity is informative about the equilibrium prices at which markets will clear in the future. Current liquidity and other microstructure variables are forward-looking measures of future preference-induced price risk.[10]

Figure 9 plots the volatility of bond returns and the associated risk premium versus the price-impact per share of order flow — defined here as $\frac{|P_1(\theta_{L1})-P_1(\min\theta_{L1})|}{\theta_{L1}-\min\theta_{L1}}$ — from Example 3. The smaller price impacts on the left of Figure 9 correspond to fully revealing volumes in Figure 6 less than the critical value $1/z'$ for which there is no future price risk or risk premium. The larger price impacts on the right correspond to partially revealing volumes larger than $1/z'$ for which there are risk and risk premia. The largest impacts at the far right of Figure 9 are associated with the intermediate volumes close to the discontinuity in Figure 6 implying that the per share price impact of order flow is non-monotone in volume in this example. Our price impact liquidity variable appears to be priced, but this is solely because of its endogenous equilibrium correlation with the distribution of the future price P_2. There are no personal trading costs for which investors must be compensated in our competitive Walrasian market. Individual investors behave as if there is *infinite* transactional liquidity at the market-clearing price.[11]

Our liquidity discovery interpretation of empirical liquidity premia avoids the Constantinides [1986] critique of the transaction cost interpretation. According to this critique, the estimated

[10]Novy-Marx [2004] makes a similar point in a model with no liquidity risk, but where market liquidity variables proxy for omitted cash flow risk factors.

[11]Liquidity discovery is yet more complicated when investors arrive in the market asynchronously, but market makers and other short-run liquidity providers are fundamentally still trying to learn about the market-clearing price given the investing public's preferences and beliefs.

liquidity premia of 6-7 percent are disproportionately large compared to transaction costs where both levels and cross-sectional variation are measured in eighths. More precisely, Lesmond, Ogden, and Trzcinka [1999] estimates total transaction costs as low as 1 percent for large stocks. Premia based on liquidity discovery avoid the disproportionality critique because the current price/order flow relation can be informative about future aggregate price risks much larger than eighths. Not only are the risk premia increasing going from the high to the low liquidity regions in Figure 9, but the roughly 2 percent premia, which are based on a reasonable short-horizon investor risk aversion $\gamma = 3$, are in the right ballpark numerically.

Common factors. Since our model has just one long-dated security, the bond represents the entire bond market. Prices and order flows should be interpreted as a price index and as a market-wide common factor in bond order flows. The existence of common factors in security prices and order flows is well-documented (see Lo and Wang [2000].) The price impact in Figure 9 should also be interpreted as a common factor in market liquidity. Chordia, Roll, and Subrahmanyam [2000] and Hasbrouck and Seppi [2001] both find statistically significant, but numerically small common factors in market liquidity. Despite its small size, Pastor and Stambaugh [2003] and Acharya and Petersen [2003] find that the common factor in liquidity is empirically priced. This is yet another example of the disproportionality critique. Liquidity discovery again provides a resolution. As is illustrated in Figure 9, small variations in the price impact of order flow — measured here in terms of percentage price changes relative to par value per percentage share of the total bond supply traded — can be associated with large variations in risk premia because of their predictive power for future liquidity risk.

Proxies for liquidity discovery. Investors arrive asynchronously in actual markets so that the process of liquidity discovery unfolds trade-by-trade along a sequence of transactions. Microstructure measures of liquidity that are often interpreted as purely transactional — such as realized bid/ask spreads and price impact coefficients — can have components relating to liquidity discovery as well as to order processing costs and adverse selection about cash-flows.

Time-variation. Markets may switch over time between regimes of higher or lower liquidity risk and more or less intense liquidity discovery. Intertemporal variation in the intensity of liquidity risk can lead to return heteroscedasticity and time-varying expected returns. Market liquidity will also co-move with the prevailing liquidity risk regime and, hence, will have predictive power for the conditional distribution of future prices.

The level of liquidity risk can change, not only in response to the arrival of public announcements about investor preferences (e.g., investor confidence surveys), but also endogenously in response to innovations in market trading. In particular, abnormal volume may Granger-cause periods of heightened liquidity risk and illiquidity. In Example 4 a volume realization that reveals (H, L, H) or (H, H, L) triggers elevated liquidity risk. Thus, liquidity discovery is a possible explanation for evidence in Lamoureux and Lastrapes [1990] that volume dominates GARCH effects in predicting future return volatility.

Price randomness. French and Roll [1986] documents that prices are more volatile during periods of trading than when exchanges are closed. Liquidity discovery provides an alternate explanation for greater price volatility during trading hours in addition to the revelation of investors' asymmetric cash flow information. Our model is also consistent with Evans and Lyons [2002] which shows that foreign exchange rate changes are correlated with foreign exchange order flow. Trading can affect asset prices via liquidity discovery even when there is no adverse selection due to private information about cash flows.

Cross-sectional asset pricing. Price randomness can, in general, be decomposed into cash flow risk and preference-induced liquidity risk. If different stocks have differing exposures to liquidity risk, then they should have different liquidity risk premia. Stocks with greater future liquidity risk and higher expected returns will also tend to have more intense ex ante liquidity discovery and, thus, higher intra-month price/order flow sensitivities. Liquidity will be priced in cross-sectional asset pricing tests, not as a transaction cost risk factor, but as an instrument for changes in the conditional distribution of future prices. Here proxy variables for changing liquidity risk play an analogous role to conditioning variables for time-varying beta in tests of the conditional CAPM (see Jagannathan and Wang [1996]).

Different investors' preferences will be disproportionately priced into different stocks if the investing public is fragmented into (possibly overlapping) clienteles for different stocks. For example, the preferences of under-diversified entrepreneurs may be more important for low market value stocks than for S&P 500 stocks. (See Heaton and Lucas [2000].) Liquidity risk may also be more important in countries with less developed financial markets or a narrower investor base. Alternatively, in a more general multi-factor economy with state-dependence in both preferences and cash flows, the precise form of investors' utility state-dependence may not be common knowledge.

Event study implications. Events that heighten uncertainty about the marginal investors' preferences will increase future price variability and the price sensitivity to order flow. Examples include changes in a company's investor base such as the split between small retail versus institutional ownership, domestic versus foreign investors or, more dramatically, the death of a large stake-holder whose heirs inherit large holdings. Bennett, Sias, and Starks [2003] finds that the hedonic preferences of mutual funds and institutional money managers change over time. Periods of volatile institutional preferences should, therefore, be associated with heightened liquidity risk and higher liquidity risk premia.

4 Conclusion

This paper has presented an integrated model of asset pricing and trading with liquidity risk and liquidity discovery given heterogenous investors with asymmetric information about each others' preferences. Preference-induced randomness in the future prices of long-dated securities exposes investors to liquidity risk when they need to retrade. Accordingly, a risk premium is required ex ante to clear the market for the long-dated securities with liquidity risk. Public trading is important because investors learn about each others' preferences by observing each others' trades through time. This is the process of liquidity discovery.

We argue that uncertainty about investors' preferences and their future securities demands are facts of life in large decentralized financial markets. As such, liquidity risk and liquidity discovery should be pervasive and empirically important phenomena. They also provide a bridge between general equilibrium asset pricing and market microstructure. Time variation in the intensity of liquidity risk may help explain empirical evidence that market microstructure variables such as volume and price/order flow impacts can predict asset pricing variables such as future price volatility and risk premia. Liquidity is priced in our model as a forward-looking measure of preference-induced risk in future market-clearing prices. As an explanation for liquidity premia in expected returns, liquidity discovery is fundamentally different from compensation for incremental trading costs.

A natural extension to our model is to allow for both cash flow risk and liquidity discovery with a cross-section of stocks. How does the required risk premium for liquidity risk change in the presence of cash flow risk? How does liquidity discovery affect the premium for cash flow risk? Can cross-sectional differences in stocks' exposures to liquidity risk be endogenized? We leave such questions for future work.

Appendix

Proof of Lemmas 1 and 2. The results follow directly from equation (9). ■

Proof of Proposition 1. In equilibrium, prices and allocations equate supply and demand curves implicitly defined by each investor's first-order conditions (6), (7), and (8) evaluated at the market-clearing conditions $\theta_{L0} + \theta_{S0} = 1$, $\theta_{L1} + \theta_{S1} = 1$ and $\theta_{S2} = 0$. After substituting in market-clearing, the first-order conditions can be expressed in terms of the short-horizon investors' initial position θ_{S0} and their equilibrium choice θ_{S1}

$$\begin{aligned} v_c(e_{S1} + P_1(\theta_{S0} - \theta_{S1}))\, P_1 &= \beta\, E_{S1} \left[v_c(e_{S2} + \tilde{P}_2 \theta_{S1}) \tilde{P}_2 \right] && (24) \\ u_c(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))\, P_1 &= \delta_1 \delta_2\, u_c(1) && (25) \\ u_c(e_{L2} - P_2 \theta_{S1}) P_2 &= \delta_2\, u_c(1). && (26) \end{aligned}$$

Existence of an equilibrium means that equations (24) – (26) always have a solution (P_1, P_2, θ_{S1}). The equilibrium is unique if the solution (P_1, P_2, θ_{S1}) is unique.

We first prove existence. Briefly, the proof involves showing that (25) implicity defines a roughly increasing function $\theta_{S1} = g(P_1)$ and that (24), after using (26) to substitute out P_2, defines a roughly decreasing function $\theta_{S1} = h(P_1)$ which intersects g at least once. The qualifier "roughly" refers to the complication that g and h are not necessarily strictly monotone. We show, however, that the non-monotonicities are not so severe as to preclude an intersection. In proving existence, we make no restrictions on the intertemporal elasticity of substitution of the long-horizon investor.

Step 1. We first establish some properties of the long-horizon investors' optimal choice. Using the implicit function theorem, the first-order condition (25) defines the net supply function $\theta_{S1} = g(P_1)$ for the long-horizon investor at date 1 where $g : (0, \infty) \to (-\infty, \overline{\theta})$ and

$$\frac{\partial \theta_{S1}}{\partial P_1} = -\frac{u_c(e_{L1} + P_1(\theta_{S1} - \theta_{S0})) + P_1(\theta_{S1} - \theta_{S0}) u_{cc}(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))}{P_1^2\, u_{cc}(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))}. \qquad (27)$$

In particular, the long-horizon investor is willing to sell $g(P_1) - \theta_{S0}$ at a price P_1. The RHS in (27) implies that $g(P_1)$ is monotone increasing when $\theta_{S1} < \theta_{S0}$ but potentially non-monotone when

$\theta_{S1} > \theta_{S0}$. However, if we let P_1^* denote the price at which the long-horizon investor does not trade at date 1 — where solving $g(P_1^*) - \theta_{S0} = 0$ gives $P_1^* = \frac{\delta_1\delta_2 u_c(1)}{u_c(e_{L1})}$ — it follows from (25) that $\theta_{S1} < \theta_{S0}$ whenever $P_1 < P_1^*$ and that $\theta_{S1} > \theta_{S0}$ when $P_1 > P_1^*$. In this sense, g is "roughly" increasing in P_1 relative to the quantity/price pair (θ_{S0}, P_1^*).

In the extremes, (25) and the Inada conditions imply that $g(P_1) \downarrow -\infty$ and $g(P_1)P_1 \downarrow -e_{L1}$ as $P_1 \downarrow 0$ and that $g(P_1)P_1 \uparrow \infty$ as $P_1 \uparrow \infty$. However, the limit of $g(P_1)$ as $P_1 \uparrow \infty$ can be either ∞ or finite depending on the relative speeds with which $u_c(e_{L1}+P_1(\theta_{S1}-\theta_{S0}))$ and $P_1(\theta_{S1}-\theta_{S0})u_{cc}(e_{L1}+P_1(\theta_{S1}-\theta_{S0}))$ in (27) go to 0 as $P_1 \uparrow \infty$. The potential boundedness of $\lim_{P_1\to\infty} g(P_1)$ and the continuity of $g(P_1)$ in turn imply that the upper bound $\overline{\theta}$ in the range of g can be finite (rather than ∞) for some utility functions u.

Step 2. Using the implicit function theorem again, the first-order condition (26) defines an ex ante net demand function $\theta_{S1} = f(P_2;\, \delta_2)$ for the long-horizon investor at date 2 for each particular realization of δ_2 where $f : (0,\infty) \to (\underline{\theta}, \infty)$ and

$$\frac{\partial \theta_{S1}}{\partial P_2} = \frac{u_c(e_{L2} - P_2\theta_{S1}) - P_2\theta_{S1}u_{cc}(e_{L2} - P_2\theta_{S1})}{P_2^2\, u_{cc}(e_{L2} - P_2\theta_{S1})}. \tag{28}$$

From (28) we see that f is monotone decreasing when $\theta_{S1} = f(P_2;\, \delta_2) > 0$ but that $f(P_2;\, \delta_2)$ is potentially non-monotone when $\theta_{S1} = f(P_2;\, \delta_2) < 0$. Solving for the price $P_2^* = \frac{\delta_2 u_c(1)}{u_c(e_{L2})}$ at which the long-horizon investor trades $f(P_2^*;\, \delta_2) = 0$, we note from (26) that $f(P_2^*;\, \delta_2) > 0$ when $P_2 < P_2^*$ and $f(P_2^*;\, \delta_2) < 0$ when $P_2 > P_2^*$. It follows from the Inada conditions and (26) that $f(P_2;\, \delta_2) \uparrow \infty$ and $f(P_2;\, \delta_2)P_2 \uparrow e_{L2}$ as $P_2 \downarrow 0$ and that $f(P_2;\, \delta_2)P_2 \downarrow -\infty$ as $P_2 \uparrow \infty$. The limit of $f(P_2;\, \delta_2)$, however, is negative but either bounded or $-\infty$ as $P_2 \uparrow \infty$ depending on the relative speeds with which $u_c(e_{L2} - P_2\theta_{S1})$ and $P_2\theta_{S1}u_{cc}(e_{L2} - P_2\theta_{S1})$ in (28) go to 0. Hence, the lower bound in the range of g is strictly negative, $\underline{\theta} < 0$, but potentially bounded depending, again, on the utility function u.

Previewing the remaining steps, we use (26) to substitute out P_2 in the RHS of (24) so as to express the short-horizon investors' first-order condition (24) as a second equation in P_1 and θ_{S1}. This defines the net demand function $\theta_{S1} = h(P_1)$ for the short-horizon investors at date 1. This construction is straightforward when the inverse demand function $P_2 = f^{-1}(\theta_{S1};\, \delta_2)$ exists. Unfortunately, there are two potential complications. First, any non-monotonicity in the demand function f for $P_2 > P_2^*$ means that some short positions $\theta_{S1} < 0$ can be absorbed at more than one market-clearing price P_2. In these cases, the inverse demand f^{-1} is a correspondence rather than a well-defined function. A second complication is that f^{-1} is only defined for quantities θ_{S1} for which there exists a market-clearing price P_2 at date 2. In other words, the optimal quantity θ_{S1} for the short-horizon investor given any price P_1 at date 1 must, by construction, be in the range $(\underline{\theta}, \infty)$ of the demand function f for the long-horizon investors at date 2. This same restriction on allowable values of θ_{S1} is also implicit in the derivation of (25) and the substitution of the date 2 FOC into the date 1 FOC for the long-horizon investors.

Step 3. When $\theta_{S1} \geq 0$, the demand function f at date 2 is monotone so that a well-defined inverse demand function f^{-1} exists for each δ_2. Hence, the RHS of (24) can be written as

$$q(\theta_{S1}) = \beta E_{S1}\left[v_c(e_{S2} + f^{-1}(\theta_{S1};\, \tilde{\delta}_2)\, \theta_{S1}) f^{-1}(\theta_{S1};\, \tilde{\delta}_2)\right] \qquad \forall\, \theta_{S1} \geq 0 \tag{29}$$

where q is a monotone decreasing, continuous function in θ_{S1}. In particular, its slope is given by

$$\begin{aligned}\frac{\partial q}{\partial \theta_{S1}} = \beta E_{S1}\Bigg[& v_c(e_{S2} + f^{-1}(\theta_{S1};\, \tilde{\delta}_2)\, \theta_{S1}) \frac{\partial f^{-1}}{\partial \theta_{S1}} + \\ & v_{cc}(e_{S2} + f^{-1}(\theta_{S1};\, \tilde{\delta}_2)\, \theta_{S1})\, f^{-1}(\theta_{S1};\, \tilde{\delta}_2) \left(\frac{\partial f^{-1}}{\partial \theta_{S1}}\, \theta_{S1} + f^{-1}(\theta_{S1};\, \tilde{\delta}_2) \right) \Bigg] < 0. \end{aligned} \tag{30}$$

where the inequality follows because, from (28), $\frac{\partial f}{\partial P_2} < 0$ and, hence, $\frac{\partial f^{-1}}{\partial \theta_{S1}} < 0$ for $\theta_{S1} > 0$ and because

$$\begin{aligned}\frac{\partial f^{-1}}{\partial \theta_{S1}}\, \theta_{S1} + f^{-1}(\theta_{S1};\, \tilde{\delta}_2) &= \frac{P_2^2\, u_{cc}(e_{L2} - P_2\theta_{S1})}{u_c(e_{L2} - P_2\theta_{S1}) - P_2\theta_{S1} u_{cc}(e_{L2} - P_2\theta_{S1})}\, \theta_{S1} - P_2 \\ &= \frac{u_c(e_{L2} - P_2\theta_{S1})\, P_2}{u_c(e_{L2} - P_2\theta_{S1}) - P_2\theta_{S1} u_{cc}(e_{L2} - P_2\theta_{S1})} > 0. \end{aligned} \tag{31}$$

Step 4. Our analysis of short positions $\theta_{S1} < 0$ is restricted to positions θ_{S1} in the domain of the inverse demand f^{-1}. This ensures that the short position $\theta_{S1} = h(P_1)$ chosen by the short-horizon investors at date 1 can be associated with a market-clearing price at date 2. For each possible δ_2, define the price $P_2^{\delta_2}$ and corresponding position $\theta^{\delta_2} = f(P_2^{\delta_2}; \delta_2)$ such that $\theta^{\delta_2} P_2^{\delta_2} = -e_{S2}$.

The Inada conditions and (26) imply that the product $f(P_2;\, \delta_2)P_2$ falls monotonely to $-\infty$ as $P_2 \uparrow \infty$ so that $P_2^{\delta_2}$ exists and is unique. In addition, define $\hat{P}_2^{\delta_2}$ as the largest price $p \leq P_2^{\delta_2}$ in the set $f^{-1}(\theta^{\delta_2}; \delta_2)$ such that, locally, the slope of the correspondence at $(\theta^{\delta_2}, \hat{P}_2^{\delta_2})$ is negative. The construction of $\hat{P}_2^{\delta_2}$ guarantees that quantity/price pairs (θ_{S1}, P_2) "to the right" of $(\theta^{\delta_2}, \hat{P}_2^{\delta_2})$ on the local portion of the correspondence — that is, for quantities $\theta_{S1} \geq \theta^{\delta_2}$ — all have products $\theta_{S1}P_2 > -e_{S2}$ so that, given the Inada conditions, $v_c(e_{S2} + f^{-1}(\theta_{S1};\, \delta_2)\theta_{S1})f^{-1}(\theta_{S1};\, \delta_2)$ is defined and finite.

For each short position $0 > \theta_{S1} > \theta^{\delta_2}$, define $P_+^{\delta_2}(\theta_{S1})$ and $P_-^{\delta_2}(\theta_{S1})$ as the maximum and minimum prices $P_2 \leq \hat{P}_2^{\delta_2}$ such that (26) holds given a particular δ_2. For quantities θ_{S1} such that the solution to (26) is unique, then $P_+^{\delta_2}(\theta_{S1}) = P_-^{\delta_2}(\theta_{S1})$. From the geometry of f^{-1}, both $P_+^{\delta_2}(\theta_{S1})$ and $P_-^{\delta_2}(\theta_{S1})$ are strictly increasing in θ_{S1}. Moreover, since $\hat{P}_2^{\delta_2}$ is on a locally negatively sloped part of the correspondence f^{-1}, the maximal prices $P_+^{\delta_2}(\theta_{S1})$ "to the right" of $(\theta^{\delta_2}, \hat{P}_2^{\delta_2})$ are less than $\hat{P}_2^{\delta_2}$. Hence, $v_c(e_{S2} + P_+^{\delta_2}(\theta_{S1})\theta_{S1})P_+^{\delta_2}(\theta_{S1})$ is defined give the Inada conditions.

Step 5. We next construct an extension $\hat{q}$ of the expectation function q from Step 3 for short positions $\theta_{S1} < 0$. This extension must be consistent with rational expectations and with the long-horizon investor's first-order condition (26). Since the goal is just to prove existence of an equilibrium, $\hat{q}$ does not need to be unique.

From Step 4 we have a critical quantities θ^{δ_2} and maximal and minimal price functions $P_+^{\delta_2}(\theta_{S1})$ and $P_-^{\delta_2}(\theta_{S1})$ for each time-preference realization δ_2. At time 1 when the expectation in (24) is taken, the short-horizon investors only have a conditional distribution over δ_2 given z. To insure that $v_c(e_{S2} + \tilde{P}_2\theta_{S1})\tilde{P}_2$ is well-defined in the expectation in (24), we initially just extend q over an interval $(\theta^m, 0)$ where $\theta_m = \max\{\theta^{\delta_2} \,|\, \mathrm{Prob}(\delta_2 \,|\, \delta_1\delta_2 = z) > 0\} < 0$. Let δ_2^m denote the value of δ_2 for which $\theta^{\delta_2} = \theta^m$. Let P_2^m and $\hat{P}_2^m$ be the associated values of $P_2^{\delta_2}$ and $\hat{P}_2^{\delta_2}$ for δ_2^m.

Consider any monotone decreasing, continuous function $\hat{q}(\theta_{S1})$ such that for all short positions $0 > \theta_{S1} > \theta^m$

$$E_{S1}\left[v_c(e_{S2} + P_-^{\tilde{\delta}_2}(\theta_{S1})\,\theta_{S1})P_-^{\tilde{\delta}_2}(\theta_{S1})\right] \leq \hat{q}(\theta_{S1}) \leq E_{S1}\left[v_c(e_{S2} + P_+^{\tilde{\delta}_2}(\theta_{S1})\,\theta_{S1})P_+^{\tilde{\delta}_2}(\theta_{S1})\right] \tag{32}$$

and $\hat{q}(\theta_{S1}) \uparrow E_{S1}\left[v_c(e_{S2} + P_+^{\tilde{\delta}_2}(\theta_{S1})\,\theta_{S1})P_+^{\tilde{\delta}_2}(\theta_{S1})\right]$ as $\theta_{S1} \downarrow \theta^m$. By having the long-horizon traders collectively randomize P_2 between $P_+^{\delta_2}(\theta_{S1})$ and $P_-^{\delta_2}(\theta_{S1})$ for each δ_2 using the appropriate probabilities $\pi^+(\theta_{S1})$ and $\pi^-(\theta_{S1}) = 1 - \pi^+(\theta_{S1})$, we can extend q to match

$$E_{S1}\left[v_c(e_{S2} + \tilde{P}_2\theta_{S1})\tilde{P}_2\right] = \hat{q}(\theta_{S1}). \tag{33}$$

Note that the short-horizon investors' expectation in $E_{S1}\left[v_c(e_{S2} + \tilde{P}_2\theta_{S1})\tilde{P}_2\right]$ is taken over both the potentially random time preference $\tilde{\delta}_2$ and also over any price randomization when P_2 from (26) given δ_2 is not unique. Since the long-horizon investors are not strategic, they do not need to be indifferent over different value of this price "sunspot." Furthermore, if the randomization over P_2 occurs at date 1, then there still is no uncertainty for the long-horizon investors about their date 2 consumption.

Step 6. There are two cases to consider to complete the existence proof. One possibility is that $P_2^m = \hat{P}_2^m$. In this case, we do not need to extend q beyond the interval $(\theta^m, 0)$. We use q to rewrite (24) as

$$v_c(e_{S1} - P_1(\theta_{S1} - \theta_{S0}))\, P_1 = q(\theta_{S1}). \tag{34}$$

Recall that $q(\theta_{S1})$ is a monotone decreasing continuous function with a domain (θ^m, ∞) where $q(\theta_{S1}) \uparrow \infty$ as $\theta_{S1} \downarrow \theta^m$. Invoking the implicit function theorem, (34) defines the net demand function $\theta_{S1} = h(P_1)$ for the short-horizon investors at date 1 where $h : (0, \infty) \to (\theta^m, \infty)$. Solving $v_c(e_{S1})\, P_1 = q(\theta_{S0})$ gives the price $P_1^{**} = \frac{q(\theta_{S0})}{v_c(e_{S2})}$ at which the short-horizon investors choose to hold $\theta_{S1} = \theta_{S0}$ bonds and, thus, trade 0 at date 1. From (34) we can show that $h(P_1)$ is roughly decreasing in P_1. In particular, $\theta_{S1} < \theta_{S0}$ when $P_1 > P_1^{**}$ and, conversely, $\theta_{S1} > \theta_{S0}$ when $P_1 < P_1^{**}$. Recall from Step 1 that the opposite is true for $g(P_1)$ from (25) relative to P_1^*. Hence, there must be at least one pair (P_1, θ_{S1}) such that $g(P_1) = \theta_{S1} = h(P_1)$. Hence, an equilibrium exists in the case where $P_2^m = \hat{P}_2^m$.

Step 7. The other possibility is $P_2^m > \hat{P}_2^m$. In this case, since the domain (θ^m, ∞) of q stops above θ^m, there may be some prices P_1 for which no solution $\theta_{S1} > \theta^m$ to (24) exists because

$$v_c(e_{S1} + P_1(\theta_{S0} - \hat{\theta}_{S1}))\, P_1 > q(\theta_{S1}) \qquad \forall\, \theta_{S1} \in (\theta^m, \infty). \tag{35}$$

In this case, we extend the function q to a quantity θ^{ϵ} corresponding to a price $P^{\epsilon} = P_2^m - \epsilon$ given the demand function f for δ_2^m. Although $v_c(e_{S2} + P_2\theta_{S1})\, P_2$ is not defined at (θ^m, P_2^m), it is defined at $(\theta^{\epsilon}, P_2^{\epsilon})$ and for all quantity/price pairs on f^{-1} between (θ^m, P_2^m) and $(\theta^{\epsilon}, P_2^{\epsilon})$. In particular, the fact that $P_2^m > \hat{P}_2^m$ implies that (θ^m, P_2^m) is on either a locally positively sloped (if $\theta^{\epsilon} < \theta^m$) or vertical (if $\theta^{\epsilon} = \theta^m$) portion of the correspondence f^{-1}.

We continue randomizing as before between $P_+^{\delta_2}(\theta_{S1})$ and $P_-^{\delta_2}(\theta_{S1})$ so as to continue the extension of q as a monotone decreasing function $\hat{q}$ between θ^m and θ^{ϵ}. Then at θ^{ϵ} for any price P_1 such that inequality (35) holds for all $\theta_{S1} > \theta^{\epsilon}$, we choose mixing probabilities such that the short-horizon investor's first-order condition (24) holds at θ^{ϵ}. Since $v_c(e_{S2} + P_2^m\theta^m)\, P_2^m = \infty$, we can always find an ϵ sufficiently small such that this is possible for any finite price P_1. At this point, the rest of proof in the $P_2^m > \hat{P}_2^m$ case uses the same arguments about $\theta_{S1} = h(P_2)$ being roughly decreasing as in the $P_2^m = \hat{P}_2^m$ case. This completes the proof of existence.

We next show that our restriction (IES) on the long-horizon investor's intertemporal elasticity of substitution $-\frac{u_c(x)}{x\, u_{cc}(x)} > 1$ is sufficient for uniqueness. Given the IES restriction, the slope of the inverse net demand function $P_2 = f^{-1}(\theta_{S1}|\delta_2)$ from (26) can now be unambiguously signed

$$\frac{\partial P_2}{\partial \theta_{S1}} = \frac{P_2^2\, u_{cc}(e_{L2} - P_2\theta_{S1})}{u_c(e_{L2} - P_2\theta_{S1}) - P_2\theta_{S1}u_{cc}(e_{L2} - P_2\theta_{S1})} < 0. \tag{36}$$

The inequality follows because our IES restriction, rewritten as $u_c(a + x) + (a + x)\, u_{cc}(a + x) > 0$ for $\forall a + x > 0$, implies that $u_c(a + x) + x\, u_{cc}(a + x) > 0$ for all $a \geq 0$. Thus, the long-horizon investor's inverse demand at date $t = 2$ is monotonically decreasing in θ_{S1}. Similarly, the slope of the net demand $\theta_{S1} = g(P_1)$ from (25) can also be signed unambiguously

$$\frac{\partial \theta_{S1}}{\partial P_1} = -\frac{u_c(e_{L1} + P_1(\theta_{S1} - \theta_{S0})) + P_1(\theta_{S1} - \theta_{S0})u_{cc}(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))}{P_1^2\, u_{cc}(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))} > 0 \tag{37}$$

where the inequality again follows from the IES restriction.

If the short-horizon investors' net demand $\theta_{S1} = h(P_1)$ given their utility v is monotone, then the equilibrium is unique. However, no additional restrictions on v (or h) are required for uniqueness given the IES restriction on the long-horizon investors. For the intersection $[\underline{\theta}, \overline{\theta}]$ of the range of $g(\cdot)$ and the domain of $f(\cdot)$, we define the composite function $h = f \circ g$ expressing P_2 as a function

$P_2 = h(P_1|\delta_2)$ of P_1 given the realized δ_2. From the chain rule, note that $\frac{\partial P_2}{\partial P_1} < 0$. Using the short-horizon investor's first order condition (24), we then define the function $k(P_1)$ as

$$k(P_1) = \beta\, E_{S1}\left[v_c(e_{S2} + \tilde{P}_2\theta_{S1})\tilde{P}_2\right] - v_c(e_{S1} + P_1(\theta_{S0} - \theta_{S1}))\, P_1 \tag{38}$$

where the functions $g(P_1)$ and $h(P_1|\delta_2)$ are substituted for θ_{S1} and $\tilde{P}_2$ respectively. If $k(\cdot)$ is monotonic in P_1, then the equilibrium must be unique. Differentiating with respect to P_1,

$$\begin{aligned}\frac{\partial k}{\partial P_1} = \beta\, E_{S1}&\left[v_c(e_{S2} + \tilde{P}_2\theta_{S1})\frac{\partial \tilde{P}_2}{\partial P_1} + v_{cc}(e_{S2} + \tilde{P}_2\theta_{S1})\tilde{P}_2\, \left(\tilde{P}_2\frac{\partial \theta_{S1}}{\partial P_1} + \theta_{S1}\frac{\partial \tilde{P}_2}{\partial P_1}\right)\right] \\ &- v_c(e_{S1} + P_1(\theta_{S0} - \theta_{S1})) + P_1\, v_{cc}(e_{S1} + P_1(\theta_{S0} - \theta_{S1}))\left(\theta_{S1} - \theta_{S0} + P_1\frac{\partial \theta_{S1}}{\partial P_1}\right).\end{aligned} \tag{39}$$

To sign $\frac{\partial k}{\partial P_1}$, we first sign the terms $\tilde{P}_2\frac{\partial \theta_{S1}}{\partial P_1} + \theta_{S1}\frac{\partial \tilde{P}_2}{\partial P_1}$ and $\theta_{S1} - \theta_{S0} + P_1\frac{\partial \theta_{S1}}{\partial P_1}$. Using (36),

$$\begin{aligned}\tilde{P}_2\frac{\partial \theta_{S1}}{\partial P_1} + \theta_{S1}\frac{\partial \tilde{P}_2}{\partial P_1} &= \frac{\partial \theta_{S1}}{\partial P_1}\left(\tilde{P}_2 + \theta_{S1}\frac{\partial \tilde{P}_2}{\partial \theta_{S1}}\right) \\ &= \frac{\partial \theta_{S1}}{\partial P_1}\left(\tilde{P}_2 + \frac{\theta_{S1}\, \tilde{P}_2^2\, u_{cc}(e_{L2} - \tilde{P}_2\theta_{S1})}{u_c(e_{L2} - \tilde{P}_2\theta_{S1}) - \tilde{P}_2\theta_{S1}u_{cc}(e_{L2} - \tilde{P}_2\theta_{S1})}\right) \\ &= \frac{\partial \theta_{S1}}{\partial P_1}\left(\frac{\tilde{P}_2\, u_c(e_{L2} - \tilde{P}_2\theta_{S1})}{u_c(e_{L2} - \tilde{P}_2\theta_{S1}) - \tilde{P}_2\theta_{S1}u_{cc}(e_{L2} - \tilde{P}_2\theta_{S1})}\right) > 0.\end{aligned} \tag{40}$$

Using (37),

$$\begin{aligned}&\theta_{S1} - \theta_{S0} + P_1\frac{\partial \theta_{S1}}{\partial P_1} \\ &\quad= \theta_{S1} - \theta_{S0} - \frac{u_c(e_{L1} + P_1(\theta_{S1} - \theta_{S0})) + P_1(\theta_{S1} - \theta_{S0})u_{cc}(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))}{P_1\, u_{cc}(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))} \\ &\quad= \frac{u_c(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))}{P_1\, u_{cc}(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))} > 0.\end{aligned} \tag{41}$$

Using (40), (41), and $\frac{\partial \tilde{P}_2}{\partial P_1} < 0$, every term in (39) is negative implying $\frac{\partial k}{\partial P_1} < 0$ and the equilibrium is unique. ■

Proof of Proposition 2. The result follows from the fact that sets of $\delta_1\delta_2$ products that pool multiple δ_2s have liquidity risk whereas sets of $\delta_1\delta_2$ products that are uniquely associated with single values of δ_2 have no liquidity risk. ■

Proof of Proposition 3. The result follows from Bayes' Rule. ■

Proof of Proposition 4. Consider a perturbation in which each original (δ_1, δ_2) pair is perturbed to a new LDI pair (δ_1', δ_2') such that $\delta_2' > \delta_2$. The resulting perturbed demand curves for the long-horizon investors at date 2 are pointwise higher. This, in turn, raises the short-horizon investors' date 1 bond demand curve provided that date 2 consumption is not a Giffen good. Since LDI perturbations leave the long-term investor's demand curve at date 1 unchanged, the short-horizon investors hold more bonds $\theta_{S1}' > \theta_{S1}$ at date 1. Since the slope of the long-horizon investor's bond supply at date 1 is positive, given the IES condition and (37), the additional short-term bond demand in turn raises the market-clearing date 1 price, $P_1' > P_1$. If the increase in the bonds inelastically sold at date 2 is not too large, then the perturbed prices at date 2 are pointwise still higher $P_2' > P_2$ in that the direct effect of $\delta_2' > \delta_2$ dominates the indirect effect of $\theta_{S1}' > \theta_{S1}$. However, if the increase demand for bonds at date 2 is sufficiently strong, then the direct effect of $\delta_2' > \delta_2$ is dominated by the large indirect effect of $\theta_{S1}' > \theta_{S1}$ and pointwise $P_2' < P_2$. The expected return result follow immediately from the price changes. ■

Proof of Proposition 5 Consider a perturbation in which some (δ_1, δ_2) pairs are perturbed to new LDI pairs (δ_1', δ_2') such that some are $\delta_2' > \delta_2$ and some are $\delta_2' < \delta_2$. The perturbed distribution of δ_2' is more volatile than the initial preferences. Suppose further that this perturbation is done so that the expected date 2 price is unchanged at the original volume θ_{S1}.

Case 1. If the precautionary savings motive raises the short-horizon investor's date 1 bond demand curve, then the short-horizon investor buys more bonds $\theta_{S1}' > \theta_{S1}$ at date 1 at — given the positive long-horizon investor slope in (37) with the IES restriction — a higher market-clearing price $P_1' > P_1$. The fact that preferences were perturbed such that $\mathrm{E}_{S1}[P_2' | \theta_{S1}' = \theta_{S1}] = \mathrm{E}_{S1}[P_2]$ and that the date 2 long-horizon demand curves are downward-sloping implies that $\mathrm{E}_{S1}[P_2' | \theta_{S1}' > \theta_{S1}] < \mathrm{E}_{S1}[P_2]$. Hence, the expected return falls.

Case 2. If the perturbation instead lowers the short-horizon investor's date 1 bond demand curve, then the short-horizon investor buys fewer bonds $\theta_{S1}' < \theta_{S1}$ at date 1 at — given the positive slope in (37) — a lower market-clearing date 1 price $P_1' < P_1$. In this case the downward-sloping date 2 long-horizon demand curves imply that $\mathrm{E}_{S1}[P_2' | \theta_{S1}' < \theta_{S1}] > \mathrm{E}_{S1}[P_2]$. Hence, the expected return rises. ■

Proof of Proposition 6 We solve for $\frac{\partial P_1}{\partial \beta}$, $\frac{\partial \tilde{P}_2}{\partial \beta}$, and $\frac{\partial \theta_{S1}}{\partial \beta}$ by totally differentiating the long-horizon and short-horizon investors' first-order conditions in (24), (25), and (26) given market-clearing. First, differentiating the short-lived investor's first-order conditions (24) gives

$$\begin{aligned} 0 \quad = \quad & E_{S1}\left[v_c(\tilde{c}_{S2})\tilde{P}_2\right] + \beta E_{S1}\left[v_c(\tilde{c}_{S2})\frac{\partial \tilde{P}_2}{\partial \beta}\right] \\ & + \beta E_{S1}\left[v_{cc}(\tilde{c}_{S2})\tilde{P}_2\left(\theta_{S1}\frac{\partial \tilde{P}_2}{\partial \beta} + \tilde{P}_2\frac{\partial \theta_{S1}}{\partial \beta}\right)\right] \\ & - v_c(c_{S1})\frac{\partial P_1}{\partial \beta} + P_1 v_{cc}(\tilde{c}_{S1})\left((\theta_{S1} - \theta_{S0})\frac{\partial P_1}{\partial \beta} + P_1\frac{\partial \theta_{S1}}{\partial \beta}\right). \end{aligned} \tag{42}$$

Next, differentiating the long-horizon investors' first-order condition (26), we can express $\frac{\partial P_2}{\partial \beta}$ in terms of $\frac{\partial \theta_{S1}}{\partial \beta}$ as

$$\frac{\partial P_2}{\partial \beta} = \frac{\partial P_2}{\partial \theta_{S1}}\frac{\partial \theta_{S1}}{\partial \beta} \tag{43}$$

where the IES restriction ensures that $\frac{\partial P_2}{\partial \theta_{S1}}$ is well-defined and, given (36), strictly negative. Using (43) and substituting in from (36), we can then rewrite the parenthetical expression in the second line of (42) as

$$\theta_{S1}\frac{\partial \tilde{P}_2}{\partial \beta} + \tilde{P}_2\frac{\partial \theta_{S1}}{\partial \beta} = \frac{\tilde{P}_2\, u_c(\tilde{c}_{L2})}{u_c(\tilde{c}_{L2}) - P_2\theta_{S1}u_{cc}(\tilde{c}_{L2})} \cdot \frac{\partial \theta_{S1}}{\partial \beta} \tag{44}$$

where, using the IES condition again, the ratio on the RHS is unambiguously positive. Similarly, differentiating the long-horizon first-order condition (25) gives

$$\frac{\partial P_1}{\partial \beta} = \frac{\partial P_1}{\partial \theta_{S1}}\frac{\partial \theta_{S1}}{\partial \beta} \tag{45}$$

where the IES restriction ensures that $\frac{\partial P_1}{\partial \theta_{S1}}$ is well-defined and strictly positive. Using (45) we can rewrite the parenthetical expression in the last term in (42) as

$$(\theta_{S1} - \theta_{S0})\frac{\partial P_1}{\partial \beta} + P_1\frac{\partial \theta_{S1}}{\partial \beta} = \frac{P_1\, u_c(c_{L1})}{u_c(c_{L1}) + P_1(\theta_{S1} - \theta_{S0})u_{cc}(c_{L1})} \cdot \frac{\partial \theta_{S1}}{\partial \beta} \tag{46}$$

where the IES restriction and the inverse of (37) imply that the ratio on the RHS is positive.

Substituting expressions (43) through (46) into (42), each term multiplying $\frac{\partial \theta_{S1}}{\partial \beta}$ is negative. Since the leading term $E_{S1}\left[v_c(\tilde{c}_{S2})\tilde{P}_2\right]$ in (42) is positive, it follows that $\frac{\partial \theta_{S1}}{\partial \beta} > 0$. The short-horizon investors' equilibrium bond holdings are increasing in β. Lastly, $\frac{\partial P_2}{\partial \beta} < 0$ since (43) implies that it has the opposite since as $\frac{\partial \theta_{S1}}{\partial \beta}$ and $\frac{\partial P_1}{\partial \beta} > 0$ since (45) implies that it has the same sign as $\frac{\partial \theta_{S1}}{\partial \beta}$. ■

Proof of Proposition 7. The utility of the long-horizon investor as a function of P_1 is given by

$$J(P_1) = u(e_{L1} + P_1(\theta_{S1} - \theta_{S0})) + \delta_1\, u(e_{L1} - P_2\theta_{S1}) + \delta_1\delta_2\, u(1)$$

where we have substituted market clearing at all dates to express portfolio choice in terms of the short-horizon investor, P_1 defined by (9), and P_2 is defined by (10).

Given the assumptions made on preferences, $J'(P_1)$ exists and is continuous. Computing $J'(P_1)$,

$$J'(P_1) = u_c(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))\left(\theta_{S1} - \theta_{S0} + P_1\frac{\partial \theta_{S1}}{\partial P_1}\right) - \delta_1 u_c(e_{L1} - P_2\theta_{S1})\left(\theta_{S1}\frac{\partial P_2}{\partial P_1} + P_2\frac{\partial \theta_{S1}}{\partial P_1}\right).$$

From the envelope theorem, $u_c(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))P_1 = \delta_1 u_c(e_{L1} - P_2\theta_{S1})P_2$, implying

$$J'(P_1) = (\theta_{S1} - \theta_{S0})u_c(e_{L1} + P_1(\theta_{S1} - \theta_{S0})) - \delta_1 u_c(e_{L1} - P_2\theta_{S1})\theta_{S1}\frac{\partial P_2}{\partial P_1}. \tag{47}$$

From the proof of Proposition 1, we know $\frac{\partial P_2}{\partial P_1} < 0$ if the elasticity of intertemporal substitution is greater than or equal to one.

When $\theta_{S1} \geq \theta_{S0}$, $J'(P_1) > 0$, while $J'(P_1) < 0$ if $\theta_{S1} \leq 0$. The sign of $J'(P_1)$ is, therefore, potentially zero only when $0 < \theta_{S1} < \theta_{S0}$. However, the set of long-horizon preferences where $J'(P_1) = 0$ on an open interval $P_1 \in (\overline{P}_1, \underline{P}_1)$ is of measure zero. To see this, suppose that there exists an interval $(\overline{P}_1, \underline{P}_1)$ where $J'(P_1) = 0$. Using optimality and equation (47) set equal to zero, θ_{S1} must satisfy

$$\theta_{S1} = \frac{\theta_{S0}}{1 - \frac{P_1}{P_2}\frac{\partial P_2}{\partial P_1}}$$

when $P_1 \in (\overline{P}_1, \underline{P}_1)$. Given θ_{S1} must also satisfy $u_c(e_{L1} + P_1(\theta_{S1} - \theta_{S0}))P_1 = \delta_1 u_c(e_{L1} - P_2\theta_{S1})P_2$

for all feasible P_1, the set of preferences that satisfy both restrictions on $P_1 \in (\overline{P}_1, \underline{P}_1)$ is measure zero. As a result, $J'(P_1)$ is generically non-zero.

Consider two long-horizon types that could pool implying $\delta_1\delta_2 = \delta_1'\delta_2'$ where $\delta_2 \neq \delta_2'$. Since in a separating equilibrium, these two types would face different equilibrium P_1s, one of these types has an incentive to mimic the other type given $J'(P_1)$ is generically non-zero. Hence, unverifiable statements by long-horizon investors about their preferences are not generically credible. ■

Proof of Lemma 3. The result follows directly from equation (21). ■

Proof of Proposition 8. Substituting market-clearing into the investors' first-order conditions gives a recursive system of equations

$$P_t\, v_c(e_{t,t} - P_t\theta_{St}) = \beta\, E_{St}\left[v_c(e_{t,t+1} + \tilde{P}_{t+1}\theta_{St})\tilde{P}_{t+1}\right] \qquad 1 \leq t \leq T-2 \quad (48)$$

$$P_1\, u_c(e_{L1} + P_1\theta_{S1}) = z_1\, u_c(e_{LT} + 1) \quad (49)$$

$$P_t\, u_c(e_{Lt} + P_t(\theta_{St} - \theta_{St-1})) = z_t\, u_c(e_{LT} + 1) \qquad 2 \leq t \leq T-2 \quad (50)$$

$$P_{T-1}\, u_c(e_{LT-1} - P_{T-1}\theta_{ST-2}) = z_{T-1}\, u_c(e_{LT} + 1). \quad (51)$$

Existence requires that these equations always have a solution.

Step 1. Suppose an equilibrium exists for dates $t, \ldots, T-1$ in which θ_{St} is given by a function $\theta_{St} = w_t(\theta_{St-1}; \underline{z}_t)$ where $\underline{z}_t$ denotes the history of revealed cumulative preferences $\{z_1, \ldots, z_t\}$ up through date t. Suppose also that $w_t(\theta_{St-1} + \Delta; \underline{z}_t) - w_t(\theta_{St-1}; \underline{z}_t) < \Delta$. Note that this is satisfied when $t = T-1$ since $\theta_{ST-1} = 0$ by assumption.

Applying the implicit function theorem to (50) gives the inverse supply function $P_t = f_t^{-1}(\theta_{St} - \theta_{St-1}; z_t)$ for long-horizon investors at date t which, given the IES restriction, is monotone increasing

$$\frac{\partial P_t}{\partial \theta_{St} - \theta_{St-1}} = -\frac{P_t^2\, u_{cc}(e_{Lt} + P_t(\theta_{St} - \theta_{St-1}))}{u_c(e_{Lt} + P_t(\theta_{St} - \theta_{St-1})) + P_t(\theta_{St} - \theta_{St-1})u_{cc}(e_{Lt} + P_t(\theta_{St} - \theta_{St-1}))} > 0. \quad (52)$$

Substituting in w_t gives the price $P_t = f_t^*(\theta_{St-1}; \underline{z}_t) \equiv f_t^{-1}(w_t(\theta_{St-1}; \underline{z}_t) - \theta_{St-1}; z_t)$ as a decreasing function of the previous young cohort's position $\theta_{S,t-1}$. If $\theta_{St-1} \uparrow \infty$, equation (50) implies that $P_t \downarrow 0$ such that $P_t(\theta_{St} - \theta_{St-1}) \downarrow -e_{Lt}$ and, thus, that $P_t\theta_{St-1} \uparrow e_{Lt} + P_t\theta_{St}$. Similarly, if

$\theta_{St-1} \downarrow \underline{\theta}^{t-1} < 0$ where $\underline{\theta}^{t-1}$ is the lower bound of the domain of f_t^{-1}, then $P_t \uparrow \infty$ so that $P_t(\theta_{St} - \theta_{St-1}) \uparrow \infty$ and, thus, $P_t\theta_{St-1} \downarrow -\infty$.

Step 2. Substituting $P_t = f_t^*(\theta_{St-1}; \underline{z}_t)$ into (48) gives an equation in θ_{St-1} and P_{t-1}

$$v_c(e_{t-1,t-1} - P_{t-1}\theta_{St-1})\, P_{t-1} = \beta\, E_{St-1}\left[v_c(e_{t-1,t} + f_t^*(\theta_{St-1}; \underline{\tilde{z}}_t)\theta_{St-1}) f_t^*(\theta_{St-1}; \underline{\tilde{z}}_t)\right] \tag{53}$$

where $f_t^*(\theta_{St-1}; \underline{z}_t)$ is potentially random due to the short-horizon investor's uncertainty at date $t-1$ about z_t in $\underline{z}_t$ given the history $\underline{z}_{t-1}$. Analogously to the proof of Proposition 1, the RHS of (53) can be expresses as a decreasing function $q_{t-1}(\theta_{St-1}; \underline{z}_{t-1})$ which lets us rewrite (53) as

$$v_c(e_{t-1,t-1} - P_{t-1}\theta_{St-1})\, P_{t-1} = q_{t-1}(\theta_{St-1}; \underline{z}_{t-1}). \tag{54}$$

Applying the implicit function theorem to (54) gives cohort $t-1$'s demand function $\theta_{St-1} = h_{t-1}(P_{t-1}; \underline{z}_{t-1})$. When P_{t-1} is below the critical price $P^*_{t-1} = q_{t-1}(0; \underline{z}_{t-1})/v_c(e_{t-1,t-1})$ corresponding to $\theta_{St-1} = 0$, their position is $\theta_{St-1} > 0$ and monotone in P_{t-1} and when $P_{t-1} > P^*_{t-1}$ then $\theta_{St-1} < 0$ but potentially non-monotone. In the limit as $P_{t-1} \downarrow 0$ we have $\theta_{St-1} \uparrow \frac{e_{t-1,t-1}}{P_{t-1}} \uparrow \infty$ since this increases $v_c(e_{t-1,t-1} - P_{t-1}\theta_{St-1})$ on the LHS from the Inada conditions and causes $E_{S1}\left[v_c(e_{t-1,t} + f_t^*(\theta_{St-1}; \underline{\tilde{z}}_t)\theta_{St-1}) f_t^*(\theta_{St-1}; \underline{\tilde{z}}_t)\right] \uparrow \infty$ on the RHS of (53) since $P_t = f_t^*(\theta_{St-1}; \underline{\tilde{z}}_t) \downarrow 0$ and since $P_t\theta_{St-1} = f_t^*(\theta_{St-1}; \underline{\tilde{z}}_t)\theta_{St-1} \uparrow \infty$ from Step 1 and, thus, $v_c(e_{t-1,t} + \tilde{P}_t\theta_{St-1}) \downarrow 0$.

Step 3. Applying the implicit function theorem to (50) for date $t-1$ gives the long-horizon investors' net demand $\theta_{St-1} = f_{t-1}(P_{t-1}; z_{t-1})$. We can show that f_{t-1} is roughly increasing relative to the size bond position θ_{St-2} being closed out by cohort $t-2$ and that f_{t-1} goes to $-\infty$ as $P_{t-1} \downarrow 0$ and that $f_{t-1}(P_{-1}) > 0$ for prices P_{t-1} sufficiently large. Hence, there must be an intersection between h_{t-1} and g_{t-1} where both θ_{St-1} and P_{t-1} are functions of θ_{St-2}.

Step 4. Writing the equilibrium short-horizon bond position as $\theta_{St-1} = w_{t-1}(\theta_{St-2}; \underline{z}_{t-1})$, we complete the recursion by showing that w_{t-1} has the required slope property assumed in Step 1 for w_t. Note that a perturbation in the incoming bonds θ_{St-2} leads to a one-to-one displacement of θ_{St-1} in the long-horizon investor's supply curve. However, since h_{t-1} is a well-defined function rather than a correspondence, its slope cannot be infinite. Thus, any vertical displacement in g_{t-1}

leads to a less than one-to-one displacement in the intersection between g_{t-1} and h_{t-1}. Hence, $w_{t-1}(\theta_{St-2} + \Delta; \underline{z}_{t-1}) - w_{t-1}(\theta_{St-2}; \underline{z}_{t-1}) < \Delta$. ■

References

Acharya, V., and L. Pedersen, 2003, Asset Pricing with Liquidity Risk, working paper, New York University.

Amihud, Y. and H. Mendelson, 1986, Asset Pricing and the Bid–Ask Spread, *Journal of Financial Economics*, 17, 223–249.

Bennett, J., R. Sias, and L. Starks, 2003, Greener Pastures and the Impact of Dynamic Institutional Preferences, *Review of Financial Studies*, 16, 1203–1238.

Brennan, M., and A. Subrahmanyam, 1996, Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns, *Journal of Financial Economics*, 41, 441-464.

Campbell, J., and J. Cochrane, 1999, By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior, *Journal of Political Economy*, 107, 205–251.

Constantinides, G., 1986, Capital Market Equilibrium with Transaction Costs, *Journal of Political Economy*, 94, 842–862.

Constantinides, G., and D. Duffie, 1996, Asset Pricing with Heterogeneous Consumers, *Journal of Political Economy*, 104, 219–240.

Chordia, T., R. Roll, and A. Subrahmanyam, 2000, Commonality in Liquidity, *Journal of Financial Economics*, 56, 3–28.

Detemple, J., 2002, Asset Pricing in an Intertemporal Partially-Revealing Rational Expectations Equilibrium, *Journal of Mathematical Economics*, 38, 219–248.

Detemple, J., and L. Selden, 1991, A General Equilibrium Analysis of Option and Stock Market Interactions, *International Economic Review*, 32, 279–303.

Duffie, D., and C.-F. Huang, 1985, Implementing Arrow-Debreu Equilibria by Continuous Trading of Few Long-Lived Securities, *Econometrica*, 53, 1337–1356.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, Is Information Risk a Determinant of Asset Returns?, *Journal of Finance*, 57, 2185–2221.

Evans, M., and R. Lyons, 2002, Order Flow and Exchange Rate Dynamics, *Journal of Political Economy*, 170–180.

French, K., and R. Roll, 1986, Stock Return Variances: The Arrival of Information and the Reaction of Traders, *Journal of Financial Economics*, 17, 5–26.

Gennotte, G., and H. Leland, 1990, Market Liquidity, Hedging, and Crashes, *American Economic Review*, 999–1021.

Glosten, L., and P. Milgrom, 1985, Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders, *Journal of Financial Economics*, 21, 123–144.

Grossman, S., 1988, An Analysis of the Implications for Stock and Futures Price Volatility of Program Trading and Dynamic Hedging Strategies, *Journal of Business*, 61, 275–298.

Grossman, S., and J. Stiglitz, 1980, On the Impossibility of Informationally Perfect Markets, *American Economic Review*, 70, 393–408.

Grundy, B., and M. McNichols, 1989, Trade and Revelation of Information through Prices and Direct Disclosure, *Review of Financial Studies*, 2, 495–526.

Hasbrouck, J., 2004, Trading Costs and Returns for US Equities: The Evidence from Daily Data, working paper, New York University.

Hasbrouck, J., and D. Seppi, 2001, Common Factors in Prices, Order Flows and Liquidity, *Journal of Financial Economics*, 59, 383–411.

Heaton, J., and D. Lucas, 2000, Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk, *Journal of Finance*, 55, 1163–1198.

Huang, M., 2003, Liquidity Shocks and Equilibrium Liquidity Premia, *Journal of Economic Theory*, 109, 104–129.

Huang, A., E. Hughson, and J. Leach, 2003, Risk Aversion, Regimes, and Returns: Revisiting the Equity Premium Puzzle, working paper, University of Colorado at Boulder.

Jagannathan, R., and Z. Wang, 1996, The Conditional CAPM and the Cross-section of Expected Returns, *Journal of Finance*, 51, 3–53.

Kraus and Smith, 1989, Market Created Risk, *Journal of Finance*, 44, 557–569.

Kraus, A., and M. Smith, 1996, Heterogeneous Beliefs and the Effect of Replicable Options on Asset Prices, *Review of Financial Studies*, 9, 723–756.

Kraus, A., and M. Smith, 1998, Endogenous Sunspots, Pseudo-Bubbles, and Belief about Beliefs, *Journal of Financial Markets,* 1, 151–174.

Kyle, A., 1985, Continuous Auctions and Insider Trading, *Econometrica*, 53, 1315–1335.

Lamoureux, C., and W. Lastrapes, 1990, Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects, *Journal of Finance*, 45, 221–229.

Leach, J. and A. Madhavan, 1992, Intertemporal Price Discovery by Market Makers: Active versus Passive Learning, *Journal of Financial Intermediation*, 2, 207–235.

Lesmond, D., J. Ogden, C. Trzcinka, 1999, A New Estimate of Transaction Costs, *Review of Financial Studies*, 12, 1113–1141.

Lo, A., and J. Wang, 2000, Trading Volume: Definitions, Data Analysis and Implications of Portfolio Theory, *Review of Financial Studies*, 13, 257–300.

Madrigal, V., 1996, Non-Fundamental Speculation, *Journal of Finance*, 51, 553–578.

Madrigal, V. and J. Scheinkman, 1997, Price Crashes, Information Aggregation, and Market Making, *Journal of Economic Theory*, 75, 16–63.

Novy-Marx, R., 2004, On the Excess Returns to Illiquidity, working paper, University of Chicago.

O'Hara, M., 2003, Liquidity and Price Discovery, *Journal of Finance* 58, 1335–1354.

Pastor, L., and R. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy*, 111, 642–685.

Radner, R., 1972, Existence of Plans, Prices, and Price Expectations in a Sequence of Markets, *Econometrica*, 44, 289–303.

Saar, G., 2001, Investor Uncertainty and Order Flow Information, working paper, New York University.

Smith, R., 1993, Market Risk and Asset Prices, *Journal of Economic Dynamics and Control*, 17, 555–569.

Strotz, R., 1956, Myopia and Inconsistency in Dynamic Utility Maximization, *The Review of Economic Studies*, 23, 165–180.

Vanden, J., 2004, Option Trading and the CAPM, *Review of Financial Studies*, 17, 207–238.

Vayanos, D. 1998, Transaction Costs and Asset Prices: A Dynamic Equilibrium Model, *Review of Financial Studies*, 11, 1–58.

Vayanos, D., 1999, Strategic Trading and Welfare in a Dynamic Market, *Review of Economic Studies*, 66, 219–245

Vayanos, D., 2001, Strategic Trading in a Dynamic Noisy Market, *Journal of Finance*, 56, 131–171.

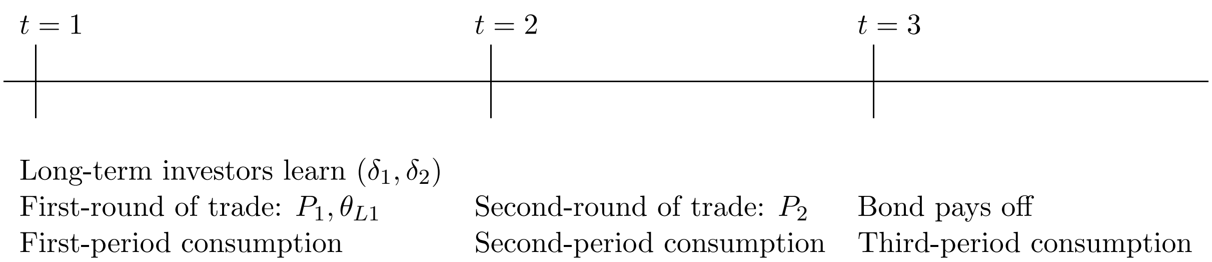

Figure 1: **Model Time-Line**

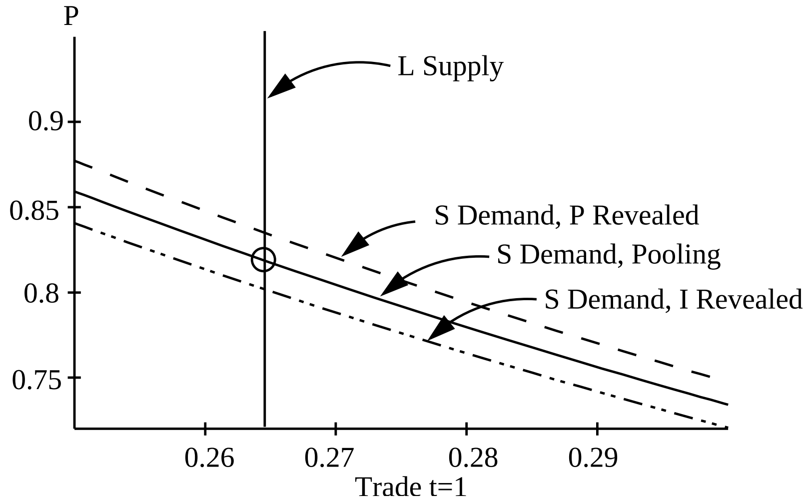

Figure 2a: **Example of a Partially Revealing Outcome — First Period.**

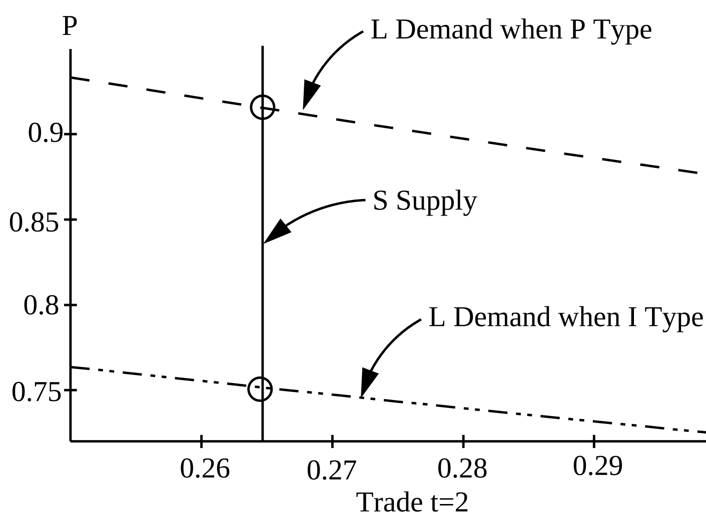

Figure 2b: **Example of a Partially Revealing Outcome — Second Period.**

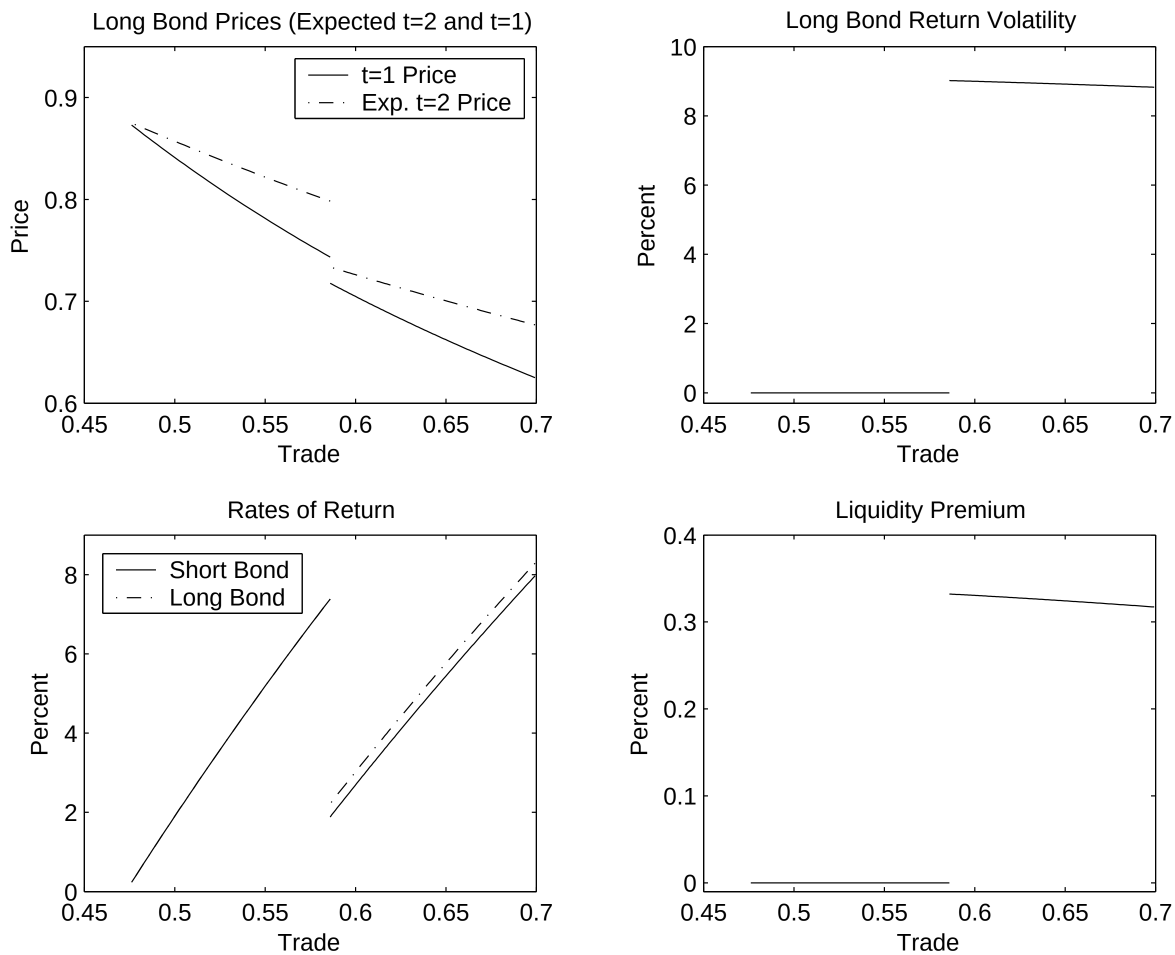

Figure 3: **Log Preferences and Uniform Priors Example.** The long-horizon and short-horizon investors have logarithmic preferences. The long-horizon investor's time preference parameter δ_2 is binomial with $\delta_2^I = 1.1$ and $\delta_2^P = 1.5$. Each δ_2 is equally likely. Conditional on $\delta_2 = 1.1$ ($\delta_2 = 1.5$), δ_1 is uniformly distributed on the interval $[1.3, 1.55]$ ($[0.95, 1.4]$). Other parameters are $\beta = 1.2$, $e_{S1} = 1.4$, $e_{S2} = 0.6$, $e_{L1} = 0.0$, and $e_{L2} = 1.0$.

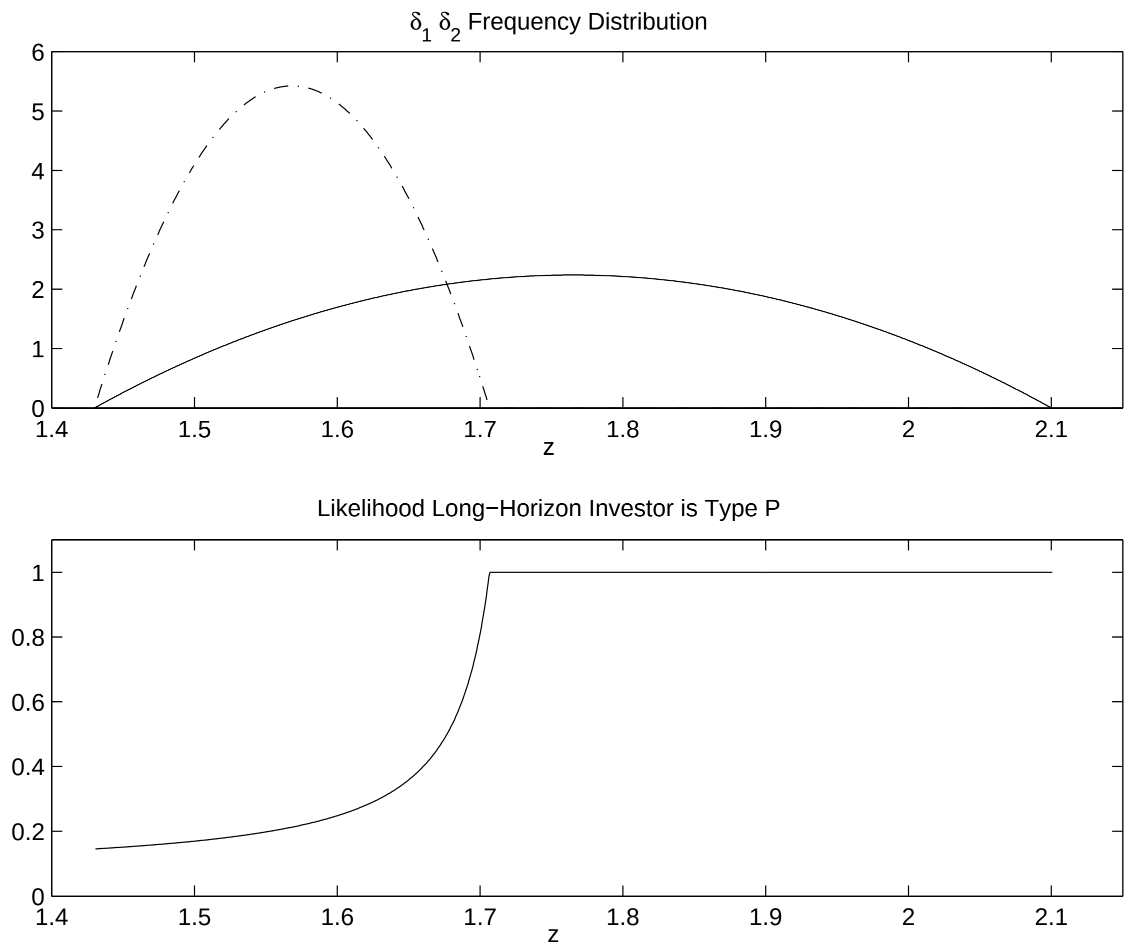

Figure 4: **Frequency Distributions for $\delta_1\delta_2$ with Beta Priors.** The time preference parameter δ_2 is binomial with $\delta_2^I = 1.1$ and $\delta_2^P = 1.5$. Each δ_2 is equally likely. Conditional on $\delta_2 = 1.1$ ($\delta_2 = 1.5$), δ_1 is beta (2, 2) distributed on the interval $[1.3, 1.55]$ ($[0.95, 1.4]$).

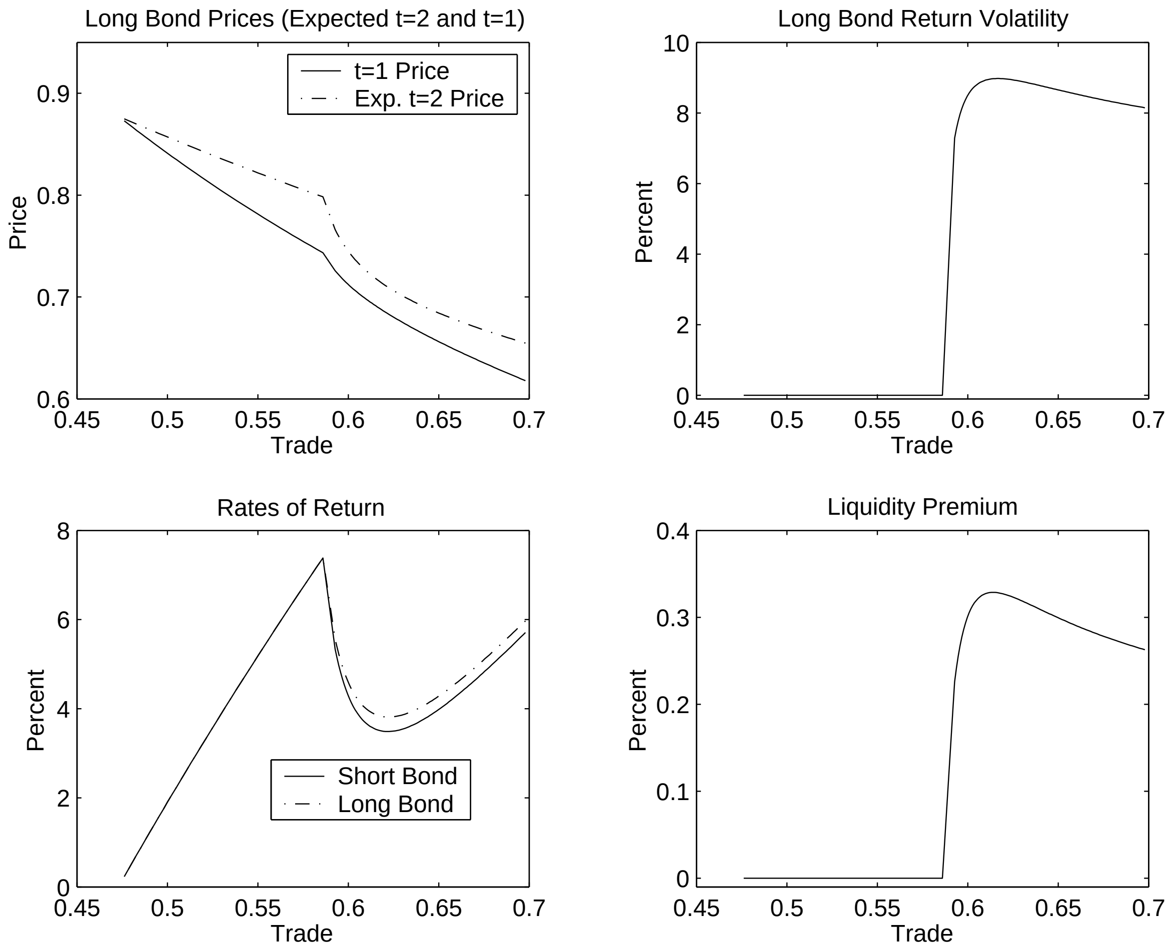

Figure 5: **Log Preferences and Beta Priors Example.** The long-horizon and short-horizon investors have logarithmic preferences. The time preference parameter δ_2 is binomial distributed with $\delta_2^I = 1.1$ and $\delta_2^P = 1.5$. Each δ_2 is equally likely. Conditional on $\delta_2 = 1.1$ ($\delta_2 = 1.5$), δ_1 is beta (2, 2) distributed on the interval $[1.3, 1.55]$ ($[0.95, 1.4]$). Other parameters are $\beta = 1.2$, $e_{S1} = 1.4$, $e_{S2} = 0.6$, $e_{L1} = 0.0$, and $e_{L2} = 1.0$.

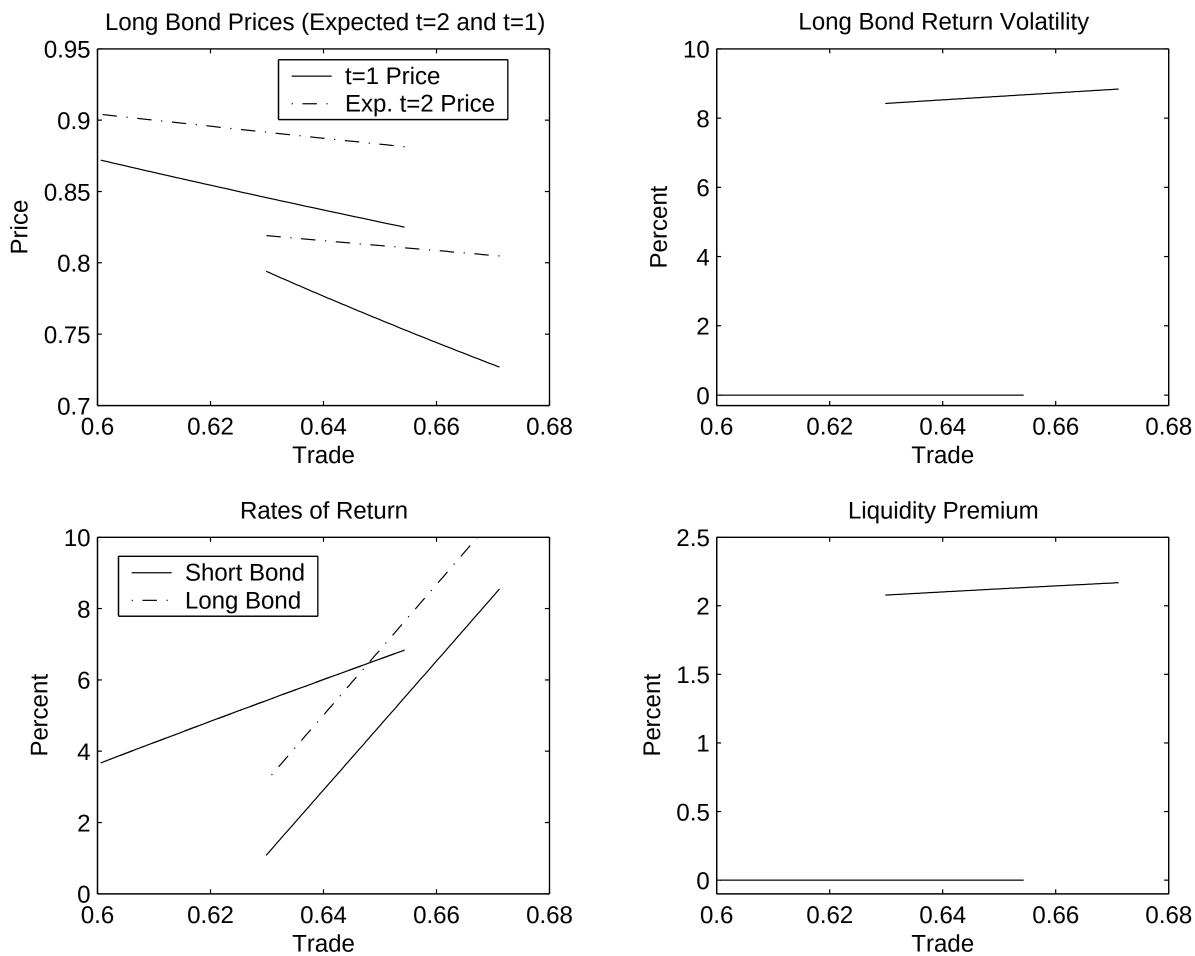

Figure 6: **Square Root/Power Preferences and Uniform Priors Example.** The long-horizon and short-horizon investors have constant relative risk averse preferences with coefficients of relative risk aversion of 0.5 and 3 respectively. The time preference parameter δ_2 is binomial distributed with $\delta_2^I = 0.8$ and $\delta_2^P = 1.0$. Each δ_2 is equally likely. Conditional on $\delta_2 = 0.8$ ($\delta_2 = 1.0$), δ_1 is uniformly distributed on the interval $[1.1875, 1.25]$ ($[0.9, 1.0]$). Other parameters are $\beta = 1.0$, $e_{S1} = 1.1$, $e_{S2} = 0.25$, $e_{L1} = 0.0$, and $e_{L2} = 1.0$.

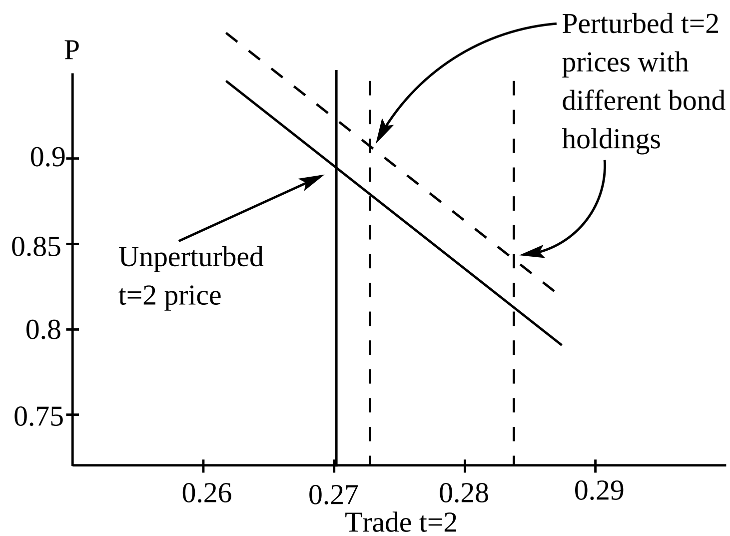

Figure 7: **Ambiguous Impact of δ_2 Perturbations on Equilibrium Prices at Date 2.**

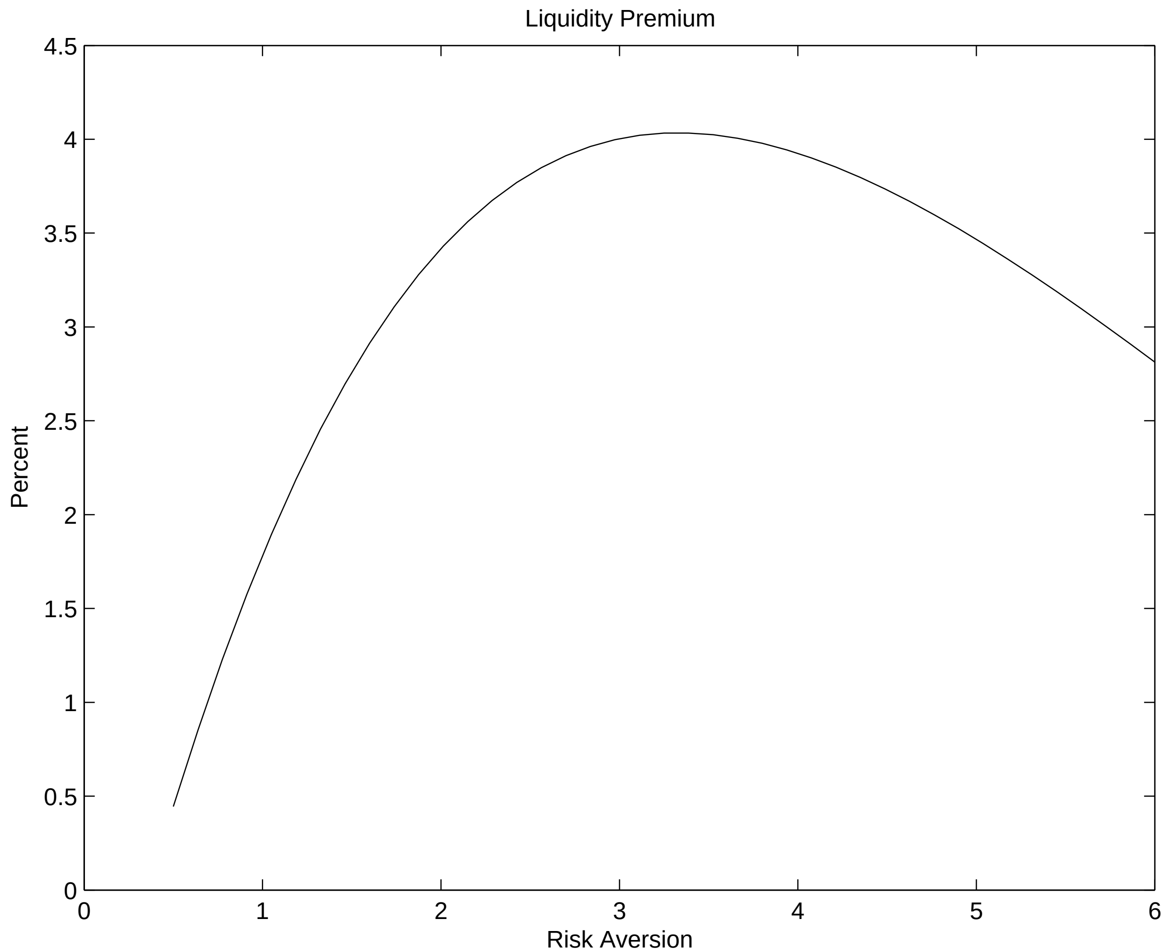

Figure 8: **Short-Horizon Investor Risk Aversion and the Liquidity Premium.** The liquidity premium of the bond is plotted as a function to the short-horizon investors' risk aversion. The long-horizon investors have constant relative risk averse preferences with a coefficient of relative risk aversion $\gamma = 0.5$. The short horizon investors have constant relative risk averse preferences with a coefficient of relative risk aversion $\gamma = [0.5, 6]$. The time preference parameter δ_2 is binomial distributed with $\delta_2^I = 1.3$ and $\delta_2^P = 2.0$. Each δ_2 is equally likely. The equilibria plotted correspond to partially revealing equilibria where $z = \delta_1 \, \delta_2 = 1.6$. Other parameters are $\beta = 1.0$, $e_{S1} = 0.4$, $e_{S2} = 0.0$, $e_{L1} = 0.0$, and $e_{L2} = 5.0$.

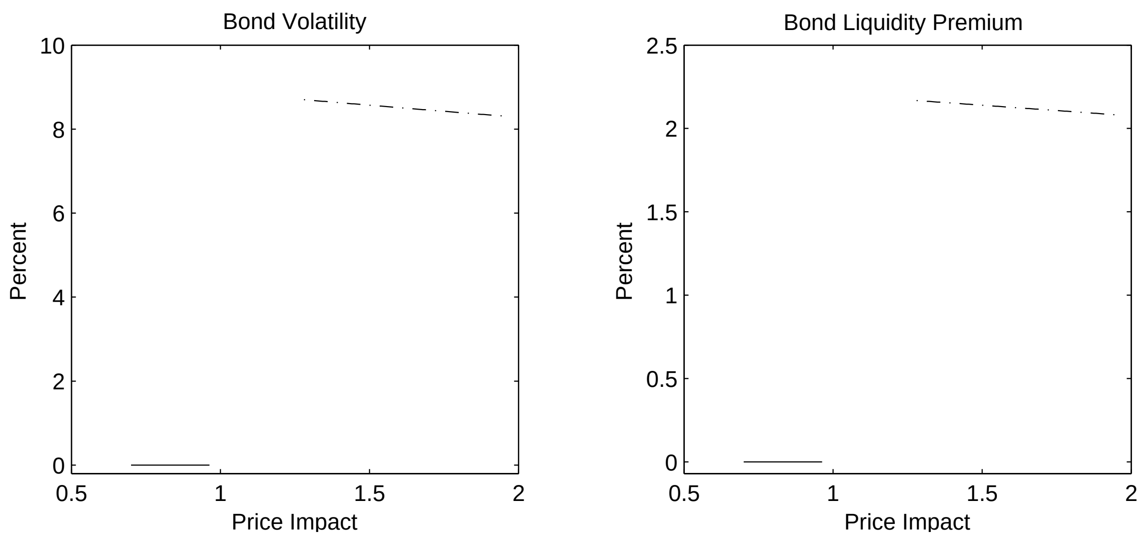

Figure 9: **Return Characteristics versus Price Impact.** The bond return volatility and liquidity premium are plotted versus the corresponding levels of the price impact of order flow $\frac{|P_1(\theta_{L1})-P_1(\min\theta_{L1})|}{\theta_{L1}-\min\theta_{L1}}$. The parameter values are the same as in Figure 6. The long-horizon and short-horizon investors have constant relative risk averse preferences with coefficients of relative risk aversion of 0.5 and 3 respectively. The time preference parameter δ_2 is binomial distributed with $\delta_2^I = 0.8$ and $\delta_2^P = 1.0$. Each δ_2 is equally likely. Conditional on $\delta_2 = 0.8$ ($\delta_2 = 1.0$), δ_1 is uniformly distributed on the interval $[1.1875, 1.25]$ ($[0.9, 1.0]$). Other parameters are $\beta = 1.0$, $e_{S1} = 1.1$, $e_{S2} = 0.25$, $e_{L1} = 0.0$, and $e_{L2} = 1.0$.