Dynamic Thin Markets[1]

Marzena Rostek[2] and Marek Weretka[3]

December 15, 2008

Abstract. Extensive empirical research has shown that in many markets institutional investors have a significant impact on prices and mitigate its adverse effects through their trading strategies. This paper develops a dynamic model of such thin markets, in which the market structure is one of bilateral oligopoly. The paper demonstrates that market thinness qualitatively changes equilibrium properties of prices and dynamic trading strategies, compared to the existing theories of asset pricing. The predictions match a number of empirical facts that are hard to reconcile with the competitive or Cournot-based models. The paper further establishes that the nonstrategic general-equilibrium approach and the strategic approach to trade via Nash in demands are dual representations of a model with endogenous price impact. The proposed approach yields an analytical framework that can be used to study dynamic markets with bilateral market power.

JEL CLASSIFICATION: D43, D53, G11, G12, L13

KEYWORDS: Thin Markets, Mispricing, Price Impact, Nash in Demands.

1 Introduction

Since trade-level data first became available two decades ago, it has now become well understood that transactions by institutional investors exert an economically significant price impact in many financial markets.[4] Because the stock positions of institutional investors constitute a sizable fraction of total trading, these trades often exceed the number of shares that the market maker is willing to trade at the quoted bid and ask prices. As a result, the trades are likely to move prices, thereby adversely affecting the terms of trade. In fact, the adverse effects of market power measured in the so-called implicit trading costs dominate the explicit costs of trading, such as commission fees and

[1] We would like to thank William Brock, Zhiwu Chen, Peter DeMarzo, Steven Durlauf, Frank Fabozzi, Christian Ghiglino, Ananth Madhavan, Antonio Mello, Matthew Pritsker, Larry Samuelson, Elu Von Thaden, and especially John Geanakoplos and Dimitri Vayanos for very helpful conversations. We are also grateful for the comments and suggestions offered by seminar participants at Chicago, Johns Hopkins, Mannheim (Finance), Oxford (Nuffield), Penn State, Princeton, Queen's, Southampton, UCL, Warwick (CRETA) and Wisconsin-Madison, and conference participants at the 2008 AFA Annual Meeting in New Orleans, the Essex Economics Workshop '2008, the 2007 ES Winter Meeting in Chicago and the 2007 ES Summer Meeting at Duke. All errors are ours.

[2] University of Wisconsin-Madison, Department of Economics, 1180 Observatory Drive, Madison, WI 53706, U.S.A. E-mail: `mrostek@ssc.wisc.edu`.

[3] University of Wisconsin-Madison, Department of Economics, 1180 Observatory Drive, Madison, WI 53706, U.S.A. E-mail: `mweretka@ssc.wisc.edu`.

[4] See, e.g., Kraus and Stoll (1972); Holthausen, Leftwich, and Mayers (1987); Chan and Lakonishok (1993, 1995); Keim and Madhavan (1995, 1996, 1998).

order-processing fees (e.g., Chan and Lakonishok [1995]; Stoll [1978]; Keim and Madhavan [1995, 1996,1998]). Techniques used to estimate market impact and facilitate trading are widespread in investment management and are available in the Market Impact Models offered by Citigroup, EQ International, ITG, MCI Barra, and OptiMark, among others. Markets in which individual trades are large relative to the average daily volume and, hence, have impact on prices are known as *thin markets*.

Widely discussed in the literature are a number of stylized facts regarding trading strategies and price behavior in thin markets that appear inconsistent with the traditional modeling of asset pricing. One common practice involves breaking up orders into *blocks*, which are then traded sequentially (Fact 1). Even in markets as deep as the NYSE, only about 20% of the value of institutional purchases and sales is completed within a single day, while more than 50% of that value takes at least four days for execution (Chan and Lakonishok [1995]). Another large body of evidence concerns the reaction of prices in thin markets to supply or demand shocks. Typical shocks examined would be forced liquidations, issuance of new debt, selling Initial Public Offerings (IPOs), inclusions of new stocks into stock market indices, such as S&P, or changes of index weights. Such shocks typically lead to a significant price change followed by a partial reversal of the price in subsequent periods. Thus, apart from the permanent effect, the resulting price adjustment also has a temporary, overshooting component (Fact 2). The two effects were first empirically discovered by Kraus and Stoll (1972) and subsequently confirmed by numerous studies for various securities. Crucially, in the data, the temporary price change occurs on the date of the shock even if the shock was pre-announced.[5] The evidence on market reaction to supply and demand shocks is striking because it points out that the trade announcement and the trade-induced price effect can be separated in time (Fact 3); and *anticipated* price changes can be observed in markets (Fact 4). In the standard competitive model, the no-arbitrage property rules out phenomena (2)-(4); Pareto efficiency after the first round of trade precludes order breakup (1) as an equilibrium prediction.

The goal of this paper is to understand the implications of market thinness for the equilibrium behavior of prices and trades. For example, we examine whether a number of empirical facts observed in financial markets, including the four above, can be explained solely by moving away from the assumption of price taking, while maintaining all of the other assumptions of a perfect-foresight general-equilibrium model, in particular, (dynamic) optimization by all traders.

To explain the bite of our approach, let us describe how it differs from the large body of research that has emerged to explore why and how price impact affects individual portfolio choices and equilibrium in financial markets. The range of different approaches in the existing literature can be grouped around two theoretical mechanisms. Traditionally, price impact has been attributed

[5] Indeed, pre-announced changes of weights in stock market indices have a significant price effect on the day of the inclusion. Such natural experiments that allow controlling for the informational component of the price change were studied for stocks and currencies or foreign equity by Kaul, Mehrotra, and Morck (2000); Hau and Rey (2004); Loderer, Cooney, and Van Drunen (1991); and Hau, Massa, and Peress (2005).

to asymmetric or private information (e.g., Glosten and Milgrom [1985]; Kyle [1985, 1989]; Easley and O'Hara [1987]; Back [1992]; Foster and Viswanathan [1996]; Holden and Subrahmanyam [1996]). Empirical studies suggest, however, that, in many trade settings, the price impact component that is due to asymmetric information can only partially account for the observed magnitudes of price changes.[6] In this paper, market thinness does not result from asymmetric information.

TM-CAPM belongs to the strand of research that has examined inventory effects as a source of price impact. The basic mechanism is as follows: when an investor's trading partners have decreasing marginal utility (i.e., risk averse traders), selling or buying requires price concessions in order for the other traders to be willing to absorb the investor's order. Within the inventory literature, incorporating price impact into asset pricing has been often accomplished by building a Monopoly/Cournot-type model with $I \geq 1$ large investors trading with a fringe of price-taking traders (e.g., Ho and Stoll [1981]; Grossman and Miller [1988]; Vayanos [2001]; Attari, Mello and Ruckes [2005]; Brunnermeier and Pedersen [2005]; Pritsker [2005]; and DeMarzo and Urošević [2006] extended by Urošević [2005]). The equilibrium implications of the Cournot-based asset pricing models do not, however, accommodate stylized facts (1)-(4) in a perfect foresight setting as long as competitive agents are rational. With a competitive fringe of buyers, the price instantaneously adjusts to a new fundamental value.[7] In fact, we demonstrate that the presence of even one rational trader who takes prices as given in a perfect-foresight model not only rules out facts (1)-(4) as equilibrium phenomena, but also implies that dynamic Cournot markets yield competitive outcomes. It follows that phenomena (1)-(4) can be ruled out by the same arguments as in the competitive model, i.e., no-arbitrage and Pareto efficiency. Therefore, allowing all traders to have price impact is essential to modeling dynamic thin markets, which is the approach taken in this paper.

Our market structure is one of bilateral oligopoly rather than Cournot. The key feature of our model, relative to prior research, is that *all* traders are large relative to the market size and are risk averse. Thus, we take the view that the equilibrium properties of prices and trading strategies are determined by the behavior of large investors who constantly monitor prices, provide liquidity and can take advantage of any price differentials that may arise in financial markets.[8] The trade by small competitive traders is seen as exogenous shocks in demand or supply, which we incorporate in the second part of the paper.

[6] See e.g., Fact 3 and Fact 4. In addition, large institutional investors do not outperform fixed benchmark portfolios, which would likely be the case if they had superior information about asset fundamentals. Madhavan and Cheng (1997) report that for the average trade value, the price impacts in the downstairs markets do not differ significantly from those in the upstairs markets, which are more transparent and less susceptible to informational asymmetries. This suggests that in both types of markets, the price impact is not mainly driven by asymmetry of information.

[7] Brunnermeier and Pedersen (2005) propose a model of predatory trading, based on Cournot, which does lead to price overshooting. In their model, the exogenous demand is generated by long-term investors who, by assumption, do not take advantage of the short-term price differentials. If the traders were optimizing dynamically, overshooting would not arise, for otherwise, the traders could make infinite profits by taking unbounded positions.

[8] Incidentally, several non-equilibrium models with price impact have been proposed, mostly by practitioners (e.g., Bertsimas and Lo [1998]; Almgren and Chriss [2000]; Almgren et al. [2005]; Subramanian and Jarrow [2001]; Dubil [2002]; Almgren [2003]; Huberman and Stanzl [2004]). These models assume exogenous price impact functions, which are then used to analyze prices and allocations in thin markets without solving for equilibrium.

In order to delineate how the mere presence of price impact affects equilibrium, we consider an otherwise standard CAPM setting with mean-variance optimizers. The standard competitive CAPM is encompassed as a limit case of our model in large markets, which we call *Thin-Markets CAPM* (*TM-CAPM*). We establish that when all traders have price impact, the equilibrium properties of prices and trades are qualitatively different from those observed in perfect-foresight competitive or Cournot models with dynamically optimizing traders. This then allows us to match a number of empirical facts that cannot be reconciled with any model that assumes competitive traders who optimize dynamically. While in this paper, the equilibrium properties are derived in the context of asset pricing, they naturally extend to other dynamic bilaterally oligopolistic markets with multi-unit demands. We provide a closed-form characterization of equilibrium trades, prices, and price impacts in dynamic trading environments with bilateral market power.

TM-CAPM is closest to Vayanos (1999), who also examined a rational expectations model with all strategic traders. Compared to the literature, our model has the following distinguishing features which allow us to uncover novel effects in dynamic thin markets.

We show that *any* exogenous demand or supply shock in a thin market modeled by TM-CAPM has two effects on prices – the *fundamental* and *liquidity* effects – which differ in their origin, persistence, and timing. The fundamental effect, which is permanent, reflects the adjustment of the fundamental value that results from the change in the average inventory in the market. This effect would be observed even in markets with price-taking liquidity providers. The fundamental effect is amplified by a temporary liquidity effect, which results from noncompetitiveness of market interactions. The permanent effect always occurs immediately after the investors learn about the shock. Consistent with the data, the temporary effect attains the maximum at the moment of trade whether or not the shock is anticipated. Furthermore, in the presence of the liquidity effect, phenomena (2)-(4) arise naturally on the equilibrium path.

We allow the trade to take place more frequently than dividend payments, which typically occur semiannually, rather than with equal frequency. In a competitive model without any shocks and discounting, allowing for multiple trading opportunities would not have any effect on the asset allocation at maturity. Strikingly, in thin markets, the higher frequency qualitatively changes the properties of equilibrium, and to emphasize this, we call it *frequent trading.* Specifically, we show that the endogenously determined market depth (price impact) is not constant, but evolves over time, even in the absence of shocks or information revelation. The market depth changes because, with frequent trading, time-to-maturity varies across trading periods and, therefore, trading rounds offer different diversification (resale) opportunities for liquidity providers prior to maturity. This, in turn, affects the investors' willingness to absorb orders placed by other traders and translates into different price concessions. We show that the endogenous nonstationarity of price impact, combined with the liquidity effect identified by TM-CAPM, generates a number of empirically documented phenomena. These include changes in price volatility unrelated to changes in fundamentals and volatility clustering. Moreover, the model can also rationalize the empirical evidence on the shape of price impact – i.e., why the permanent price impact function is typically estimated as a linear function of the block size, while its temporary counterpart appears to be a concave function of the size of a block.[9]

Our model allows for many risky assets. This is particularly important in determining asset pricing formulas for assets traded in thin markets (e.g., blockage discount). TM-CAPM allows deriving formulas that account for the adverse effect of market thinness on the market value of the entire portfolio. Additionally, with many assets, the model identifies that simultaneous shocks in different markets increase the overall price volatility via cross-market price impact effects.

Finally, our work offers several methodological advances to research on price impact. We retain a general-equilibrium structure of the model. TM-CAPM is, thus, related to the literature on noncompetitive general equilibrium, initiated by Negishi (1960) and followed by Arrow and Hahn (1971), Hahn (1977), and Hart (1995). Unlike these models, the equilibrium in our model is determinate, which allows us to give predictions about the behavior of prices and trades in thin markets. Furthermore, although not defined as a Nash in the game, the noncompetitive equilibrium from this paper can be re-cast as Nash in a game in which traders submit demand functions. The game of Nash in demands (or supplies) was introduced by Grossman (1981) and further developed in the seminal papers by Kyle (1989) in finance and by Klemperer and Meyer (1989) in industrial organization. We establish that, within the CAPM setting, there exists a bijection between Nash equilibrium in linear demands and the noncompetitive equilibrium. The result demonstrates that the strategic and general-equilibrium approaches to market interactions are dual representations. The duality implies that all the results established in this paper apply to and extend those obtained in the literature on games in demand functions.

2 A Model of Frequent Trading in Thin Markets

2.1 Market Microstructure

There are I traders, also called *liquidity providers*, where I can be a small number. With the usual abuse of notation, I will also denote the set of traders. Investment opportunities include N risky assets and one riskless asset (e.g., a bond). Investors can trade for T trading rounds, after which assets mature and dividends are paid. The dividends from risky assets, are distributed normally according to $\mathcal{N}(A, \mathcal{V})$, where A is the vector of the expected asset payoffs and $\mathcal{V}$ is the (symmetric and positive definite) variance-covariance matrix of payoffs. For notational convenience, we assume that the interest rate on the riskless asset is zero. Alternatively, the riskless asset can be interpreted as money.

Liquidity providers enter each period $t = 1, 2, \ldots, T$ with stocks of risky assets $\theta^i_{t-1} \in \mathbb{R}^N$ and bonds $\theta^i_{b,t-1} \in \mathbb{R}$. After they trade $\Delta\theta^i_t$ and $\Delta\theta^i_{b,t}$ in stocks and bonds, respectively, the liquidity

[9] The following academic and non-academic papers considered various classes of the functional forms of price impact: Bertsimas and Lo (1998); Almgren and Chriss (2000); Subramanian and Jarrow (2001); Dubil (2002); Almgren (2003); and Obizhaeva and Wang (2005).

providers end the trading period t with holdings of $\theta_t^i = \theta_{t-1}^i + \Delta\theta_t^i$, and $\theta_{b,t}^i = \theta_{b,t-1}^i + \Delta\theta_{b,t}^i$, with which they enter $t+1$. Trades $\Delta\theta_t^i$ and $\Delta\theta_{b,t}^i$ denote net demands in period t. $(\theta_0^i, \theta_{b,0}^i)$ denotes the exogenously given initial inventory of trader i. Investors choose their trades to maximize the expected CARA utility functions. By the standard argument, such assumptions are jointly equivalent to assuming that investors are mean-variance optimizers; that is, investor's i indirect utility function, expressed in terms of after-trade portfolios, is linear in bond holdings and quadratic in risky assets,

$$U(\theta_T^i, \theta_{b,T}^i) = \theta_{b,T}^i + A \cdot \theta_T^i - \frac{\alpha}{2}\theta_T^i \cdot \mathcal{V}\theta_T^i. \tag{1}$$

It is useful to define a *market participation rate* as

$$\gamma \equiv 1 - \frac{1}{I-1}. \tag{2}$$

The closer γ is to one, the more participants trade in the market and the more competitive the market interaction. With only two traders, γ is equal to zero. Equilibrium does not exist in this case.[10] In hindsight, traders' price impacts are shown to be mutually reinforcing, and since, with bilateral trade, the reinforcement occurs without any discounting (i.e., the buys or sells of one trader can be potentially absorbed by only one other trader), this leads to infinite price impacts. Let θ^{Av} denote an *average inventory*, which is defined as the portfolio held by all liquidity providers, evaluated in *per capita* terms,

$$\theta^{Av} \equiv \frac{1}{I}\sum_{i \in I} \theta_0^i. \tag{3}$$

Finally, $\{x_t\}_t$ denotes the sequence $\{x_1, ..., x_T\}$, $\{x^i\}_i$ denotes the sequence $\{x^1, ..., x^I\}$, and $\{x_t^i\}_{i,t}$ stands for the sequence $\{\{x^i\}_i\}_t$. Throughout, a bar "$-$" denotes equilibrium.

2.2 Equilibrium

In TM-CAPM, trades of the liquidity providers are large relative to the market size and can, therefore, exert a non-negligible impact on prices. The price impact of trader i is formalized as an $N \times N$ matrix $\mathcal{M}_t^i$, in which a typical element (n, m) characterizes the price change of asset m resulting from a marginal increase in demand for asset n. When $N = 1$ the matrix becomes a scalar and is equal to the slope of a one-dimensional residual demand. As long as the *price impact matrix* $\mathcal{M}_t^i$ is non-zero, the asset demands on which investor i operates are not perfectly elastic.

In this paper, the price impacts of each trader are not exogenous but are determined in equilibrium jointly with trades and prices. With equilibrium in every period t being now a triple $(\bar{p}_t, \Delta\bar{\theta}_t, \bar{\mathcal{M}}_t)$, the standard conditions – traders' optimization and market clearing – are supplemented by a consistency condition of price impacts.

[10] The non-existence of equilibrium with two traders is also present in closely related models by Kyle (1989), Klemperer and Meyer (1989) with a vertical demand, and Wilson (1979). Alternatively, the outcome for markets with two investors can be viewed as an equilibrium in which price impacts are infinite and imply that the trading costs exceed the potential gains to trade, making *status quo* (no trade) optimal.

Given the price observed in the market, p_t, a portfolio traded at this price, $\Delta\bar{\theta}^i_t$, and price impact matrix, $\mathcal{M}^i_t$, the investor faces a demand function,

$$p_{p_t,\Delta\bar{\theta}^i_t,\mathcal{M}^i_t}(\Delta\theta^i) = p_t + \mathcal{M}^i_t(\Delta\theta^i_t - \Delta\bar{\theta}^i_t). \tag{4}$$

In equilibrium, the conjectured price impacts should correspond to the traders' true price impacts. This condition defines a consistency restriction on matrices $\mathcal{M}^i_t$, formally stated below. For the sake of transparency, we first define equilibrium, assuming the notion of consistency; then the definition of consistency follows. Definition 1 modifies the concept of competitive equilibrium by allowing non-zero price impacts.

Definition 1 *For any period $t = 1, ..., T$, a vector $(\bar{p}_t, \Delta\bar{\theta}_t, \bar{\mathcal{M}}_t)$ is a* noncompetitive equilibrium *if*

(i) Asset markets clear, $\sum_i \Delta\bar{\theta}^i_t = 0$;

(ii) For any i, the trade $\Delta\bar{\theta}^i_t$ is optimal, given demand function $p_{\bar{p}_t,\Delta\bar{\theta}^i_t\bar{\mathcal{M}}^i_t}(\cdot)$;

(iii) For any i, price impact matrix $\bar{\mathcal{M}}^i_t$ is consistent with $\bar{\mathcal{M}}^{-i}_t$.

Next, we endogenize price impacts. Fix an arbitrary profile of price impacts of all traders but i, $\bar{\mathcal{M}}^{-i}_t$. To conceptualize the consistency of $\bar{\mathcal{M}}^i_t$, we consider how the market reacts to an investor's i deviation from his equilibrium trade $\Delta\bar{\theta}^i_t$ to any trade $\Delta\theta^i_t$. Competitive models assume that the effects of such a deviation on prices are at most negligible, and hence they can be ignored. Here, by contrast, the price decreases sufficiently to encourage the other traders to optimally absorb the additional asset supply. The extra units are thus sold at a price concession. As a result, all markets clear and all the other investors respond optimally to prices given assumed price impacts $\bar{\mathcal{M}}^{-i}_t$, even if investor i is trading a suboptimal quantity of shares. We say that any deviation $\Delta\theta^i_t$ by trader i triggers a subequilibrium that is defined as follows:

Definition 2 *Given $\bar{\mathcal{M}}^{-i}_t$, vector $(p^*_t, \Delta\theta^{-i*}_t, \bar{\mathcal{M}}^{-i}_t)$ is a* subequilibrium *triggered by trade $\Delta\theta^{i*}_t$ if:*

(i) Markets clear with the deviation, $\Delta\theta^{i*}_t + \sum_{j\neq i} \Delta\theta^{j*}_t = 0$;

(ii) For any $j \neq i$, trade $\Delta\theta^{j*}_t$ is optimal given demand functions $p_{p^*_t,\Delta\theta^{j*}_t,\bar{\mathcal{M}}^j_t}(\cdot)$.

If every deviation $\Delta\theta^{i*}_t$ of i triggers a unique subequilibrium, then trader i is effectively facing a downward sloping residual demand $p^i(\Delta\theta^{i*}_t)$, which assigns the market clearing price to any $\Delta\theta^{i*}_t$, and the slope of which measures i's price impact. Consistent price impact reflects the price change needed to clear the market for any possible deviation $\Delta\theta^{i*}_t$, given that the other traders respond optimally to market prices.

Definition 3 *$\bar{\mathcal{M}}^i_t$ is* consistent *with $\bar{\mathcal{M}}^{-i}_t$ if, for any deviation $\Delta\theta^{i*}_t$ of i, there exists a unique subequilibrium $(p^*_t, \Delta\theta^{-i*}_t, \bar{\mathcal{M}}^{-i}_t)$ such that*

$$p^*_t - \bar{p}_t = \bar{\mathcal{M}}^i_t\left(\Delta\theta^{i*}_t - \Delta\bar{\theta}^i_t\right). \tag{5}$$

In other words, the price impact of investor i is endogenized by being equated with the actual market responses needed to absorb his deviations. We then say that the profile of price impacts $(\bar{\mathcal{M}}_t^1, ..., \bar{\mathcal{M}}_t^I)$ is *consistent* if, for every i, $\bar{\mathcal{M}}_t^i$ is consistent with $\bar{\mathcal{M}}_t^{-i}$. Since, in the definition of equilibrium, the consistency is required to hold for all i, the consistency restriction is a fixed-point condition (in matrices), which will appear in predictions as a mutual interdependence of market power. The dynamic equilibrium is found by backward induction. Section 3, derives the unique symmetric dynamic equilibrium in closed form.

Interpretation of Trading Behavior and Equilibrium. One way to see how the equilibrium considered in this paper relates to the competitive equilibrium is to note that the traders are slope-takers rather than price-takers. That is, an investor assumes that changing his trading position does not affect the slope of his residual demand; the investor will affect the price as long as that slope is not zero. Similarly, at the level of individual behavior, the consistency condition implies that, when traders reoptimize in response to price concessions following a deviation of investor i, they assume that their price impact is as in equilibrium. If the endogenously derived price impacts are equal to zero, the competitive equilibrium obtains. This will be the case if there are infinitely many traders ($\gamma = 1$) or if traders are risk neutral ($\alpha = 0$). For interior parameter values, however, the predictions of TM-CAPM will differ from those in the competitive model.

In Section 8, we show that, within the CARA-Normal (or, more generally, a quadratic) framework, the equilibrium from this paper is outcome-equivalent to a Nash equilibrium in a game in which traders submit linear demand functions (e.g., Kyle [1989], Vayanos [1999]). One implication of the result is that our noncompetitive equilibrium, although not defined as a Nash equilibrium in a game, does have a game-theoretic foundation. Thus, TM-CAPM could be re-cast as a game in (net) demand functions. We have chosen to retain the general-equilibrium structure to highlight that the analysis can also be couched in terms of quantities, which allows us to make direct comparisons with the standard CAPM. Our approach is also more general than the linear Nash equilibrium – the latter does not survive beyond the quadratic setting,[11] whereas the equilibrium from this paper is well defined in an arbitrary quasilinear environment. Therefore, our approach permits studying the robustness of results to the assumption of quadratic utilities.

Since within CAPM, the relation between the representations is, in fact, one-to-one, the result further uncovers that our noncompetitive general-equilibrium approach and the strategic approach of trading in demands are dual representations of a model with endogenous market power. The link is revealed because of the manner in which we endogenize price impact. Section 8 explains the connection in detail and discusses broader implications of the duality. Here, we describe the interpretation of the trading behavior introduced by the dual representation. In the strategic as well as the non-strategic representation of our model, a trader can be viewed as trading against the market represented by a supply function, which is a horizontal sum of the demands of other

[11] That is, when other bidders submit linear bids, the best response is not linear. By contrast, outside of the quadratic setting, the concept of the noncompetitive equilibrium from this paper still pins down predictions and equilibrium is not linear.

traders. Strikingly, we show that the only information any investor needs to have in order to respond optimally to equilibrium orders, and in fact to the *arbitrary* orders of others, is his own preference and the slope of his own residual demand. In particular, no information about the number – let alone the utility functions, identities, or trading strategies of his trading partners – is required. Remarkably, the traders do not need to know the game (model). This property is particularly attractive in a one-period market; knowing one's preference in dynamic trading requires knowing one's value function (see Section 8). Our model thus fit particularly well anonymous markets in which investors have no other information but their past trades and market prices and discover their market power through statistical inference; each investor's price impact summarizes the information about the residual market against which he trades.

3 Market Thinness and Equilibrium

In this section, we derive the globally unique symmetric equilibrium in a model with two trading periods ($T = 2$) after which assets mature. We use the two-period model to explain the mechanisms operating in thin markets. Throughout, we highlight the differences with the competitive benchmark. In Appendix I, we derive equilibrium for an arbitrary T. The fundamental value profile $v \in \mathbb{R}^N$ is defined as the average marginal utility from risky assets,

$$v \equiv A - \alpha \mathcal{V} \theta^{Av}. \tag{6}$$

It is straightforward to show that the fundamental values coincide with the vector of prices from the competitive CAPM. We solve for equilibrium by backward induction, starting with the last trading period $t = 2$.

3.1 Equilibrium in the Last Period ($t = 2$)

Suppose that after trade in period one, trader i enters the second trading period with risky portfolio $\bar{\theta}_1^i$ and bonds $\bar{\theta}_{b,1}^i$. At $t = 2$, the indirect utility function of a mean-variance optimizing trader as a function of trade $(\Delta\theta_2^i, \Delta\theta_{b,2}^i)$ is given by

$$V_2^i(\Delta\theta_2^i, \Delta\theta_{b,2}^i) = \underbrace{\left(\bar{\theta}_{b,1}^i + \Delta\theta_{b,2}^i\right)}_{\theta_{b,2}^i} + A \cdot \underbrace{\left(\bar{\theta}_1^i + \Delta\theta_2^i\right)}_{\theta_2^i} - \frac{\alpha}{2} \underbrace{\left(\bar{\theta}_1^i + \Delta\theta_2^i\right)}_{\theta_2^i} \cdot \mathcal{V} \underbrace{\left(\bar{\theta}_1^i + \Delta\theta_2^i\right)}_{\theta_2^i} \tag{7}$$

The indirect utility function is linear in trade of bonds $\Delta\theta_{b,2}^i$ and, because of risk aversion, it is quadratic in the trade of risky assets $\Delta\theta_2^i$.

Portfolio Choice. The optimal trades in assets are characterized by the equality between each investor's marginal utility and marginal payment (or revenue if he sells), which, unlike in the competitive market, depends on the quantity traded.[12] Given his – for now, taken to be exogenous – price impact $\mathcal{M}_2^i$, trader i behaves as a residual monopsonist (or monopolist). At the optimum, his marginal payment from trading each asset exceeds the prices by $\mathcal{M}_2^i \Delta\bar{\theta}_2^i$,

$$\partial V_2^i(\cdot)/\partial \Delta\theta_2^i \equiv A - \alpha\mathcal{V}\left(\bar{\theta}_1^i + \Delta\bar{\theta}_2^i\right) = p_2 + \mathcal{M}_2^i \Delta\bar{\theta}_2^i. \tag{8}$$

With a positive definite price impact matrix $\mathcal{M}_2^i$, one can solve (8) for the individual noncompetitive asset demand of investor i in period two as a function of price and his price impact

$$\Delta\bar{\theta}_2^i(\cdot) = (\mathcal{M}_2^i + \alpha\mathcal{V})^{-1}(A - \alpha\mathcal{V} \cdot \bar{\theta}_1^i - p_2). \tag{9}$$

Compared to the competitive market, as long as the investor's price impact is positive, his inverse demand becomes steeper and the investor reduces his order – for any given price, he buys or sells less.

Price impacts. We now endogenize price impacts $\bar{\mathcal{M}}$. Let the price impacts of all other traders but i be fixed, and suppose that investor i decides to sell $\Delta\theta_2^i \neq \Delta\bar{\theta}_2^i$, thus offering extra shares above his equilibrium trade. For markets to clear, the price must adjust, so that the other investors are willing to purchase the additional shares at the new price

$$\Delta\theta_2^i + \sum_{j \neq i} \Delta\bar{\theta}_2^j(p_2, \mathcal{M}_2^j) = 0. \tag{10}$$

Along with the derived noncompetitive demands of other traders (9), the market clearing condition (10) determines the price that assures market clearing for any possible deviation $\Delta\theta_2^i$. Thus, market clearing and optimization by the residual market implicitly define the inverse demand function faced by trader i. Substituting demands (9) of traders other than i into (10), solving for prices p_2, and using the equilibrium market clearing yields trader i's residual demand

$$p_{\bar{p}_2, \Delta\bar{\theta}_2^i, \bar{\mathcal{M}}_2^i}(\Delta\theta^i) = \bar{p}_2 + \underbrace{\left(\sum_{j \neq i} (\mathcal{M}_2^j + \alpha\mathcal{V})^{-1}\right)^{-1}}_{\bar{\mathcal{M}}_t^i} (\Delta\theta_2^i - \Delta\bar{\theta}_2^i). \tag{11}$$

The equilibrium price impact of investor i, as a function of price impacts of other traders, is given by

$$\begin{aligned} \bar{\mathcal{M}}_2^i &= \left(\sum_{j \neq i} (\bar{\mathcal{M}}_2^j + \alpha\mathcal{V})^{-1}\right)^{-1} = \\ &= (1-\gamma)\mathcal{H}(\bar{\mathcal{M}}_2^j + \alpha\mathcal{V} | j \neq i), \end{aligned} \tag{12}$$

[12]The condition equalizes the marginal rate of substitution between the risky assets and bonds with the ratio of their marginal revenues. The marginal utility and the marginal revenue of the riskless asset are both equal to one.

where $\mathcal{H}(\cdot)$ is the harmonic mean operator.[13]

Equation (12) reveals several of the model's central predictions. It first uncovers an interesting mathematical structure about equilibrium market power. Namely, the price impact of investor i is characterized as a harmonic average of the convexities of other investors' value functions, $\alpha\mathcal{V}$, augmented by their price impact, $\bar{\mathcal{M}}_2^j$, and discounted by a factor of $1-\gamma$. To better understand why (12) holds, recall that price impact represents price concessions that are sufficient for the asset markets to clear following unilateral deviations of trader i from his equilibrium trade. The greater the price impact assumed by other traders, the larger the price concession needed to make them willing to absorb the additional risky assets. This reasoning explains how market power mutually reinforces among investors. That the reinforcement is governed by a harmonic (and not, say, arithmetic) average implies that traders with flatter individual demands (with smaller $\bar{\mathcal{M}}^j + \alpha\mathcal{V}$) are relatively more important in determining i's market power. Indeed, these traders will require lower price concessions to purchase a given amount of the risky assets.

The relation between investor's i price impacts in equation (12) takes the impacts of traders other than i as given. To pin down the endogenous price impact for each trader, we find a symmetric solution to the system of consistent matrices described by I nonlinear equations (12), one for each trader. In the unique symmetric solution ($\bar{\mathcal{M}}_2^i = \bar{\mathcal{M}}_2^j$ for all $i, j \in I$), the consistent price impact for any $i \in I$ is equal to

$$\bar{\mathcal{M}}_2^i = \frac{1-\gamma}{\gamma}\alpha\mathcal{V}. \tag{13}$$

The essential determinant of an investor's i price impact is the concavity of the preferences (value functions) of his trading partners, which in static TM-CAPM is determined by bidders' risk aversion and payoff volatility. Given the finite number of traders ($\gamma < 1$) and strictly decreasing marginal utility ($\alpha > 0$), the traders are affected by absorbing the buys or sells of other investors. In addition, the shares are risky, and buying them increases the overall level of riskiness of ultimate holdings, amplified or weakened by cross-market impact, as specified by the variance-covariance matrix $\mathcal{V}$. It follows that the price concessions are required, and the equilibrium price impacts are non-negative. What is less apparent in a symmetric solution, TM-CAPM predicts that it is other investors' risk aversion that enters directly into trader's i price impact, whereas i's own risk aversion enters only by the mutual reinforcement mechanism; more risk averse trading partners are more reluctant to increase their holdings of risky assets, which implies larger price concessions in trading. As expected, price impact is partially mitigated by the number of potential liquidity providers, captured by γ.[14] The predictions for the last trading period approach the competitive outcome when investors are approximately risk neutral ($\alpha \sim 0$) or when the number of traders is

[13] Recall that the harmonic mean of K numbers x_k, $k = 1, ..., K$, is defined as $\mathcal{H}(x_1, ..., x_K \mid k = 1, ..., K) = K(\sum_{k=1,...,K} x_k^{-1})^{-1}$. In (**??**), the harmonic mean operator is applied to matrices.

[14] It is useful to observe that this effect is absent from Cournot. There, the price impact of one of the large investors is equal to the slope of the demand formed by the competitive traders and is independent of the number of other large investors. In our model, when the number of traders increases, the effect that orders of any given trader have on the average marginal utility becomes weaker because each of the other traders absorbs a smaller fraction of these orders.

large ($\gamma \sim 1$).

Asset Prices. The equilibrium asset prices can be derived from the market clearing condition by substituting demands (9)

$$0 = \frac{1}{I}\sum_{i\in I}\Delta\bar{\theta}^i(\bar{p}_2, \bar{\mathcal{M}}_2^i) = v - \bar{p}_2, \tag{14}$$

where we used the symmetry of $\bar{\mathcal{M}}_2^i + \alpha\mathcal{V}$ and the definition of v, (6). The equilibrium prices in the last trading period coincide with the competitive prices

$$\bar{p}_2 = v \equiv A - \alpha\mathcal{V}\theta^{Av}. \tag{15}$$

It is somewhat surprising that, despite the market power of the traders, the prices are as in the competitive CAPM. Intuitively, with symmetric price impacts, the market power of buyers and sellers is balanced, and the buyers and sellers reduce their demand and supply for each asset by the same factor γ. Consequently, thin markets clear at the competitive prices, even though the trades are not competitive. This result, however, relies on the joint assumption of (i) quadratic and (ii) homogenous utility functions, and (iii) the deterministic structure of the model. Relaxing one of the three assumptions would introduce price effects.

Optimal Portfolios. In a perfectly competitive CAPM with symmetric bidders, in equilibrium each bidder sells $\bar{\theta}_1^i$ and replaces it with the average portfolio θ^{Av}. To see how investors rebalance their portfolios in thin markets, we substitute (13) and (15) into (9)

$$\Delta\bar{\theta}_2^i = \gamma(\theta^{Av} - \bar{\theta}_1^i). \tag{16}$$

Investor i sells a fraction γ of a portfolio with which he entered the trading round and replaces it with γ of the average inventory. Interestingly, the fraction is determined solely by the market participation rate γ; in particular, it is independent of risk aversion α. Given the equilibrium trades of risky assets (16), the changes of bond holdings at $t = 2$ is such that trader's budget is balanced

$$\Delta\theta_{b,2}^i = -v \cdot \Delta\bar{\theta}_2^i = -\gamma v \cdot (\theta^{Av} - \bar{\theta}_1^i). \tag{17}$$

In summary, when markets are thin, the prices in the last period will be equal to a fundamental value, while the trade of each investor will be reduced by a factor of γ relative to the infinitely deep market.

3.2 Equilibrium in the First Period $(t = 1)$

Given the optimal trading strategy at $t = 2$, (16) and (17), the ultimate holdings of risky assets as a function of the first-period trade $\Delta\theta_1^i$ become

$$\bar{\theta}_2^i = (1-\gamma)\bar{\theta}_0^i + (1-\gamma)\Delta\theta_1^i + \gamma\theta^{Av}. \tag{18}$$

In response to their price impact, in the last trading period, investors rebalance less than they would in competitive markets. Therefore, unlike in a competitive model, the ultimate portfolio depends on round-one trade $\Delta\theta_1^i$, because increasing risky holdings at $t = 1$ exposes the trader to idiosyncratic risk at maturity. As we now show, the less than full diversification at $t = 2$ has direct consequences for the market depth at $t = 1$.

Substituting (18) into (1) gives the indirect utility function of trader i at $t = 1$, as a function of trade in this period. The presence of period-one trade $\Delta\theta_1^i$ in the final portfolio implies that the value function has a quadratic term and is strictly concave and that the investor is risk averse at $t = 1$ as well. The coefficient of effective risk aversion, $(1-\gamma)^2\alpha$, depends on the market participation rate γ, which happens, because only $1-\gamma$ of the first-period trade survives till maturity, while the remaining fraction γ is liquidated at $t = 2$. Higher competitiveness of the market in the last period improves the hedging possibility in that period, which weakens the impact of trade $\Delta\theta_1^i$ on the final holdings $\bar{\theta}_2^i$.

The equalization of marginal utility and marginal payment determines the optimal trading strategy for i, given any $\mathcal{M}_1^i$

$$\partial V_1^i(\cdot)/\partial\Delta\theta_1^i \equiv v - (1-\gamma)^2\alpha\mathcal{V}\left(\bar{\theta}_0^i + \Delta\bar{\theta}_1^i - \theta^{Av}\right) = p_1 + \mathcal{M}_1^i\Delta\bar{\theta}_1^i, \tag{19}$$

which yields the first-period trade as a function of price and price impact

$$\Delta\bar{\theta}_1^i(\cdot) = \left(\mathcal{M}_1^i + (1-\gamma)^2\alpha\mathcal{V}\right)^{-1}\left(v - p_1 - (1-\gamma)^2\alpha\mathcal{V}\left(\bar{\theta}_0^i - \theta^{Av}\right)\right). \tag{20}$$

Following the steps from the previous section, one can derive the residual supply for trader i to find its slope $\mathcal{M}_1^i = (1-\gamma)\mathcal{H}(\mathcal{M}_1^j + (1-\gamma)^2\alpha\mathcal{V}|j \neq i)$. The system of I such conditions has a unique symmetric solution

$$\bar{\mathcal{M}}_1^i = \frac{(1-\gamma)^3}{\gamma}\alpha\mathcal{V}. \tag{21}$$

The possibility of re-trade at $t = 2$ lowers the price concessions required at $t = 1$, and the first-period price impact is smaller than in the second period by a factor of $(1-\gamma)^2$.

The equilibrium price is equal to the fundamental value $\bar{p}_1 = v$ and the equilibrium trade

$$\Delta\bar{\theta}_1^i = \gamma(\theta^{Av} - \bar{\theta}_0^i). \tag{22}$$

As at $t = 2$, in the first period, the trader sells γ of its current portfolio and acquires γ of the average portfolio.

To recap, in the model with two trading periods, traders' price impact is positive in both periods, and is smaller in the earlier trading round. It is key to observe why. The last-period price impact results from the risk aversion of liquidity providers, which makes them willing to absorb the risky assets of other traders only at price concessions. By rendering full diversification suboptimal at $t = 1$, it is the thinness of the second-period market that makes the bidders averse to absorbing risky shares in the first period, which induces price impact at $t = 1$ as well. On the other hand, having a trading opportunity in the future, the liquidity providers are more willing to absorb risky assets from other traders, knowing that they will be able to partially diversify their positions in the next period. The required price concessions at $t = 1$ are smaller and so is the price impact.

It is worth reiterating one important lesson from the two-period model: whether or not the market is thin at $t = 1$, depends on the thinness of the market at $t = 2$. In the extreme case, if we departed from our model by assuming that the traders are price takers at $t = 2$, then the risky holdings at maturity would not depend on the first-period trade $\Delta\theta_1^i$ and would be equal to θ^{Av}. The investors would become effectively risk neutral at $t = 1$ and the consistent price impacts would be zero, implying competitive markets in this period. In the next section, we show that the fact that *all* liquidity providers have price impact is essential to studying dynamic thin markets.

4 Dynamic Origins of Market Thinness

TM-CAPM reveals an intertemporal mechanism that governs market (non)competitiveness – future market thinness begets present market thinness. This section demonstrates that the presence of even one price-taking trader breaks this link and makes the early-period market perfectly competitive. As a result, the outcome of a dynamic equilibrium is competitive. We also argue that these implications hold not only in TM-CAPM but also in a large class of perfect-foresight models that are based on a uniform price mechanism. The argument we present, thus, further motivates our modeling choice of allowing all traders to have price impact and elucidates the origins of market thinness.

Let us depart from our model by assuming that one of the traders is a price-taker. By definition, the price impact of this trader (j) is equal to zero in both periods, $\mathcal{M}_1^j = \mathcal{M}_2^j = 0$. We maintain the assumption that all the other traders have the correct assessments of their price impact – that is, these assessments are also consistent with the competitive belief of trader j. Let $\bar{p}_2$ denote an equilibrium price in period $t = 2$. In that last period, the competitive investor j will trade to arrive at the portfolio that equalizes his marginal utility and $\bar{p}_2$. Crucially, his ultimate holdings of risky assets at the end of $t = 2$ depend only on the price $\bar{p}_2$ and are independent of his trade of risky assets at $t = 1$. Since the trader becomes insensitive to the riskiness of assets traded in period one, despite $\alpha > 0$, his value function at $t = 1$ becomes linear in trade $\Delta\theta_1^j$ with the coefficient of $\bar{p}_2$,

$$V_1^j(\Delta\theta_1^j, \Delta\theta_{b,1}^j) = \Delta\bar{\theta}_{b,1}^j + \bar{p}_2 \cdot \Delta\theta_1^j + c, \tag{23}$$

where c is a constant that does not depend on the trades in period one. (With perfect foresight, an increase of $\Delta\theta_1^j$ results in the change of the holdings of bonds by $\bar{p}_2 \cdot \Delta\theta_1^j$ and it does not affect risky holdings $\bar{\theta}_2^j$.) Intuitively, since, on the optimal path, any increase in the demand for risky assets at $t = 1$ would cause a one-to-one increase in the sales of risky assets at $t = 2$, the first-period trade does not affect the riskiness of the ultimate holdings and the trader behaves as if he were risk neutral in the first period.

It follows from j's first-order condition that his constant marginal utility (in terms of $\Delta\theta_1^j$), combined with his competitive belief at $t = 1$ ($\mathcal{M}_1^j = 0$), makes his demand at $t = 1$ perfectly elastic. In the first round, investor j is, thus, willing to absorb arbitrary trades from other investors without any price concession from $\bar{p}_2$. Consequently, the other traders' consistent price impacts are equal to zero and all traders behave competitively at $t = 1$. The overall outcome in the two-period model is competitive – the allocation is efficient at the average portfolio θ^{Av} for all traders, prices are equal to v in both periods, and there is no trade in the second round. This happens even though, at $t = 2$, the market is *not* competitive as all traders' actual price impacts are strictly positive.

The argument that the demand of a price-taking trader is perfectly elastic at $t = 1$ does not rely on the consistency of price impacts of other investors. While in the context of TM-CAPM, the demonstration of the implied competitiveness of the outcome does appeal to consistency, we now argue that the conclusion extends to dynamic Cournot-type models. That is, in such models, the ultimate portfolios are fully diversified (i.e. Pareto efficient) and prices are competitive throughout all trading periods; the residual supply faced by each strategic trader is perfectly elastic in all but the last period. We close this section by formalizing this result for a Cournot model with $T \geq 2$ trading rounds.

Consider a model as in Section 2, in which traders are divided into two groups: (1) non-strategic, price-taking traders ($\bar{\mathcal{M}}^i \equiv 0$) who submit downward-sloping demands (provide liquidity), which then define the market demand for strategic players; and (2) strategic players ($\bar{\mathcal{M}}^j > 0$) whose strategies are quantities $\Delta\theta_t^i$, i.e., market orders, which do not depend on prices. We continue to allow all traders to place arbitrarily large orders. Such a model has the microstructure of the standard Cournot model, applied to financial markets. For a competitive trader, let $\Delta\bar{\theta}_t^i$ denote his equilibrium demand evaluated at the equilibrium price and, for a strategic trader, let it denote his equilibrium trade. Proposition 1 characterizes subgame-perfect Nash equilibria in a model with $T \geq 2$ trading rounds.

Proposition 1 (Dynamically Competitive Markets) *In any subgame-perfect Nash equilibrium, in every period $t = 1, ..., T$, prices are equal to v, and trades $\left\{\Delta\bar{\theta}_t^i\right\}_{i,t}$ are such that $\sum_i \Delta\bar{\theta}_t^i = 0$ and $\bar{\theta}_T^i = \theta^{Av}$.*

The result is striking. It demonstrates that even one competitive rational trader who ignores its impact on prices makes the markets perfectly competitive in all but the last period. The result is robust to modeling assumptions as it can be extended to a large class of noncompetitive models based on a uniform price, such as uniform-price auctions, monopoly, etc. Because the mechanism through which the market outcome becomes competitive arises in dynamic trading, we call such models *dynamically competitive*. The key difference between TM-CAPM and dynamically competitive models is that, in those models, there is at least one competitive trader capable of providing liquidity to the market by placing arbitrarily large limit orders, whereas in TM-CAPM, the liquidity is provided by institutional traders who recognize their impact on prices. (Competitive traders who submit market orders can be easily incorporated in TM-CAPM without affecting the dynamic thinness of markets.)

We conclude that, in order to model perfect-foresight dynamic markets that are thin, one needs to allow all traders who provide liquidity – that is, traders who make choices contingent on prices and define residual demands for other traders – recognize their price impact. This is the approach taken in this paper.

5 Model Predictions

This section describes the model's predictions about trading strategies, prices and price impacts in dynamic thin markets. We show that the equilibrium properties of prices and trades in bilaterally oligopolistic markets are qualitatively different from Cournot and other dynamically competitive markets (in the sense of Proposition 1).

Optimal trading strategies. The first result characterizes how market thinness affects the optimal execution of trade. The competitive CAPM predicts that investors instantaneously sell their initial inventories and rebalance their holdings within one period; and that they invest in a combination of the market portfolio and the riskless asset (Two-Fund Separation Theorem). In a deterministic setting, no trade takes place in subsequent periods, as the investors' risky holdings become efficient already in the first trading period. By contrast, TM-CAPM predicts that the optimal handling of large orders in thin markets ($\gamma < 1$) involves trading in blocks. The adverse effects of price impact induce investors to break up their orders into smaller blocks and place them on the market sequentially.

In addition, order break-up takes a particular, easy-to-execute form that results in a *Three-Fund Separation*. Namely, every time they trade, investors sell a constant fraction of (the remaining part of) their initial portfolios to invest between the average inventory and the riskless asset.

Proposition 2 (Three-Fund Separation) *For every trading period $t = 1, ..., T$, the risky part of the optimal portfolio is a convex combination of the initial and average inventories, θ_0^i and θ^{Av}, and the weight assigned to θ^{Av} is monotonically increasing over time:*

$$\bar{\theta}_t^i = (1-\gamma)^t \theta_0^i + (1-(1-\gamma)^t)\theta^{Av}. \tag{24}$$

The remaining wealth is invested in the riskless asset, $\theta^i_{b,t}$.

Order break-up (Fact 1) is a common practice among large investors. Table 1 reports a summary of findings by Chan and Lakonishok (1995), who examined the time structure of the institutional orders placed for 29 months on the NYSE. Only 20% of the total volume of all institutional purchases and sales is completed within one day, and more than 30% of the orders takes at least 6 days to execute.

TABLE 1. ORDER BREAK-UP

	1 Day	2-3 Days	4-6 Days	> 6 Days
Buy	20.1%	26.7%	21.7%	31.5%
Sell	22.1%	27.2%	20.5%	30.2%

(25)

Data: All trades of NYSE and AMEX stocks by 37 investment management firms from July 1, 1986, to December 30, 1988 (October 1987 excluded). A buy/sell package is defined as successive purchases/sales of a stock with at most a 5-day break between consecutive trades. The numbers are percentages of the total volume of trade in $. (Chan and Lakonishok [1995, Table 1])

PARTIAL DIVERSIFICATION OF RISK. TM-CAPM further predicts that when the number of trading periods is sufficiently large, the portfolios will converge to the competitive holdings, but for any finite number of periods, the portfolios will be distinct. Consequently, at any point in time, the idiosyncratic risk will not be perfectly hedged. On the other hand, the allocation can be arbitrarily close to efficiency, provided that the time to maturity $(T - t)$ is sufficiently long.

When markets are deeper, individual risky holdings converge more quickly to the competitive outcome, i.e., the average inventories held by all investors (see Figure 1). Somewhat surprisingly, not only does γ affect the speed of trade, but it actually fully determines it. In particular, the speed of trade does *not* depend on risk aversion α, as long as $\alpha > 0$. Intuitively, higher α is associated with greater gains to trade, and hence encourages more aggressive hedging through faster trading. It also, however, amplifies the price impacts of all traders, making the interactions less competitive and reducing the trade. In a quadratic symmetric model, the two effects of risk aversion offset each other. Thus, even if large institutional traders are almost risk-neutral, as is sometimes assumed in the finance literature, they will choose to trade slowly.

DELAY OF TRADE WITH (DESPITE) COMPLETE INFORMATION. Why do investors trade after the first period despite the absence of shocks or information disclosure? As explained in Section 3.2, market thinness in the last trading period T implies that the markets in $T-1$ are thin as well. By a recursive argument, the thinness of markets in the last trading period implies market thinness in all earlier periods, including round one. In response to the noncompetitiveness of the first-period market, investors reduce their orders and the gains to trade are not exhausted in the first trading round (nor are they in any subsequent period), which leads to re-trade.

Our results raise the question 'What does one "trading period" stand for in the model and what does it represent empirically?' This question does not arise in competitive models without discounting, since there, with no shocks and information disclosure, even one trading round suffices to exhaust all gains to trade. This is not the case in thin markets where the number of trading opportunities does affect the final allocation of assets. In a thin market, a trading period represents the time needed for a price to be formed and trades to take place at *this* price. Each additional trading round allows for better diversification of portfolios and, hence, market thinness gives rise to an alternative-to-discounting meaning of "time is money" in the following sense. The investors strictly prefer to trade (rather than being indifferent or preferring to wait) whenever there is a trading opportunity. If there is only one trading period, the price-taking agents would rebalance their entire initial holdings, while the noncompetitive investors diversify γ of these holdings. As demonstrated in Proposition 1, in dynamically competitive markets, the equilibrium trade is not pinned down in the absence of discounting. The optimal execution of orders is uniquely determined in TM-CAPM even without introducing any preference for urgency. In Section 6.3, we suggest a test to empirically identify the length of time that corresponds to a trading period.

Non-stationarity of Price Impact. The noncompetitive CAPM predicts that the price impact is not constant across the trading periods, but instead increases as the time approaches maturity – the further from maturity, the more opportunities to diversify and re-trade, the less costly it is for the investors to depart from their current holdings, and the smaller the price concessions required. That mechanism is apparent in the value function (equation (41) in Appendix I), which becomes more and more concave over time to reflect the traders' increasing effective risk aversion, and decreasing willingness to buy risky assets at given price concessions; this generalizes the intuition from the two-period model in Section 3.2. To the extent that market depth can proxy the level of market competitiveness, TM-CAPM predicts that markets are least competitive just prior to maturity.

Proposition 3 (Time Structure of Price Impact) *In TM-CAPM, the price impact exponentially decreases with time-to-maturity $T-t$,*

$$\bar{\mathcal{M}}_t^i = \frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\alpha\mathcal{V}. \tag{26}$$

Remarkably, the derived schedule of price impact matrices, $\bar{\mathcal{M}}_t^i$, is directly proportional to the variance-covariance matrix of returns, $\mathcal{V}$. Thus, the markets for less risky assets are deeper. Moreover, formula (26) shows that there are cross-market price-impact effects. When the payoffs of two stocks are positively correlated, the sale of one asset inflicts a downward pressure on the price of other assets. In addition, time-to-maturity, $T-t$, and market participation rate, γ, weaken the effect of asset riskiness, $\mathcal{V}$, and risk aversion, α, which are the determinants of concavity of the investors' preferences in TM-CAPM, on price impact.

One implication of the nonstationarity of price impact is that liquidity is correlated across assets even if their returns are independent.[15] That mechanism, dubbed 'commonality in liquidity', has already been widely documented in the empirical literature (e.g. Hasbrouck and Seppi (2001); Huberman and Halka (2001); Chordia, Roll, and Subrahmanyam (2002); the survey by Amihud, Mendelson and Pedersen [2005]).[16] The non-stationarity of the frequent trading model is not surprising *per se* in a model with a finite number of trading periods and varying strength of the end-time effect across trading rounds. That the market depth is nonstationary has, however, striking implications for price behavior when the model is extended to include exogenous supply or demand shocks, which we discuss in Section 6.3.

The derived monotonic time structure of price impact suggests that market depth can endogenously become higher prior to news announcements.[17] Similarly, Proposition 3 implies that, in thin markets, asset maturity might become an active instrument in stabilizing markets; the increasing maturity of an asset lowers the price impact in the market for that asset and increases the level of competitiveness in all trading periods.

Security Market Line. One of the most celebrated and controversial results in the standard CAPM is the Security Market Line, which asserts that the return of an asset can be explained solely by the covariance of its return with the return of the market portfolio. Analyzing the tradeoff between risk and return is much harder in the noncompetitive model, as the asset prices no longer coincide with the marginal utilities of traders, and, moreover, marginal utilities typically differ across agents.

Nevertheless, we have shown that, under our assumptions, the equilibrium asset prices derived in (51) coincide with the fundamental values of assets and are, therefore, identical to the competitive prices in every period. It follows that asset returns are the same random variables as in the competitive model and their expectations lie on the Security Market Line, spanned by the riskless

[15] Market thinness can be viewed as a particular source of market illiquidity, with price impact measuring illiquidity. Domowitz, Hansch, and Wang (2005) show empirically that liquidity commonality is due to co-movements in supply and demand that are induced by cross-sectional correlation in order types (market and limit orders), while return commonality is caused by correlation in order flows (order direction and size). Thus, stocks that do not correlate in returns can feature liquidity co-movement because return co-movement and liquidity co-movement are caused by different economic factors. The authors conclude that liquidity co-movement does indeed pose a challenge for traditional diversification strategies that are based solely on return interactions.

[16] It might seem that, apart from predicting commonality in liquidity, Propositions 2 and 3 imply that the volume of trade and liquidity (measured by the inverse of price impact) should be positively correlated in time series data. Yet, although the empirical support for the cross-sectional relation of liquidity across markets is strong, the evidence on the dynamic relation is mixed (see, e.g., Johnson [2007]). Notice, however, that TM-CAPM predicts that, in any given trading period, investors will rebalance a fraction γ of the remaining part of their initial holdings, and this fraction depends solely on the participation rate and not on the price impact or time. That the absolute volume of trade appears correlated with price impact is an artefact of the perfect correlation of gains to trade across traders in the first trading period. Although we have not solved for that, we believe that, with endowment shocks, the correlation would break down.

[17] For that prediction, let us reinterpret maturity as follows. Recall that T is defined as the period at the end of which assets mature and the uncertainty about dividends is fully resolved. Our analysis implies that if the disclosure of information is introduced to the model, T would alternatively be interpreted as a moment at the end of which some (partial) information about dividends is revealed. Our predictions are consistent with the extensively documented fact that volatility tends to increase prior to scheduled news announcements and decrease on the announcement day as uncertainty is resolved by the market participants.

return and the return on the average inventory. Let R^{Av} denote the expected return of an average inventory, let $\beta_n = \mathcal{V}_{Average,n}/\mathcal{V}_{Average}$ be the beta of asset n, and let R_n be the expected return of asset n.

Proposition 4 (Security Market Line) *In thin markets with I liquidity providers, the expected returns of assets in any period $t = 1, ..., T$ are located on the Security Market Line,*

$$R_n - R = \beta_n (R^{Av} - R). \tag{27}$$

We should stress here that in our model, the average inventory is defined as the *per capita* risky portfolio held by a possibly small group of liquidity providers who are trading in a given asset market. Therefore, the standard approach to testing the predictions of CAPM, based on the whole market portfolio, should not be applied in this instance. To empirically test TM-CAPM, one should first properly identify a thin market.

Proposition 4 implies that thin trading *per se* does not necessarily distort asset returns. It suggests that in order to explain liquidity premia within CAPM, one needs other distortions (on that, see Section 9).

6 Price Effects in Thin Markets

TM-CAPM suggests that, in the absence of shocks and with sufficiently long trading horizon, thin markets feature an essentially competitive outcome – almost perfectly diversified portfolios and competitive prices throughout. As we demonstrate in this section, however, thin markets respond very differently than dynamically competitive markets to exogenous shocks in asset supply/demand, or when investors need to quickly liquidate their portfolios. We show that taking into account market thinness may explain a number of empirical phenomena that are hard to reconcile within the dynamically competitive models; these include asset price overshooting (Facts 2 and 4, Section 6.1); the existence of instruments for asset valuation that account for price impact (Section 6.2); and stylized facts about return volatility (Section 6.3). Examples of supply shocks in financial markets include forced liquidation, issuance of new debt, or selling Initial Public Offerings (IPOs). Other exogenous demand or supply shocks that have been extensively studied are an inclusion of an asset into the S&P index or a change of index weights – index funds invest a constant fraction of wealth in companies that are included in an index, regardless of the performance of an asset; therefore, a change of index weights induces a demand shock that is not associated with the new information about the fundamental value of an asset. Alternatively, exogenous shocks could capture a net trade of small competitive traders who do not monitor prices and are unable to take advantage of price differentials.

To examine how thin markets react to exogenous shocks in asset supply (or demand), we enrich the model with unanticipated as well as an anticipated exogenous sale of a large block of shares by a trader other than I liquidity providers.

6.1 Fundamental and Liquidity Effects

As evidenced by the empirical literature, the exogenous shocks in asset supply result in price overshooting. Even if the shock is preannounced, on the actual event date, the price drops below the new fundamental value to attain that value only in subsequent periods. This phenomenon, often referred to as *price overshooting*, should not occur in (dynamically) competitive markets. Otherwise, price-taking investors could make infinite profits by placing an unbounded buying order when the price is depressed and a selling order after the price reverts deterministically. This is why, in dynamically competitive models, the presence of even one price-taking trader prevents overshooting from occurring on the equilibrium path. What should be observed, instead, is that the prices adjust to the new fundamental value immediately following the shock or its announcement and remains there till maturity. Our model reveals that, in thin markets, the price change resulting from any exogenous supply or demand shock has a *fundamental* component, which is permanent, and a *liquidity* component, which is temporary. Proposition 5 demonstrates that price overshooting is a direct consequence of these two effects. As we proceed to explain, the two effects differ not only in their origin and persistence, but also in timing and magnitude dynamics.

Consider an unanticipated one-time shock in asset supply. In period t^*, a portfolio $\hat{\theta}$ is being liquidated along with the trade by liquidity providers. Since all liquidity providers learn about the shock in period t^*, the average inventory attains a new post-shock level $\theta^{Av*} \equiv \theta^{Av} + \hat{\theta}/I$, and the post-shock fundamental value changes to $v^* \equiv A - \alpha\mathcal{V}\theta^{Av*}$. The change in the fundamental value represents the fundamental effect

$$\Delta^F \equiv -\alpha\mathcal{V}\frac{\hat{\theta}}{I}, \tag{28}$$

which results from the permanent change in the average holdings of risky assets by liquidity providers. Given the investors' risk aversion (decreasing marginal utility), the supply shock $\hat{\theta} > 0$ lowers the average marginal valuation.

The fundamental effect would also be observed in a model with price-taking liquidity providers, as long as the number of such providers remained small (so that the *per-capita* shock $\hat{\theta}/I$ is not negligible). It is the liquidity effect, which lowers the price at t^* further below v^*, that is due to the noncompetitive nature of trade. Why do the traders demand price concessions beyond the drop in the fundamental value? On the equilibrium path (with or without the shock), each trader equalizes his marginal utility and his marginal payment, which implies that the average marginal payment always coincides with the fundamental value. Without the shock, the marginal payment is equal to the equilibrium price. With a positive net supply of risky assets at t^*, $\hat{\theta}$, investors demand, on average, positive amounts of shares, and investors' average marginal payment exceeds the market price by $\bar{\mathcal{M}}^i_{t^*}\hat{\theta}/I > 0$. It then follows from optimality that the price is below the average marginal utility. For the sake of intuition, suppose that the investors' portfolios are fully diversified on the day of the shock; then, as a residual monopsonist, each investor is willing to absorb $\hat{\theta}/I$ only at a price lower than his marginal utility v^*.[18] The liquidity effect is equal to

$$\Delta^L \equiv -\frac{(1-\gamma)^{2(T-t^*)+1}}{\gamma}\alpha\mathcal{V}\frac{\hat{\theta}}{I}. \tag{29}$$

Proposition 5 describes the price behavior on the equilibrium path following a supply shock.

Proposition 5 (Asset Price Overshooting) *Following an unanticipated liquidity shock, $\hat{\theta}$, in period t^*, the prices adjust by $\Delta^F + \Delta^L$. In period $t^* + 1$, the prices revert by Δ^L to the post-shock fundamental value v^* and remain at this level in all following periods.*

Thus, TM-CAPM predicts overshooting as the equilibrium reaction of prices to unanticipated shocks in asset supply. (Section 7 extends the result to anticipated shocks.) The overshooting price pattern described in Proposition 5 was first empirically documented by Kraus and Stoll (1972) and subsequently confirmed by many studies (e.g., Harris and Gurel [1986]; Holthausen, Leftwich and Mayers [1990]; Chan and Lakonishok [1995]; Beneish and Whaley [1996]; Keim and Madhavan [1996]; Lynch and Mendenhall [1997]; and Greenwood [2005]). The name "overshooting" has become popular more recently. Earlier studies referred to "long-" versus "short-run" or "permanent" and "temporary" effects of shocks on prices, which correspond to the fundamental and liquidity effects in TM-CAPM.

Why does the liquidity effect not persist like the fundamental effect? Since no other shocks occur after t^*, in all periods following t^*, the average trade of the liquidity providers becomes equal to zero (by market clearing) and the price is equal to the new fundamental value $\bar{p}_t = v^*$. Observe that the fundamental effect Δ^F does not depend on the timing of the shock t^*; due to the nonstationarity of price impact, the magnitude of overshooting Δ^L is greater when the time to maturity is shorter. These predictions are strongly supported by the methodology recently implemented by Citigroup to estimate price impact.[19] Figure 2 depicts both effects for the exogenous shock $\hat{\theta}$ in period t^*. Panel A shows the path for the trade of an asset, and panel B, the price of an asset $\bar{p}_t$.

The influential paper by Brunnermeier and Pedersen (2005) explained price overshooting by "predatory trading". When a trader needs to quickly liquidate a portfolio, other investors sell and subsequently buy back the asset. This strategy lowers the price at which they can obtain the liquidated portfolio. The mechanism arises due to the presence of long-run investors who define a downward-sloping demand, buying assets when they are expensive and selling when assets are cheap. Our explanation of overshooting is complementary in that predatory trading does not occur in our model, since all traders maximize their preferences. In addition, while predatory trading can

[18] When portfolios are fully diversified, this can easily be seen by taking the average of equation (44) across all traders and observing that the left-hand side (the average marginal utility) is equal to v_{t^*} in all periods $t \geq t^*$, while the right-hand side (the average marginal revenue) is equal to $p_{t^*} + \bar{\mathcal{M}}^i_{t^*}\hat{\theta}/I$ in t^* and p_t from $t^* + 1$ on.

[19] In that program, the price impact is decomposed into the following: (a) a permanent component ("reflects the information transmitted to the market by the buy/sell imbalance"), which is believed to be roughly independent of the trade scheduling; and (b) a temporary component ("reflects the price concession needed to attract counterparts within a specified short time interval"), which is highly sensitive to trade scheduling (Almgren et al. [2005]).

rationalize price overshooting as a response to unanticipated shocks, our model can also explain delayed overshooting.

When the asset returns are correlated, overshooting in one market spills over to other markets. Apart from the permanent adjustment of the fundamental value in substitute or complement asset markets,

$$\theta \cdot \Delta^F = \frac{\alpha}{I} Cov(A\hat{\theta}, A\theta), \tag{30}$$

an exogenous sale of one asset induces liquidity effects in these markets at t^*

$$\theta \cdot \Delta^L = \frac{(1-\gamma)^{2(T-t^*)+1}}{\gamma} \frac{\alpha}{I} Cov(A\hat{\theta}, A\theta). \tag{31}$$

LIMITS TO ARBITRAGE. The overshooting result from Proposition 5 may seem puzzling at first. Since the liquidity providers know at t^* that the prices will revert in the next period, why do they not arbitrage the temporary price differential between t^* and t^*+1, like they would in the competitive or Cournot models? In addition, could outside investors benefit from the anticipated price change? In answering these questions, we show that another qualitative change in thin markets involves endogenously arising limits to arbitrage.

A fundamental paradigm of the classical asset pricing models is that shocks can have only negligible effects on asset prices. With price-taking agents, anticipated price differentials create infinite profit opportunities and the flows of speculative capital immediately drive the price back to the fundamental value. By contrast, in thin markets, the potential profits from entering the market are bounded due to price impact, which limits the benefits from arbitrage and reduces the incentives to enter the market. To see that, suppose that, at t^*, an (outside) entrant purchases a block of assets to be sold in the next period. The benefit from such a market operation is bounded for two reasons: first, due to the price impact at t^*, buying a few more shares than $\hat{\theta}$ (let alone taking an unbounded position) would result in a strictly negative profit, as the purchase would drive the price above the fundamental value in this period; secondly, selling the shares in period t^*+1 would also have adverse effects on the price, further magnified by the non-stationarity of price impact. Still, for any overshooting effect, there exists a sufficiently small trade $\left\{\hat{\theta}_t\right\}_{t \geq t^*}$ satisfying $\sum_{t \geq t^*} \hat{\theta}_t = 0$ (i.e., a round-trip trade) that gives a positive profit.[20] Given the boundedness of profits, unlike in a competitive model, even small fixed entry costs may prevent outsiders from arbitraging the price overshooting. These entry costs include explicit trading costs, such as transactions costs, but also the costs associated with learning the characteristics of the stocks. Mitchell, Pedersen and Pulvino (2007) document that it may take months for outside capital to bid prices back to the fundamental value.[21] This slow entry is attributed by the authors to information barriers and the

[20] In this argument, we assume that the liquidity providers do not anticipate the trades by an outsider.

[21] The study examines price behavior in the convertible bond market in 2005 and around the collapse of LTCM in 1998 and merger targets in the 1987 market crash. During these events, natural liquidity providers were themselves forced to liquidate their holdings, which depressed the prices below the fundamental values, despite the fact that

costs of maintaining dormant financial and human capital in a state of readiness when arbitrage opportunities arise.

The argument of entry fees does not apply to liquidity providers. Why, as shown in Proposition 5, do liquidity providers not take advantage of the anticipated price change? Compared to the benefits from arbitraging by outsiders, for any given round-trip placed by an insider, there is an additional effect. By placing a round-trip, liquidity providers would exert a negative (first-order) externality by increasing the cost of their current equilibrium purchases. In a noncompetitive equilibrium, the marginal benefit from inter-temporal arbitrage is exactly offset by the marginal externality cost.[22]

Once one acknowledges that traders who can place large orders will have price impact, limits to arbitrage arise. The argument behind limits to arbitrage with non-price taking behavior differs in two ways from that in the competitive model. First, the externality exerted by arbitrage on other trades introduces a difference in arbitrage possibilities between insiders and outsiders. Secondly, for insiders as well as outsiders, the profits from arbitrage are bounded for any round-trip trade.

Price Manipulation. Given that traders can affect prices, a natural question is whether the investors have any incentive to manipulate the price. More precisely, so far, we have examined whether it would be profitable for an investor to arbitrage the price differential created by an exogenous liquidity shock. We now ask whether any investors would want to destabilize the market and generate such shocks themselves by submitting a sequence of market orders and then take advantage of the resulting price differentials. If each block in the sequence is interpreted by the liquidity providers as a once-and-for-all shock, whereas the manipulator knows the whole sequence of the shock and thus knows the price path, such asymmetry in information might potentially lead to a positive profit from price manipulation. Proposition 6 establishes that, even though it is possible for the investors to affect the price in a thin market, doing so can never be profitable. A *price manipulation* is defined as a non-zero sequence of trades of risky assets $\left\{\hat{\theta}_t\right\}_t$ such that $\sum_t \hat{\theta}_t = 0$ (i.e., a *round trip*) for which $\sum_t \bar{p}_t \cdot \hat{\theta}_t > 0$.

Proposition 6 (Price Manipulation) *For any round-trip trade $\left\{\hat{\theta}_t\right\}_t$, the net profit is negative $\sum_t \bar{p}_t \cdot \hat{\theta}_t < 0$, where $\{\bar{p}_t\}_t$ is the vector of equilibrium prices with an unanticipated sequence of shocks $\left\{\hat{\theta}_t\right\}_t$.*

there was little change in the overall fundamentals. In the convertible bond markets, the prices deviated from the fundamental values, reaching the maximum discount of 2.7% in 2005 (2.5 standard deviations from the historical average), and 4% in 1998 (4 standard deviations from the average). During the crash of 1987, the median merger arbitrage deal spreads increased to 15.1%. In all cases, it took several months for traders to increase their capital or for better-capitalized traders to enter.

[22] Suppose, for the sake of simplicity, that the holdings of risky assets are fully diversified. If in t^*, an investor increases his trade by ε and sells the same amount in the next period, the benefit from arbitraging the price differential is equal to $\varepsilon \times \bar{\mathcal{M}}\hat{\theta}/I$. At the same time, the additional demand increases the price in t^* by $\bar{M} \times \varepsilon$, which adversely affects the terms of trade of $\hat{\theta}/I$. On the equilibrium path, the marginal benefit is exactly equal to the marginal negative externality inflicted on today's trade.

The key feature of the model underlying the robustness of thin markets to price manipulation is time-independence of the fundamental component of price impact. With a time varying fundamental effect price manipulation could be profitable. The price change induced by the liquidity effect lasts only for one period and always works against the manipulator, irrespective of whether he buys or sells.[23]

6.2 Market Value and Blockage Discount

In a perfectly competitive market, the cash value of a block of shares, $\hat{\theta}$, is simply equal to the quantity of shares times the corresponding prices currently observed on the market, $\bar{p} \cdot \hat{\theta}$. When markets are thin, selling a large block of shares exerts a downward pressure on prices, and the market value no longer reflects the actual amount of cash that would be obtained by selling block $\hat{\theta}$. The problem of apprizing assets traded in thin markets has been recognized by valuation specialists, who apply the so called *blockage discount*. The blockage discount is defined as a "deduction from the actively traded price of a stock because the block of stock to be valued is so large relative to the volume of actual sales on the existing market that the block could not be liquidated within a reasonable time without depressing the market price" (*Handbook of Advanced Business Valuation*, p. 140). These are distinct from (though sometimes confused with) *restricted stock discounts* due to the difficulty in selling that is caused by regulatory or contractual constraints. In practice, blockage discounts are applied not only to stocks, but also to real estate, personal property (e.g., collections of art, antiques and manuscripts), charitable gifts, etc. The discounts have typically been estimated to range between 0 and 15 percent. The IRS has acknowledged the concept of blockage discount since 1937.[24] According to Federal Tax Regulations, the burden of demonstrating that a blockage discount is justified lies on the taxpayer. Yet, there is no equilibrium asset pricing model that would provide guidance on how to assess the cash value of assets and the appropriate amount of blockage discounts. Practitioners have developed a range of heuristic methods for how to adjust the values of assets (e.g., Estabrook [1999, 2001]), and these methods have been adopted in appraisal businesses and valuation consulting. Our results can quite directly be applied to formally address asset valuation in thin markets and thus derive blockage discounts.

The challenge in formalizing appraisals when markets are thin arises because assets are often transferred outside of the market, or because the transfer is only hypothetical. For example, a typical instance where blockage discounts are applied involves a transfer of a property in a case of a divorce. It is in the interest of the divorcees to claim a large price impact (and blockage discount), which implies a large tax discount. The relevant question is: what would be the value of the property if it were sold on the market (even though it will *not* be)? This counterfactual

[23] Huberman and Stanzl (2004) also decompose an (exogenous) price impact into the permanent and the temporary component to identify conditions on the two price impact functions under which price manipulation is not feasible. Proposition 6 provides an alternative to their argument; our setting is less general in that the permanent price impact is constant over time and is more general in that it allows for an arbitrary number of assets.

[24] Estabrook (1999, 2001) and Pratt (2001) provide a summary of U.S. Tax Court decisions involving blockage discounts.

reasoning corresponds to how the price impact is modeled in TM-CAPM. Let $\bar{p}_t$ be the observed market price and $\hat{p}_t$ be the hypothetical price that would be obtained if the block were offered on the market. The blockage discount is equal to $BD \equiv \hat{\theta} \cdot (\bar{p}_t - \hat{p}_t) = -\hat{\theta} \cdot \Delta\bar{p}_t$, where $\Delta\bar{p}_t$ is as in Proposition 5. Consequently, the blockage discount becomes

$$BD = \underbrace{\frac{\alpha}{I} Var(A \cdot \hat{\theta})}_{\hat{\theta} \cdot \Delta^F} + \underbrace{\frac{(1-\gamma)^{2(T-t^*)+1}}{\gamma} \frac{\alpha}{I} Var(A \cdot \hat{\theta})}_{\hat{\theta} \cdot \Delta^L}. \tag{32}$$

In the derivation of formula (32), we made two implicit assumptions: that the block is being sold all at once, and that the owner does not have any other assets but the considered block. In practice, traders break up large packages into smaller blocks and sell them gradually over time to mitigate the adverse effects that result from market thinness. Therefore, formula (32) is likely to overestimate the value of a blockage discount, and it should be interpreted as an upper bound on the discount. The lower bound for the blockage discount is the fundamental effect, as that effect is present even if the trade is spread over time. If a trader has other assets which are not included in $\hat{\theta}$ and whose payoffs are positively (negatively) correlated with the liquidated portfolio, then the liquidation also affects the values of these assets. The blockage discount should then be adjusted upwards (downwards) accordingly, applying (30).

6.3 Price Volatility

The following robust findings about price volatility have long been documented in empirical finance: the magnitude of price volatility is not justified by the volatility of asset fundamentals; changes in volatility are largely unrelated to changes in fundamentals; volatility exhibits persistence, i.e., clustering of large and small moves; and unconditional distribution of asset returns has heavy tails. Section 6.1 showed that, in thin markets, apart from changing the fundamental value, any exogenous shock has an additional liquidity effect on the equilibrium price. This section demonstrates that the existence of the liquidity effect, combined with the equilibrium behavior of price impact, allow us to shed light on the empirical findings about price volatility. Excess volatility is a direct consequence of the existence of the liquidity effect itself.

To study volatility, suppose that instead of a once-and-for-all shock $\hat{\theta}$, we observe a sequence of i.i.d. mean-zero shocks in asset supply $\hat{\theta}_t$ (generated, for example, by noise traders), and that the shocks have not been anticipated by the traders.[25] As explained in Section 6.1, by increasing the inventories, positive realizations of $\hat{\theta}_t$ permanently affect traders' average marginal utility (the fundamental value). This fundamental effect makes the fundamental value follow a random walk. The conditional variance of the fundamental value is constant and is a linear transformation of the variance of a shock. The proportionality coefficient is monotonically decreasing in γ and is equal to

[25] This argument is only heuristic, as a proper formulation of the problem would require that agents take the randomness of the supply into account when trading. The extra risk would result in a more concave value function (for a model with uncertainty about future prices, see, e.g., Vayanos [1999]). Still, this would not change qualitatively the effects of supply shocks on prices.

zero with $\gamma = 1$; with infinitely many traders, the *per capita* shock is negligible and so is its impact on the fundamental value. These properties hold whether or not traders are price-takers.

The volatility of prices depends on two independent factors: market participation rate and whether or not traders take into account their own impact on prices. In a market with price-taking liquidity providers, the equilibrium price coincides with the fundamental value, the conditional price change $\Delta\bar{p}_t$ is i.i.d. and the price follows a random walk. By contrast, in a thin market, the additional liquidity effect of shocks increases the volatility of prices above of the volatility of the fundamental value. Excess volatility is thus a direct consequence of the investors' price impact in every trading period. In addition, the overall price volatility is further magnified by cross-market effects: as long as the asset payoffs are not independent, the price impact matrix is not diagonal and shocks in one market generate liquidity effects also in other asset markets. Since the liquidity effect of any shock lasts for only one period, in a thin market, the price is a sum of two stochastic processes, a random walk and a mean-reverting process. Next, we offer a heuristic argument to illustrate how the remaining three empirical facts about price volatility can be explained by the behavior of the liquidity effect governed by the dynamics of the equilibrium price impact in our model of frequent trading.

Because the equilibrium price impact varies over time, price volatility changes independently of changes in fundamentals. If, in addition, the monotonic time structure of the equilibrium price impact is taken into account, the high (low) volatility of price is clustered at the end (in the beginning) of the trading horizon. Finally, due to the endogenous time-dependence of the liquidity effect, the conditional distribution of $\Delta\bar{p}_t$ evolves over time and its variance increases exponentially when time approaches maturity. It follows that the price change in trading periods close to T puts a large mass on realizations that are far from typical variability over the trading horizon. If the distribution of $\Delta\bar{p}_t$ is estimated under the assumption that the price change is i.i.d., such an approach might lead to an empirical distribution that has heavy tails, even if the kurtosis of the distribution in every period is less than 3.

Are thin markets efficient in that they fully reflect all available information and no trader can benefit from trading on information? Typically, the concept of market efficiency of prices is formalized by the martingale property. In TM-CAPM with unanticipated shocks, equilibrium prices are not martingales, not even in the weakest sense – namely, with respect to information about past prices. Nevertheless, as argued in Section 6.1, despite some price changes being anticipated, no liquidity provider can benefit from arbitraging. Therefore, if market efficiency is defined with respect to liquidity providers, who monitor the prices, thin markets are efficient even though the prices are not martingales.

Finally, in TM-CAPM, statistical properties of equilibrium prices are different at different frequencies: unlike the sequences of prices from every trading period, the sequences of prices taken every two or more periods are martingales. This suggests a test to determine the length of a period in TM-CAPM from data: two trading rounds correspond to the shortest time for which prices are martingales.

6.4 The Functional Form of Price Impact

A much-discussed question in empirical finance has been about the shape of price impact functions, defined as the magnitude of price changes as a function of a block size. It is now standard to distinguish between permanent and temporary price impact functions, which are based on the long-run and short-run price changes, respectively. In the data, the permanent impact function is typically estimated as a linear function of blocks (e.g., Almgren et al. [2005]),[26] whereas the temporary price impact appears to be a concave function of the block size, which by now is a robust and well-documented result (e.g., Keim and Madhavan (1996); Kempf and Korn [1999]; Plerou et al. [2002]; Almgren et al. [2005]). Much of the empirical evidence has been established, assuming that price impact is *time-independent*.

Our model of thin markets with unanticipated i.i.d. shock predicts that, in any trading period, the derived permanent and temporary price effects are both linear in the block size (cf. (28) and (29)). The permanent price effect is stationary in TM-CAPM – the permanent price effect depends on the magnitude of the shock realizations and price volatility, but shocks per share are constant across all trading periods. If price impact indeed evolves over the trading horizon, as predicted by TM-CAPM, and large blocks are liquidated when the price impact is small, while small ones are liquidated when it is large, then the estimation of the temporary price impact function that assumes stationary price impact would lead to a spurious concavity of the temporary price effect, even if that effect were linear in every period. We now use TM-CAPM to illustrate this mechanism.

In order to obtain the negative correlation between the block size and price impact, we depart from the i.i.d. assumption by introducing a smart noise $\hat{\theta}_t$ – a less extreme (and more realistic) version of a stochastic noise trade from Section 6.3 – the magnitude of which adapts to market conditions. Suppose that, in every period, apart from liquidity providers, who are present in all trading periods, we observe *occasional traders*. We assume that such traders enter the financial market only once and choose their position to maximize their mean-variance preferences

$$u_t\left(\hat{\theta}, \hat{\theta}_b\right) = \hat{\theta}_{b,t} + A \cdot \left(\theta_t^o + \hat{\theta}_t\right) - \alpha(\theta_t^o + \hat{\theta}_t) \cdot \mathcal{V}(\theta_t^o + \hat{\theta}_t), \tag{33}$$

where θ_t^o is the stochastic endowment of an occasional trader who trades in t. For simplicity, we assume that occasional traders place market orders (i.e., orders that are not contingent on prices). It follows that the presence of the occasional traders does not affect the price impacts of other investors, and we can normalize their number to one in any trading period without loss of generality. The optimal trade of the occasional trader is equal to

$$\hat{\theta}_t(\cdot) = (\bar{\mathcal{M}}_t^o + \alpha\mathcal{V})^{-1}(A - \bar{p} - \alpha\mathcal{V}\theta_t^o), \tag{34}$$

where $\bar{\mathcal{M}}_t^o$ stands for the price impact of the occasional trader in t.[27] Although stochastic, the

[26] It is also consistent with the findings of Loeb (1983), and Chan and Lakonishok (1995), who show that order size is critical for the price effect.

[27] The price effect of the trade placed by the occasional investor is greater, as, by changing $\hat{\theta}^{Av}$, the trade also

occasional trade does depend on market conditions and, in particular, on market depth.

Given that θ_t^o is i.i.d. with $E(\theta_t^o) = \theta^{Av}$, the unconditional expected price is constant and equal to the fundamental value, and the block size $\hat{\theta}_t$ is negatively correlated with the price impact $\bar{\mathcal{M}}_t^o$. Consequently, large blocks, which are typically observed in earlier trading periods when the price impact is smaller typically have a smaller temporary price impact per share than the small blocks– observed towards maturity when price impact is higher. An estimation based on the assumption of stationary price impact will then lead to a concave temporary price impact function, despite the function being linear in every period.

7 Anticipated Shocks

So far, we have considered the supply shocks that were announced on the day of the shock. Many shocks in financial markets are, however, announced long before they actually occur (e.g., changing the weights of a stock market index, including a new stock into an index). Extensive empirical evidence shows that such preannounced shocks have price effects not only on the day of the announcement, but also on the actual day of the shock.[28] The observed price effects thus cannot be attributed to any revelation of information about the fundamental value, which should have been incorporated on the day of the announcement. In this section, we discuss how prices are affected by the timing between the announcement and the occurrence of the shock and by the anticipated break-up of the shock into blocks.

Separation of the Shock Announcement and Price Effect. To examine the effect of preannouncing the shock, suppose that in period t^*, liquidity providers learn that an extra supply of assets $\hat{\theta}$ will become available in period t^{**}. The fundamental value instantaneously adjusts to its post-shock level v^* in the announcement period t^* and remains there till T, as no new information about the shocks is revealed. In the model with price takers, the price would follow the path of the fundamental value. Since the fundamental effect Δ^F, given by (28), is defined with respect to the expectation in t about the average holdings at the end of period T, we conclude that the effect occurs at the moment of the announcement *and not* at the moment when the shock is realized. In TM-CAPM, traders recognize their impact on prices and we observe an additional liquidity effect Δ^L given by (29) evaluated at t^{**}, which, unlike the unanticipated shock, takes place only in the actual period of the shock, t^{**}. This happens because the liquidity effect is not driven by information disclosure but rather by the effect that the absorption of the extra assets has on the average marginal payment. By an argument similar to the one for unanticipated shock, since the traders are, on average, buying on the shock day, the average marginal payment, which coincides with the fundamental value, is above the equilibrium price and the liquidity providers have no

induces a fundamental effect.

[28] For example, Newman and Rierson (2004) found that new bond issuance in the European telecommunication sector increased yield spreads of other firms in the sector. The effect was transitory, significant, and peaked on the day of issuance, not on the day of announcement. In addition, the severity of the effect was enhanced by asset riskiness and correlation – a finding that corresponds to our model's prediction.

incentives to arbitrage (Fact 4). The new feature of price behavior for pre-announced shocks is the time separation of the fundamental and liquidity effects (Fact 3).

In addition to the fundamental and liquidity price adjustments, when the shock is pre-announced, we observe the third effect – in all the periods between the announcement and the shock occurrence, the price is depressed by $\gamma\Delta^L$. This happens because the providers anticipate the drop in price in t^{**}, which lowers their willingness to buy the assets in all periods prior to t^{**}. The formalization of the three effects is a special case of Proposition 7. The evolution of price response to an announced once-and-for-all shock is depicted in Figure 3.

Multiple Anticipated Blocks. In practice, portfolios are often liquidated in blocks. In Section 6.3, we have already studied the price effects that result from the sales of multiple blocks; there, we assumed that, in each period, new shocks came as a surprise to liquidity providers. Here, we examine the price effects of sequential trading when the entire sequence of trades is credibly announced prior to trade at $t = 1$. We study the effect of the liquidation of a sequence of blocks $\{\hat{\theta}_t\}_t$, the total liquidated portfolio being $\hat{\theta} \equiv \sum_{t=1}^{T} \hat{\theta}_t$.

As in the case of a single unanticipated shock, following the announcement of the sequence, the fundamental value adjusts to $v^* = A - \alpha\mathcal{V}\theta^{Av} - \alpha\mathcal{V}\hat{\theta}/I$. The last term represents the cumulative fundamental effect of the sequence Δ^F; it is equal to the sum of the fundamental effects of the individual blocks, $\Delta^F = \sum_{t=1}^{T} \Delta_t^F$, where $\Delta_t^F = -\alpha\mathcal{V}\hat{\theta}_t$. The total fundamental effect is independent of how the portfolio $\hat{\theta}$ is partitioned into blocks. It follows that, if liquidity providers are price takers, the cash obtained by liquidating $\hat{\theta}$ is independent of the partition. In thin markets, however, the price path and hence the cash obtained do depend on the order of block sizes–unlike the fundamental effect, the liquidity effect is not additive.

Proposition 7 (Anticipated Multiple Blocks) *Consider a sequence of anticipated sales* $\left\{\hat{\theta}_t\right\}_t$. *In any trading period* t*, the price is equal to*

$$\bar{p}_t = v^* + \Delta_t^L + \gamma \sum_{l=1}^{T-t} \Delta_{t+l}^L, \tag{35}$$

where the liquidity effect in period t *is given by*

$$\Delta_t^L = \frac{(1-\gamma)^{2(T-t)+1}}{\gamma} \alpha\mathcal{V}\hat{\theta}_t/I. \tag{36}$$

In any period, the price departs from the fundamental value by the current liquidity effect, which is reinforced by the fraction γ of the cumulative effect of all the subsequent liquidity effects. The current liquidity effect in (35) results from the noncompetitive trading, as explained in Section 6.1. The cumulative effect generalizes the third effect discussed above in the context of a single anticipated shock; the anticipation of the depressed prices in the future reduces the incentives to buy and strengthens the incentives to sell today, which lowers the price. Interestingly, the cumulative liquidity shock affects today's price with weight γ, irrespective of how far in the future the liquidity shock occurs. The constant weight appears due to the balancing of two effects: the farther in the future the liquidity shock is from today, the smaller the fraction of today's trade that maintains a lower price until the shock period. On the other hand, the cumulative effects influence all prices between today and the period of the shock, which increases the weight.

In sum, just as when the sequence of sales is not anticipated, the asset prices in the long run are not affected by how the portfolio is divided into smaller blocks or the time at which the trade takes place. By contrast, the price path is sensitive to how the portfolio is partitioned. This occurs because future sales depress the price during the whole period between the announcement and the shock, and the effects on prices are cumulative. As a result, if the sales are anticipated (and credible), the liquidator can be expected to concentrate most of the trade in the first period.

Proposition 7 (equation (35)) further suggests that the liquidator has strong incentives not to announce the liquidation. If the sequence to be placed is announced in advance and the whole portfolio $\hat{\theta}$ is being liquidated, the fundamental value instantaneously drops by Δ^F. Furthermore, the ability to spread the current liquidity effect is reduced, as the anticipation of future liquidity effects adversely affect prices today. If instead the sequence is not announced, the fundamental value of the portfolio decreases slowly in each period when the blocks are traded. This benefits the liquidator, who receives a better price for the initial blocks. In addition, the cumulative liquidity effects are not present, and hence the sequence being placed can be arbitrarily long, making the current liquidity effects negligible.

8 Duality of Equilibrium Representation

In this section, we examine the relation between TM-CAPM from this paper and the literature on strategic market interactions in which investors submit linear (net) demand functions. The game of Nash with price-quantity schedules as strategies was introduced by Grossman (1981) and further developed in the seminal papers by Kyle (1989) and Vayanos (1999) in finance and by Klemperer and Meyer (1989) in industrial organization.[29] A dynamic game with supply functions adequately describes strategic environments in many financial markets where every liquidity provider submits a downward-slopping demand (a collection of limit orders). The existing literature on Nash in demand/supply functions focused on markets with one good and here we generalize the game of Nash in demands to markets with N goods (assets) and T trading rounds.

8.1 Strategic Representation of Noncompetitive Equilibrium

We demonstrate that within the framework with quadratic preferences, such as CARA-Normal, (dynamic) noncompetitive equilibrium from this paper is equivalent to a (subgame-perfect) Nash equilibrium in a game in which traders place linear demand orders in every period. Hence, the result (1) reveals that noncompetitive trading with endogenous price impact has a dual representation

[29] For a comprehensive review, see Vives (1999).

as a general-equilibrium model and, like in the strategic noncompetitive literature, as a Nash in demand schedules; (2) enables extending the results in that literature to a larger class of trading environments; and (3) offers new insights about trading behavior (described in Section 2.2).

We now introduce the game of Nash in demand functions that was studied in finance and industrial organization. Consider a game in which, in each trading round, investors simultaneously submit linear demand schedules $\Delta\theta_t^i(\cdot) : R^N \rightarrow R^N$, specifying their net demands for N risky assets as a function of price vector p. The literature on Nash in demands assumes that the submitted demands are downward-sloping. This assumption generalizes to the case of N assets as a condition that the Jacobians of the submitted demand functions, $D\Delta\theta_t^i(\cdot)$, are negative definite. A market maker aggregates the submitted schedules and finds the price vector $\tilde{p}_t \in R^N$ that clears markets for all assets $\sum \Delta\theta_t^i(\tilde{p}_t) = 0$. The payoffs are as follows: in every period t, trader i buys $\Delta\theta_t^i(\tilde{p}_t)$ risky assets and pays $\Delta\theta_t^i(\tilde{p}_t) \cdot \tilde{p}_t$ in terms of the riskless asset. The trader's payoff is then determined by his utility evaluated at the ultimate holdings of assets after T rounds of trade. To close the game, we assume that, given the submitted demand functions, when the market-clearing price $\tilde{p}_t$ does not exist or there is more than one such price vector, no trade takes place and each trader receives payoff equal to their utility without trade.

A well-known problem with the static game in demand functions is that it yields a continuum of equilibria in deterministic settings. Intuitively, except at the equilibrium price, Nash does not put any restrictions on the quantities specified by a player's submitted demand. Therefore, the shape (slope) of the best response is not determined. Since the player's submitted demands define the residual supply faced by other traders, the quantities optimal for the other traders are functions of the slopes of these demands, i.e., of trades at prices other than the equilibrium price. The extra degree of freedom in selecting best responses translates into the multiplicity of Nash equilibria.

In order to refine the set of equilibria, it is common in the literature to add noise trade, ε, the distribution of which is then shrunk to zero. Thereby, the literature focuses on the linear Nash equilibrium in demands that is *robust* to adding any noise with an arbitrary distribution. When selecting the demand function to submit, investors do not know the particular realization of ε but only the distribution of ε. Intuitively, the noise introduces uncertainty about the residual supply and induces each player to respond optimally to *all* prices. This determines the slope of the submitted best response and pins down a unique Nash equilibrium.

Just as Nash equilibrium gives weak predictions in games with demand functions, the subgame-perfect Nash is not determinate in dynamic games in demands. Therefore, we now extend the equilibrium selection argument to dynamic games with T periods. We say that a profile of demand functions $\{\Delta\theta_t^i(\cdot)\}_{t,i}$ is a *robust* subgame-perfect Nash equilibrium if, for any $\bar{t} = 1, ..., T$ and any subgame starting at $\bar{t}$, $\{\Delta\theta_t^i(\cdot)\}_{t \geq \bar{t}}$ is a Nash equilibrium that is robust to adding noise at $\bar{t}$. Observe that this preserves the perfect-foresight structure of the model, as, in every subgame, the uncertainty is present only in the first trading round.

To formally state the equivalence result, for any profile of demands $\{\Delta\theta_t(\cdot)\}_t$, we define an *associated profile* $\{\tilde{p}_t, \Delta\tilde{\theta}_t, \tilde{\mathcal{M}}_t\}_t$: Let $\{\tilde{p}_t\}_t$ be the market clearing prices ($\sum \Delta\theta_t^i(\tilde{p}_t) = 0$ for every

t), and let $\left\{\Delta\tilde{\theta}_t^i\right\}_t$ be the corresponding equilibrium trades of investor i, where $\Delta\tilde{\theta}_t^i \equiv \Delta\theta_t^i(\tilde{p}_t)$. Finally, we define sequence of price impacts $\left\{\tilde{\mathcal{M}}_t^i\right\}_t$. The residual supply for trader i in t, $p_t^s(q_i)$ is the inverse of the negative of $\sum_{j\neq i}\Delta\theta_t^j(p)$, and the trader's i price impact matrix in t, $\tilde{\mathcal{M}}_t^i$ is a slope of the residual supply, $\tilde{\mathcal{M}}_t^i \equiv Dp_t^s\left(\Delta\tilde{\theta}_t^i\right)$. In a robust subgame-perfect Nash equilibrium, the associated profile exists and is unique. Let $\Delta\bar{\theta}_t^i(\cdot,\mathcal{M}_t)$ denote the noncompetitive demand function of trader i, which generalizes (9) (and is derived in Appendix I, equation (45)).

Proposition 8 (Equivalence of Equilibrium Representation) *For any robust subgame-perfect Nash equilibrium in demands $\{\Delta\theta_t(\cdot)\}_t$, its associated profile $\{\tilde{p}_t, \Delta\tilde{\theta}_t, \tilde{\mathcal{M}}_t\}_t$ constitutes a dynamic noncompetitive equilibrium. Conversely, for any dynamic noncompetitive equilibrium $\{\bar{p}_t, \Delta\bar{\theta}_t, \bar{\mathcal{M}}_t\}_t$, functions $\Delta\theta_t^i(\cdot) \equiv \Delta\bar{\theta}_t^i(\cdot, \bar{\mathcal{M}}_t)$ constitute a robust subgame-perfect Nash equilibrium.*

So far, we have focused on the implications of Proposition 8 for the noncompetitive model presented in this paper. The result has implications for the literature on strategic trading in demands as well. Namely, it implies that the robust equilibrium of Nash in demands has a general-equilibrium representation. Hence, all results about noncompetitive trading established in this paper (and other papers on noncompetitive equilibrium) carry over to models based on games in demand functions. In addition, the setting in the demand-functions literature is extended to dynamic games with an arbitrary number of goods and traders who can *all* be strategic and asymmetric in their endowments.

8.2 Competitive, Noncompetitive and Nash Equilibria

Using the strategic representation of our model, we close by comparing the equilibrium conditions in deterministic models of market interactions based on a uniform price, such as the competitive equilibrium, the noncompetitive equilibrium from this paper, and the Nash equilibrium. We show how the different concepts of equilibrium can be related using two conditions: market clearing and optimization by strategic players. We restrict attention to one-period markets.

We begin by observing that although the competitive outcome can be rationalized by Nash equilibrium, competitive bidding by all traders in a game in which a finite number of risk averse players submit demand functions is not Nash (and hence cannot rationalize the competitive outcome). The outcome of a competitive equilibrium can be rationalized by a game of Nash (but not robust Nash) in demands. For instance, a game in which (at least) two players submit perfectly elastic bids (as in Bertrand) results in competitive outcome; the actual price impacts $\tilde{\mathcal{M}}^i$ of all traders are then equal to zero. Can a competitive bidding that equalizes the marginal utility with a price (for any price), which defines the bidding schedule $\Delta\theta^i(p)$ rationalize competitive outcome as well? The competitive bids of risk averse agents are downward sloping and, hence, in a finite market, the residual supply faced by trader i is not perfectly elastic at the market clearing price $\tilde{p}$, but it has a slope of $\tilde{\mathcal{M}}^i > 0$. Note that $\tilde{\mathcal{M}}^i$ is the actual price impact in the market with I traders who take prices as given (i.e., it is consistent with the competitive beliefs of other traders in the sense of Definition 3). Nash behavior dictates that each trader best-responds given others' bids. In other words his bidding schedule equalizes his marginal utility and his marginal revenue only *at* the market clearing price $\tilde{p}$; that is, the associated trades $\Delta\tilde{\theta}^i \equiv \Delta\theta^i(\tilde{p})$ and prices $\tilde{p}$ satisfy

$$A - \alpha\mathcal{V}\left(\theta_0^i + \Delta\tilde{\theta}^i\right) = \tilde{p} + \tilde{\mathcal{M}}^i\Delta\tilde{\theta}^i. \tag{37}$$

Clearly, the trades realized from the competitive bids $\Delta\tilde{\theta}^i$ and the competitive price $\tilde{p}$ do not satisfy the best-response condition (37), because the competitive bid $\Delta\theta^i(\cdot)$ is determined by the condition $A - \alpha\mathcal{V}\left(\theta_0^i + \Delta\theta^i(p)\right) = p$, for any price, including $p = \tilde{p}$.

In deterministic models, condition (37) is, in fact, necessary (and, with linear bids, sufficient) for any collection for bid functions of the players $\{\Delta\theta^i(\cdot)\}_i$ to constitute a Nash equilibrium in any game based on a uniform price, such as Bertrand, Cournot, and Nash in demands. Unless the game itself pins down price impacts,[30] Nash equilibrium is not determinate. In games with demands as strategies, Nash equilibrium does not pin down the slopes of individual bids of the bidders, which implies that the consistent price impacts $\tilde{\mathcal{M}}^i$ are not determined. In light of the general-equilibrium representation of the noncompetitive equilibrium, the multiplicity of Nash equilibria arises because one of the three elements of equilibrium is free.

The game-theoretic representation of noncompetitive equilibrium (robust Nash) requires that condition (37) holds at any price and not just at the equilibrium price. This identifies the additional restriction on the bidding behavior placed by the noncompetitive equilibrium, which refines the set of Nash equilibria of the demand game. Unlike in Nash, investors equalize their marginal utilities and revenues at all prices given their assessments $\tilde{\mathcal{M}}^i$, which pins down the slope of their own bids. The restriction is embedded in the equilibrium selection based on adding noise in the game-theoretic representation and the one based on endogenous price impacts in the general-equilibrium representation.

There are noncompetitive general-equilibrium models, in which traders have exogenously given assessments of their price impacts $\left\{\mathcal{M}^i\right\}_i$ (e.g., Negishi [1960], Huberman and Stanzl [2004]). Such models generalize the notion of competitive equilibrium by allowing positive price impacts. As in the competitive equilibrium, the outcomes (as opposed to bidding) in these models can be rationalized by Nash equilibrium in demands – for any profile of exogenous price impacts $\{\mathcal{M}^i\}_i$, one can find bidding schedules $\{\Delta\theta^i(\cdot)\}_i$ that induce price impacts that coincide with the exogenous ones.

Is noncompetitive bidding consistent with exogenous price impacts a Nash equilibrium? In parallel with competitive bidding, suppose that players submit bids that, for any price, equalize marginal utility with marginal payment given their exogenously given assessments $\left\{\mathcal{M}^i\right\}_i$; that is, they bid according to $A - \alpha\mathcal{V}\left(\theta_0^i + \Delta\theta^i\right) = p + \mathcal{M}^i\Delta\theta^i$, for any price p, and submit

$$\Delta\theta^i(p) = \left(\mathcal{M}^i + \alpha\mathcal{V}\right)^{-1}\left[A - \alpha\mathcal{V}\theta_0^i - p\right]. \tag{38}$$

[30] For example, the Cournot game defines the price impact of all players to be equal to the slope of the market demand; the Bertrand game defines the players' price impacts to be equal to zero.

Such bidding captures trading behavior in many financial markets, where investors trade assuming some estimated price impact. Since, generically, $\tilde{\mathcal{M}}^i \neq \mathcal{M}^i$, bids (38) will typically not constitute a Nash equilibrium. The bids will be Nash if and only if the exogenously assumed price impacts coincide with the consistent price impacts from the noncompetitive equilibrium.

It is interesting to contrast the informational requirements behind the robust Nash (and, by Proposition 8, noncompetitive) and other Nash equilibria. To implement a noncompetitive equilibrium as a Nash in demands, it suffices that, apart from his own preference, each player knows his own consistent price impact $\bar{\mathcal{M}}^i$ and submits the strategy (38) with $\bar{\mathcal{M}}^i$. This would not suffice to implement other Nash equilibria – submitting (38) with the exogenously given $\mathcal{M}^i$ would induce different actual price impacts than $\left\{\mathcal{M}^i\right\}_i$. Nash equilibrium could be implemented if, in addition, each trader knows the slope of his own bid (which rationalizes the price impacts of others). The mild informational requirement of the noncompetitive equilibrium provides an argument for why the selection of Nash that is implicit in our model might be attractive in modeling of financial markets.

9 Discussion

TM-CAPM suggests that even if price impact does not affect equilibrium returns, accounting for the very presence of price impact can help one understand order break-up, asset price overshooting, limits to arbitrage, commonality in liquidity, cross-market liquidity effects, the empirical evidence on the behavior of return volatility, the shape of permanent and temporary price impact functions, the existence of valuation instruments, such as the blockage discount, etc. The empirical literature on illiquid markets has demonstrated that expected asset returns are higher for illiquid assets, as the liquidity premium compensates for the low marketability of an asset. Therefore, our finding that market impact does not affect asset prices, and hence their returns, is not confirmed by the data. Notice, however, that in the setting studied in this paper, there is nothing that would represent the traders' concern about having to liquidate part of their portfolio prior to maturity. Rostek and Weretka (2008) model traders who, in every period, assign a positive (and arbitrarily small) probability to a distress situation in which they receive only the liquidation value of their holdings. Crucially, if the traders were price takers, introducing the liquidity concern would have no effects on prices and allocations, as the cash value from liquidation would coincide with the value at maturity. It is the interaction of price impact and the liquidity concern that introduces new effects the main result of which is the derivation of liquidity premia that are time-varying and depend on the equilibrium dynamics of the price impact.

Appendix I

Proof. (Proposition 1: Dynamically Competitive Markets) In the last period, a competitive trader j will equalize his marginal utility with the price $\bar{p}_T$. Hence, portfolio $\bar{\theta}_T^j$ does not depend on the trade in any period $t < T$. By the perfect foresight of the deterministic price in T, the value function of trader j in any period $t < T$ is given by

$$V_t^j\left(\Delta\theta_t^j, \Delta\theta_{b,t}^j\right) = \Delta\theta_{b,t}^j + \bar{p}_T \cdot \Delta\theta_t^j + c_t. \tag{39}$$

Thus, trader j behaves as risk neutral in all periods $t < T$. The (competitive) demand of the trader in $t < T$ is perfectly elastic, and it defines a horizontal residual supply for other traders. It follows that, in all periods $t < T$, the price impacts of all traders other than j are equal to zero and the prices are equal to $\bar{p}_T$. We now argue that, given identical prices in both periods, after the trade in $T-1$ takes place, the risky holdings coincide across traders and hence are equal to θ^{Av} for all traders. In period T, trader j equalizes his marginal utility with $\bar{p}_T$, which implies that his marginal utility in T is equal to the marginal valuation of any other trader in $T-1$, who equalizes his marginal valuation with $\bar{p}_T$ in this period. Since these traders are non-competitive in T, this further implies that their portfolios at the end of $T-1$ are fully diversified (otherwise their portfolios would not be the same after the final round of trade). Given that the portfolios of all traders, including j, are equal to θ^{Av} after round $T-1$, the prices in all periods are equal to $\bar{v}$. ■

Derivation of Noncompetitive equilibrium with T trading periods (Propositions 2, 3, and 4). We derive the unique symmetric equilibrium in TM-CAPM using the following strategy. We first conjecture the functional form of the value function in period t, find the noncompetitive equilibrium in t, and finally determine recursively the coefficients of the value function.

For each investor i, define an auxiliary portfolio $\tilde{\theta}_t^i$ as a convex combination of the current and the average inventory

$$\tilde{\theta}_t^i \equiv (1-\lambda_t)\theta^{Av} + \lambda_t\bar{\theta}_{t-1}^i, \tag{40}$$

where the scalar λ_t is given by $(1-\gamma)^{T-t}$. Observe that $\tilde{\theta}_t^i$ is predetermined at t. For any period t, we propose the following value function

$$V_t^i(\Delta\theta_t^i, \Delta\theta_{b,t}^i) = \Delta\theta_{b,t}^i + ((1-\lambda_t)v + \lambda_t A) \cdot \Delta\theta_t^i - \frac{\alpha}{2}(\tilde{\theta}_t^i + \lambda_t\Delta\theta_t^i) \cdot \bar{\mathcal{V}}(\tilde{\theta}_t^i + \lambda_t\Delta\theta_t^i) + c_t, \tag{41}$$

where constant c_t is independent of trades $(\Delta\theta_t^i,\, \Delta\theta_{b,t}^i)$, and the policy function is

$$\Delta\bar{\theta}_t^i = \gamma(\theta^{Av} - \theta_{t-1}^i). \tag{42}$$

As a check, in the last trading period, T, the value function coincides with the utility function (1), given that $\lambda_T = 1$ and $c_T = \bar{\theta}_{b,T-1}^i$. In hindsight, for $t < T$, the functional form in (41) can be motivated as follows. As we show, in any given trading period, it is optimal for each investor to partially replace his risky holdings with the average inventory. From the perspective of period $t < T$, only a fraction λ_t of the current trade $\Delta\theta_t^i$ survives until maturity in T. Hence, only λ_t of the current trade adds to the riskiness of the ultimate portfolio $\bar{\theta}_T^i$ and determines the effective concavity of the value function at t. The remaining part, $1-\lambda_t$, is liquidated in subsequent periods at market prices equal to v. We suggestively refer to λ_t as a *survival rate*. For the candidate value function (41), we derive the equilibrium in period t and determine the corresponding parameters λ_t and c_t.

The individual portfolio choice is affected by the presence of price impact. Consider the budget set of investor i, who chooses trades $\Delta\theta_t^i$ and $\Delta\theta_{b,t}^i$ in period t, given the residual demands $p_{p_t,\Delta\bar{\theta}_t^i,\mathcal{M}_t^i}(\cdot)$,

$$p_{p_t,\Delta\bar{\theta}_t^i,\mathcal{M}_t^i}(\Delta\theta_t^i)\cdot\Delta\theta_t^i+\Delta\theta_{b,t}^i\leq 0. \tag{43}$$

Because the prices of risky assets are functions of the quantities demanded $\Delta\theta_t^i$, the budget constraint (43) is quadratic rather than linear in $\Delta\theta_t^i$. Optimal trades of assets are characterized by the equality between marginal utilities and marginal revenues, which, unlike in the competitive market, depend on the quantity traded.[31] At the optimum, the marginal revenues from selling each asset exceed the prices by $\mathcal{M}_t^i\Delta\bar{\theta}_t^i$,

$$(1-\lambda_t)v+\lambda_t A-\lambda_t\alpha\mathcal{V}\cdot(\tilde{\theta}_t^i+\lambda_t\Delta\bar{\theta}_t^i)=p_t+\mathcal{M}_t^i\Delta\bar{\theta}_t^i. \tag{44}$$

With a positive semidefinite $\mathcal{M}_t^i$, we solve (44) for the individual asset demand of investor i as a function of price and price impact

$$\Delta\bar{\theta}_t^i(p_t,\mathcal{M}_t^i)=(\lambda_t^2\alpha\mathcal{V}+\mathcal{M}_t^i)^{-1}((1-\lambda_t)v+\lambda_t A-\alpha\lambda_t\mathcal{V}\cdot\tilde{\theta}_t^i-p_t). \tag{45}$$

If price impacts $\mathcal{M}_t^i$ are positive, a trader reduces his order, i.e., for any given price, he buys or sells less than the competitive trader. Note in passing that, at the equilibrium prices $\bar{p}_t$ and price impacts $\bar{\mathcal{M}}_t^i$, the net trades of all traders must sum to zero: $\sum_{i\in I}\Delta\bar{\theta}_t^i(\bar{p}_t,\bar{\mathcal{M}}_t^i)=0$.

Next, we determine $\bar{\mathcal{M}}_t$. Let $\Delta\theta_t^i$ be an arbitrary off-equilibrium deviation of trader i. For markets to clear, the price must decrease, so that the other investors are willing to purchase the additional shares. By market clearing and optimization of other traders, the required price change satisfies

$$\Delta\theta_t^i+\sum_{j\neq i}\Delta\bar{\theta}_t^j(p_t,\mathcal{M}_t^j)=0. \tag{46}$$

Individual demands (45) for $j\neq i$ and (46), define price p_t for any deviation,

$$p_{\bar{p}_t,\Delta\bar{\theta}_t^i,\bar{\mathcal{M}}_t^i}(\Delta\theta^i)=\bar{p}_t+\underbrace{(\sum_{j\neq i}(\mathcal{M}_t^j+\lambda_t^2\alpha\mathcal{V})^{-1})^{-1}}_{\bar{\mathcal{M}}_t^i}(\Delta\theta_t^i-\Delta\bar{\theta}_t^i). \tag{47}$$

Thus, in any period the price impact is

$$\bar{\mathcal{M}}_t^i=(\sum_{j\neq i}(\bar{\mathcal{M}}_t^j+\lambda_t^2\alpha\mathcal{V})^{-1})^{-1}=(1-\gamma)\mathcal{H}(\bar{\mathcal{M}}_t^j+\lambda_t^2\alpha\mathcal{V}|j\neq i), \tag{48}$$

where $\mathcal{H}(\cdot)$ is the harmonic mean operator for positive definite matrices.

The profile of consistent price impact $\bar{\mathcal{M}}_t$ is the fixed point of the map defined by I equalities (48), one for each trader. In the unique symmetric solution ($\bar{\mathcal{M}}_t^i=\bar{\mathcal{M}}_t^j$), the consistent price impact is the same for all i and has a closed-form solution

$$\bar{\mathcal{M}}_t^i=\frac{(1-\gamma)}{\gamma}\lambda_t^2\alpha\mathcal{V}. \tag{49}$$

[31] Geometrically, condition (44) corresponds to the familiar tangency between an indifference curve and the budget set, the latter being represented by a parabola, the slope of which is given by the vector of the–now quantity-dependent–marginal revenues. The strict convexity of both the objective function and the budget set assures that this condition is sufficient as well as necessary for optimality.

The equilibrium asset prices can be derived from the market-clearing condition by substituting demands (45)

$$0 = \frac{1}{I} \sum_{i \in I} \Delta\bar{\theta}^i(\bar{p}_t, \bar{\mathcal{M}}_t^i) = v - \bar{p}_t, \tag{50}$$

where we used the symmetry of $\bar{\mathcal{M}}_t^i + \lambda_t^2 \alpha \mathcal{V}$, the definition of v, (6), and the fact that $\frac{1}{I}\sum_{i \in I} \tilde{\theta}_t^i$ is equal to the average inventory. The equilibrium prices in each period are

$$\bar{p}_t = v \equiv A - \alpha \mathcal{V} \theta^{Av}. \tag{51}$$

The portfolios in each period are obtained by substituting (51) and (49) into (45)

$$\Delta\bar{\theta}_t^i = \gamma(\theta^{Av} - \bar{\theta}_{t-1}^i). \tag{52}$$

Survival rate λ_t. We show that the candidate value function (41) is valid in an arbitrary period $t < T$ applying an inductive argument. We have already verified that (41) holds in T. If (41) is satisfied for any $t' > t$, the policy function (52) implies that for any trade $\Delta\theta_t^i$, the ultimate risky portfolio after trade in T can be written as

$$\bar{\theta}_T^i = (1 - (1-\gamma)^{T-t})\theta^{Av} + (1-\gamma)^{T-t}(\bar{\theta}_{t-1}^i + \Delta\theta_t^i). \tag{53}$$

With λ_t given by

$$\lambda_t \equiv (1-\gamma)^{T-t}, \tag{54}$$

portfolio $\bar{\theta}_T^i$ can be written as $\bar{\theta}_T^i = \tilde{\theta}_t^i + \lambda_t \Delta\theta_t^i$. The bond holdings after trade are given by

$$\bar{\theta}_{b,T}^i = \Delta\bar{\theta}_{b,t}^i + v(1-\lambda_t)\Delta\bar{\theta}_t^i + c_t, \tag{55}$$

where the constant is equal to

$$c_t = \bar{\theta}_{b,t}^i - v(1-\lambda_t)(\theta^{Av} - \bar{\theta}_{t-1}^i). \tag{56}$$

Applying (53), (55) and (56) in the utility function (1) establishes (41) as the functional form of the value function in period t. Propositions 2, 3, 4 follow from the derivation of the equilibrium.

Proof. (Proposition 2: Three-Fund Separation) The assertion is implied by (52). ∎

Proof. (Proposition 3: Time Structure of Price Impact) The result is implied by (49) and (54). ∎

Proof. (Proposition 4: Security Market Line) The result follows from two observations: first, by (51), the prices, and hence the asset returns, are as in the competitive model; second, formula (27) holds in the competitive model. ∎

The proof of Proposition 5 is presented after the proof of Proposition 7.

Proof. (Proposition 6: Price Manipulation) Consider an unanticipated round-trip trade of N risky assets, i.e., a non-zero sequence of trades $\left\{\hat{\theta}_t\right\}_t$ such that $\sum_t \hat{\theta}_t = 0$. The price vector in each period is given by

$$\bar{p}_t = v - \frac{\alpha}{I}\left(\frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\mathcal{V}\hat{\theta}_t + \sum_{k\leq t}\mathcal{V}\hat{\theta}_k\right), \tag{57}$$

where v is the fundamental value in the absence of the round trip, the first element in parentheses is the liquidity effect in t, while the second element corresponds to the fundamental effects of the demand or supply shocks induced by the round trip up to t. The cash obtained from the round-trip trade is equal to

$$\begin{aligned}\sum_{t=1}^{T}\bar{p}_t\cdot\hat{\theta}_t &= \sum_{t=1}^{T}\left[v - \left(\frac{\alpha}{I}\frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\mathcal{V}\hat{\theta}_t + \sum_{k\leq t}\mathcal{V}\hat{\theta}_k\right)\right]\cdot\hat{\theta}_t = \\ &= -\frac{\alpha}{I}\sum_{t=1}^{T}\frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_t - \frac{\alpha}{I}\sum_{t=1}^{T}\left(\sum_{k\leq t}\mathcal{V}\hat{\theta}_k\right)\cdot\hat{\theta}_t, \end{aligned}\tag{58}$$

where we have eliminated the element $v\sum_{t=1}^{T}\hat{\theta}_t$ by the round-trip assumption. The last sum can be decomposed as follows

$$\sum_{t=1}^{T}\left(\sum_{k\leq t}\mathcal{V}\hat{\theta}_k\right)\cdot\hat{\theta}_t = \sum_{t=2}^{T}\sum_{k<t}\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_k + \sum_{t}\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_t. \tag{59}$$

The sum $\sum_{t=2}^{T}\sum_{k<t}\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_k$ consists of all (h,k) combinations of $\hat{\theta}_h\cdot\mathcal{V}\hat{\theta}_k$ such that $h\neq k$ and each combination enters exactly once. In addition, since the elements are symmetric ($\hat{\theta}_h'\mathcal{V}\hat{\theta}_k = \hat{\theta}_k'\mathcal{V}\hat{\theta}_h$), the sum, augmented by $\frac{1}{2}\sum_t\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_t$, can be written as

$$\sum_{t=2}^{T}\sum_{k<t}\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_k + \frac{1}{2}\sum_{t}\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_t = \frac{1}{2}\sum_{k=1}^{T}\hat{\theta}_k\cdot\mathcal{V}\sum_{t=1}^{T}\hat{\theta}_t = 0, \tag{60}$$

where the final equality follows, again, from the round-trip assumption, $\sum_{t=1}^{T}\hat{\theta}_t = 0$. We obtain

$$\sum_{t=1}^{T}\bar{p}_t\cdot\hat{\theta}_t = -\frac{\alpha}{I}\sum_{t=1}^{T}\left(\frac{(1-\gamma)^{2(T-t)+1}}{\gamma} + \frac{1}{2}\right)\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_t. \tag{61}$$

Using that $\mathcal{V}$ is positive definite, we have $\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_t \geq 0$, with a strict inequality for $\hat{\theta}_t\neq 0$. Since $\hat{\theta}_t\neq 0$ for at least one t, the proposition's assertion follows

$$\sum_{t=1}^{T}\bar{p}_t\cdot\hat{\theta}_t < 0. \tag{62}$$

■

Proof. (PROPOSITION 7: ANTICIPATED MULTIPLE BLOCKS)[32] Let $(\hat{\theta}_1,...,\hat{\theta}_T)$ be a sequence of anticipated shocks. For the vector of shocks equal to zero, this proof can be applied to establish all of the paper's results that characterize the dynamic noncompetitive equilibrium. Therefore, this proof focuses on the steps that pertain specifically to Proposition 7. We proceed by induction. In the last trading period, T, the value function is

[32] Recall that $\theta_t^{Av} = \frac{1}{I}\sum_i\bar{\theta}_t^i$ is the average inventory at the beginning of the trading period t, v^* is the average marginal utility (i.e., the fundamental value) after the last period of trade, $v^* = A - \alpha\mathcal{V}\theta_{T+1}^{Av}$, and the temporary effect is equal to $\Delta_t^{Temp} = \alpha\mathcal{V}(1-\gamma)^{2(T-t)+1}/\gamma$.

$$V_T^i(\Delta\theta_T^i, \Delta\theta_{b,T}^i) = \Delta\theta_{b,T}^i + \bar{\theta}_{b,T-1}^i + A \cdot (\bar{\theta}_{T-1}^i + \Delta\theta_T^i) - \frac{\alpha}{2}(\bar{\theta}_{T-1}^i + \Delta\theta_T^i) \cdot \mathcal{V}(\bar{\theta}_{T-1}^i + \Delta\theta_T^i). \tag{63}$$

Setting the marginal utility and the marginal revenue equal to each other yields

$$A - \alpha\mathcal{V}(\bar{\theta}_{T-1}^i + \Delta\bar{\theta}_T^i) = \bar{p}_T + \bar{\mathcal{M}}_T^i \Delta\bar{\theta}_T^i. \tag{64}$$

Solving for the optimal trade $\Delta\bar{\theta}_T^i(\bar{p}_T, \bar{\mathcal{M}}_T^i)$ from (64) and summing the derived functions for all trading partners of i, we obtain i's residual supply, the slope of which gives i's price impact equal to $\bar{\mathcal{M}}_T^i = (1-\gamma)\mathcal{H}(\alpha\mathcal{V} + \bar{\mathcal{M}}_T^j | j \neq i)$; in a symmetric equilibrium, $\bar{\mathcal{M}}_T^i = \alpha\mathcal{V}(1-\gamma)/\gamma$. Averaging the F.O.C. (64) across all trades, using $\sum_{i\in I} \Delta\bar{\theta}_T^i = \hat{\theta}_T$ and substituting for the price impact gives

$$A - \alpha\mathcal{V}(\theta_T^{Av} + \frac{\hat{\theta}_T}{I}) = \bar{p}_T + \frac{1-\gamma}{\gamma}\alpha\mathcal{V}\frac{\hat{\theta}_T}{I}, \tag{65}$$

from which we can derive the equilibrium prices in period T,

$$\bar{p}_T = v^* - \frac{1-\gamma}{\gamma}\alpha\mathcal{V}\frac{\hat{\theta}_T}{I} = v + \Delta_T^L. \tag{66}$$

Substituting the derived prices (66) back into the F.O.C. (64), we obtain policy functions for the equilibrium trades,

$$\Delta\bar{\theta}_T^i = \gamma(\theta_T^{Av} - \bar{\theta}_{T-1}^i) + \frac{\hat{\theta}_T}{I}. \tag{67}$$

This demonstrates that the formulas for policy function, price, and price impact proposed in Proposition 7 hold in T. Suppose now that the formulas hold in all periods from $t+1$ to T. We will show, then, that they also hold in t. The ultimate portfolio, as a function of trades in period t, is given by

$$\bar{\theta}_T\left(\Delta\theta_t^i, \Delta\theta_{b,t}^i\right) = (1-(1-\gamma)^{T-t})\theta_{t+1}^{Av} + \underbrace{(1-\gamma)^{T-t}}_{\lambda_t}(\bar{\theta}_{t-1}^i + \Delta\theta_t^i) + \sum_{l=1}^{T-t}\frac{\hat{\theta}_{t+l}}{I}. \tag{68}$$

In addition, the trade of risky assets in the l^{th} period following t, as a function of trades in t, is equal to

$$\Delta\theta_{t+l}^i = -\gamma\,(1-\gamma)^{l-1}\left(\bar{\theta}_{t-1}^i + \Delta\theta_t^i\right) + c_1^i, \tag{69}$$

where c_1^i is a constant that does not depend either on $\Delta\theta_t^i$ or $\Delta\theta_{b,t}^i$. This implies that the holdings of bonds in T are given by

$$\bar{\theta}_{b,T}^i\left(\Delta\theta_t^i, \Delta\theta_{b,t}^i\right) = \bar{\theta}_{b,t-1}^i + \Delta\theta_{b,t}^i + \sum_{l=1}^{T-t}\Delta\theta_{t+l}^i \cdot \bar{p}_{t+l} = \bar{\theta}_{b,t-1}^i + \Delta\theta_{b,t}^i + \left(\bar{\theta}_{t-1}^i + \Delta\theta_t^i\right)\sum_{l=1}^{T-t}\bar{p}_{t+l}\gamma\,(1-\gamma)^{l-1} + c_2^i, \tag{70}$$

where c_2^i is independent of $\Delta\theta_t$ and $\Delta\theta_{b,t}$. Applying (68) and (70) to (1), we observe that the value function is linear in the trade of bonds $\Delta\theta_{b,t}^i$ and quadratic in $\Delta\theta_t^i$. The derivative of the value function with respect to $\Delta\theta_t^i$ is given by

$$\frac{\partial V_t^i(\cdot)}{\partial\Delta\theta_t^i} = \lambda_t\left[A - \alpha\mathcal{V}\left((1-\lambda_t)\theta_{t+1}^{Av} + \lambda_t(\bar{\theta}_{t-1}^i + \Delta\theta_t^i) + \sum_{l=1}^{T-t}\frac{\hat{\theta}_{t+l}}{I}\right)\right] + \sum_{l=1}^{T-t}\bar{p}_{t+l}\gamma\,(1-\gamma)^{l-1}. \tag{71}$$

Using the quasilinearity of the value function, the first-order (necessary and sufficient) optimality condition is $\partial V_t^i(\cdot)/\partial\Delta\theta_t^i = \bar{p}_t + \bar{\mathcal{M}}_t^i\Delta\theta_t^i$, for any $\bar{p}_t$ and $\bar{\mathcal{M}}_t^i$. This allows solving for the optimal trade $\Delta\bar{\theta}_t^i(\bar{p}_t, \bar{\mathcal{M}}_t^i)$. Summing the derived trade functions for all $j \neq i$, we find i's price impact to be equal to $\bar{\mathcal{M}}_t^i = (1-\gamma)\,\mathcal{H}((\lambda_t)^2\,\alpha\mathcal{V} + \bar{\mathcal{M}}_t^j | j \neq i)$; in a symmetric equilibrium, $\bar{\mathcal{M}}_t^i = \alpha\mathcal{V}(1-\gamma)^{2(T-t)+1}/\gamma$, as desired. Applying the derived price impacts $\bar{\mathcal{M}}_t^i$ and the definition of liquidity effect Δ_t^L in F.O.C. and averaging the F.O.C. across all traders, we arrive at

$$\underbrace{\lambda_t\left(A + \alpha\mathcal{V}\left(\theta_{t+1}^{Av} + \sum_{l=1}^{T-t}\frac{\hat{\theta}_{t+l}}{I}\right)\right)}_{\text{Term 1}} + \underbrace{\sum_{l=1}^{T-t}\bar{p}_{t+l}\gamma\,(1-\gamma)^{l-1}}_{\text{Term 2}} = \bar{p}_t - \Delta_t^L. \tag{72}$$

Term 1 is equal to $\lambda_t v^*$ and Term 2 is a weighted sum of prices for the periods following t. Since all prices are linear functions of v^* and Δ_{t+l}^L for all $l > 0$, Term 2 is also a linear function of those variables. We next determine the coefficients that multiply v^* and Δ_{t+l}^L for any $l > 0$. Using the fact that v^* enters the prices in all periods in Term 2, we find the coefficient that multiplies v^*

$$v^*\sum_{l=1}^{T-t}\gamma\,(1-\gamma)^{l-1} = \gamma\left(\frac{1-(1-\gamma)^{T-t}}{1-1-\gamma}\right) = v^*\left(1-(1-\gamma)^{T-t}\right) = v^*\,(1-\lambda_t)\,. \tag{73}$$

For any $k = t+1, ..., T$, the liquidity effect Δ_k^L enters Term 2 through all prices between t and $k-1$ (multiplied by coefficient γ) and the price in period k (with the coefficient of 1). Crucially, this term is not present in the prices following period k (see (35)). This allows us to find the sum of all the components that contain Δ_k^L in Term 2 as

$$\gamma\Delta_k^L\left(\gamma\sum_{l=1}^{k-1-t}(1-\gamma)^{l-1} + (1-\gamma)^{k-t-1}\right) = \gamma\Delta_k^L\left(\gamma\frac{1-(1-\gamma)^{k-t-1}}{\gamma} + (1-\gamma)^{k-t-1}\right) = \gamma\Delta_k^L. \tag{74}$$

Observe that this holds for any $k = t+1, ..., T$. Hence,

$$\text{Term 2} = v^*\,(1-\lambda_t) + \gamma\sum_{l=1}^{T-t}\Delta_{t+l}^L. \tag{75}$$

Therefore, the averaged F.O.C. (72) simplifies to

$$\underbrace{\lambda_t v^*}_{\text{Term 1}} + \underbrace{v^*\,(1-\lambda_t) + \gamma\sum_{l=1}^{T-t}\Delta_{t+l}^L}_{\text{Term 2}} = v^* + \gamma\sum_{l=1}^{T-t}\Delta_{t+l}^L = \bar{p}_t + \Delta_t^L. \tag{76}$$

Solving for the price

$$\bar{p}_t = v^* + \Delta_t^L - \gamma\sum_{l=1}^{T-t}\Delta_{t+l}^L, \tag{77}$$

establishes that the equilibrium price in t is as asserted by Proposition 7. To complete the proof, we verify that the policy function for risky assets holds as well. Using the equilibrium price impact, we find that the F.O.C. is given by

$$\lambda_t\left(A-\alpha\mathcal{V}\left(\left(1-\lambda_t\right)\theta^{Av}_{t+1}+\lambda_t\left(\bar{\theta}^i_{t-1}+\Delta\theta^i_t\right)+\sum_{l=1}^{T-t}\frac{\hat{\theta}_{t+l}}{I}\right)\right)+\sum_{l=1}^{T-t}\bar{p}_{t+l}\gamma\left(1-\gamma\right)^{l-1} \tag{78}$$
$$= \bar{p}_t+\frac{\left(1-\gamma\right)^{2(T-t)+1}}{\gamma}\alpha\mathcal{V}\Delta\theta^i_t.$$

Substituting for the equilibrium prices and the value of v^*, we obtain the policy function in t,

$$\Delta\theta^i_t=\gamma\left(\theta^{Av}_t-\bar{\theta}^i_{t-1}\right)+\frac{\hat{\theta}_t}{I}. \tag{79}$$

■

Proof. (Proposition 5: Asset Price Overshooting) In Proposition 7, normalize $t^*=1$ (w.l.o.g.) and take the sequence of shocks equal to $\{\hat{\theta}_t\}_t=\{\hat{\theta}_1,0,\ldots,0\}$. ■

Appendix II

We first prove an auxiliary lemma that characterizes a robust Nash equilibrium in linear demands, given the value function in t. Then, we demonstrate the equivalence of robust subgame-perfect Nash and dynamic noncompetitive equilibria. In the proof, we use the following notation: $\mathcal{M}=\{\mathcal{M}^1,\ldots,\mathcal{M}^I\}$ is an array of I, $N\times N$ positive, semi-definite matrices, and $\mathbf{M}$ is the set of all arrays $\mathcal{M}$ (note that $\mathbf{M}\subset\mathbb{R}^{I\times N\times N}$); for any period t, function $\mathcal{F}_t:\mathbf{M}\to\mathbf{M}$ is defined as

$$\mathcal{F}_t(\mathcal{M})\equiv(1-\gamma)(\mathcal{H}^1_t(\mathcal{M}),\ldots,\mathcal{H}^I_t(\mathcal{M})), \tag{80}$$

where each component $\mathcal{H}^i_t:\mathbf{M}\to\mathbb{R}^{N\times N}$ is the harmonic mean of positive definite matrices $\mathcal{M}^j+\lambda_t^2\alpha^j\mathcal{V}$ for all traders but i,

$$\mathcal{H}^i_t(\mathcal{M})\equiv\mathcal{H}(\mathcal{M}^j+\lambda_t^2\alpha^j\mathcal{V}|j\neq i). \tag{81}$$

We prove Proposition 8, by extending the argument from Weretka (2007) to a dynamic setting. We first characterize a robust Nash equilibrium in period t, assuming that value function (41) holds in this period, and, hence, the parameter λ_t is given.

Lemma 1 (Characterization of Robust Nash Equilibrium in Demands) $\Delta\theta_t(\cdot)$ *is a robust Nash equilibrium in demands in t, if and only if, for each bidder i, the strategy* $\Delta\theta^i_t(\cdot)$ *is given by*

$$\Delta\theta^i_t(p)=(\bar{\mathcal{M}}^i_t+\lambda_t^2\alpha\mathcal{V})^{-1}(A-\lambda_t^2\alpha\mathcal{V}\bar{\theta}^i_0-p), \tag{82}$$

where $\bar{\mathcal{M}}^i_t$ *is the fixed point of the map* $\mathcal{F}_t$.

Proof. (Lemma 1: Characterization of Nash Equilibrium in Demands) Fix an arbitrary nondegenerate distribution of ε. Steps 1 and 2 derive the Nash best response for player i, $\Delta\theta^i_t(p)$.
Step 1. (Prices p^ε perfectly reveal the realization of ε.) Let $\Delta\theta^i_t(\cdot)$ be the Nash equilibrium strategy of trader i, $\Delta\theta^i_t(\cdot)$. For any realization ε, the equilibrium price vector p^ε assures market clearing and hence is

uniquely determined by

$$\varepsilon = \sum_{i \in I} \Delta\theta_t^i(p^\varepsilon). \tag{83}$$

By assumption, the Jacobians $D\Delta\theta_t^i$ are negative definite; demands are thus downward-sloping and, hence, invertible. The relation $\varepsilon(p^\varepsilon)$ is a linear bijection. Since the traders make choices contingent on p^ε, it follows that they will respond optimally to each price p^ε and, hence, for each realization of ε.

Step 2. (Derivation of the optimal demand function for trader i, $\Delta\theta_t^i(\cdot)$, given the realization ε.) Fix equilibrium strategies $\Delta\theta_t^j(\cdot)$ for bidders $j \neq i$, and a realization ε. Since the Jacobians $D\Delta\theta_t^j$ are negative definite, the market clearing condition

$$\Delta\theta_t^i + \sum_{j \neq i} \Delta\theta_t^j(p) + \varepsilon = 0 \tag{84}$$

implicitly defines a linear residual supply function faced by trader i, with a deterministic slope given by $\mathcal{M}_t^i \equiv -(\sum_{j \neq i} D\Delta\theta_t^i)^{-1}$ and a stochastic intercept that depends on the realization of ε. For any realization of ε, the marginal payment for an asset is equal to $p^\varepsilon + \mathcal{M}_t^i \Delta\theta_t^i$. The necessary and sufficient optimality condition for such a realization requires that the marginal utility coincide with the marginal payment, evaluated at the optimal trade,

$$A - \lambda_t^2 \alpha \mathcal{V}(\Delta\theta_t^i + \bar{\theta}_{t-1}^i) = p^\varepsilon + \mathcal{M}_t^i \Delta\theta_t^i. \tag{85}$$

Condition (85) can be solved for the optimal demand $\Delta\theta_t^i$ as a function of p^ε.

$$\Delta\theta_t^i(p^\varepsilon) = (\mathcal{M}_t^i + \lambda_t^2 \alpha \mathcal{V})^{-1}(A - \lambda_t^2 \alpha \mathcal{V} \theta_{t-1}^i - p^\varepsilon). \tag{86}$$

The demand function is a best response to a family of residual supplies with the slope $\mathcal{M}_t^i$ and a stochastic intercept. In a robust Nash equilibrium, trader i responds optimally to any realization of p^ε. It follows that the functional form of the best response of i, given the bidding strategies of others, is as in (82),

$$\Delta\theta_t^i(p) = (\mathcal{M}_t^i + \lambda_t^2 \alpha \mathcal{V} \bar{\theta}_t^i)^{-1}(A - \lambda_t^2 \alpha \mathcal{V} \bar{\theta}_{t-1}^i - p). \tag{87}$$

Step 3. (In a Nash equilibrium in demands, $\mathcal{M}_t$ is a fixed point of $\mathcal{F}_t$.) To prove the "only if" part, observe that, if all traders $j \neq i$ follow the optimal strategy (87), the slope of the residual supply for i is given by

$$\mathcal{M}_t^i \equiv -(\sum_{j \neq i} D\Delta\theta_t^j)^{-1} = (\sum_{j \neq i} (\mathcal{M}_t^j + \lambda_t^2 \alpha \mathcal{V})^{-1})^{-1} = (1-\gamma)\mathcal{H}_t^i(\mathcal{M}_t),$$

where, for the first equality, we used the definition of $\mathcal{M}_t^i$, and, for the second equality, we applied $D\Delta\theta_t^j = \left(\mathcal{M}_t^j + \lambda_t^2 \alpha \mathcal{V}\right)^{-1}$ and the optimality of (87) for all traders $j \neq i$. Since this equality holds for any i, $\mathcal{M}_t$ is a fixed point of $\mathcal{F}_t$ (and hence $\mathcal{M}_t = \bar{\mathcal{M}}_t$).

Step 4. (If $\mathcal{M}_t$ is a fixed point of $\mathcal{F}_t$, then the profile $\Delta\theta_t(\cdot)$ is a Nash equilibrium in demands.) For the "if" part, assume that all bidders $j \neq i$ follow (82) where $\bar{\mathcal{M}}_t$ is a fixed point of $\mathcal{F}_t$. We show that (82) defines a best response for player i. Since $\bar{\mathcal{M}}_t$ is a fixed point, $\bar{\mathcal{M}}_t^i$ satisfies

$$\bar{\mathcal{M}}_t^i = (1-\gamma)\mathcal{H}_t^i(\bar{\mathcal{M}}_t) = (\sum_{j\neq i}(\bar{\mathcal{M}}_t^j + \lambda_t^2\alpha\mathcal{V})^{-1})^{-1}. \tag{88}$$

If traders $j \neq i$ follow (82), the derivatives of their strategies w.r.t. p are equal to $D\Delta\theta_t^j = (\bar{\mathcal{M}}_t^j + \lambda_t^2\alpha\mathcal{V})^{-1}$, and, hence, (88) gives

$$\bar{\mathcal{M}}_t^i = -(\sum_{j\neq i} D\Delta\theta_t^j)^{-1}, \tag{89}$$

and $\bar{\mathcal{M}}_t^i$ is the deterministic slope of the residual supply faced by trader i, defined by the bids of other traders. Consequently, (82) is a best response to the bids of other traders. Since this is true for any i, $\Delta\theta_t(\cdot)$ defined by (82) is a robust Nash equilibrium in demands.

Step 5. (The Nash equilibrium is robust.) Since the derived Nash equilibrium strategies are independent of the distribution of $F(\varepsilon)$, they remain the best responses for any distribution of ε, in particular for $\varepsilon = 0$. Thus, the characterization holds for a robust Nash equilibrium. ■

Proof. (Proposition 8: Equivalence of Equilibrium Representation)

Step 1. (For a Nash equilibrium in demands $\Delta\theta_t(\cdot)$, the associated profile $(\tilde{p}_t, \Delta\tilde{\theta}_t, \tilde{\mathcal{M}}_t)$ defines a noncompetitive equilibrium in period t.) Assume that the value function (41) holds. Let $\Delta\theta_t(\cdot)$ be a robust Nash equilibrium in demands in t. By Lemma 1, individual bids $\Delta\theta_t^i(\cdot)$ can be represented as (82) where $\bar{\mathcal{M}}_t$ is a fixed point of $\mathcal{F}_t$. As shown in (89), the fixed point satisfies $\bar{\mathcal{M}}_t^i = -(\sum_{j\neq i} D\Delta\theta_t^j)^{-1}$, and hence coincides with the associated price impact $\tilde{\mathcal{M}}_t^i$ ($\bar{\mathcal{M}}_t^i = \tilde{\mathcal{M}}_t^i$ for any i).

To show that $(\tilde{p}_t, \Delta\tilde{\theta}_t, \tilde{\mathcal{M}}_t)$ defines a noncompetitive equilibrium, we need to verify three conditions, i.e., market clearing, optimization by all traders, and consistency of price impacts. For market clearing, observe that, by definition, $\Delta\tilde{\theta}^i \equiv \Delta\theta_t(\tilde{p}_t)$ and the price $\tilde{p}_t$ is defined as the market-clearing price $\sum_i \Delta\theta_t^i(\tilde{p}_t) = 0$, hence, $\sum_i \Delta\tilde{\theta}_t^i = 0$. Given the value function (41), the necessary and sufficient condition for the optimality of the associated trade $\Delta\tilde{\theta}_t^i$ at $\tilde{p}_t$ given the price impact $\tilde{\mathcal{M}}_t^i$ is the equality of the marginal payment and the marginal utility at $\Delta\tilde{\theta}_t^i$,

$$A - \lambda_t^2\alpha\mathcal{V}(\Delta\tilde{\theta}_t^i + \bar{\theta}_{t-1}^i) = \tilde{p}_t + \tilde{\mathcal{M}}_t^i\Delta\tilde{\theta}_t^i. \tag{90}$$

Given the equality $\bar{\mathcal{M}}_t^i = \tilde{\mathcal{M}}_t^i$ and the fact $\tilde{p}_t = \bar{p}^{\varepsilon=0}$, equality (90) follows from the Nash equilibrium bid function $\Delta\theta(\cdot)$. Since $\tilde{\mathcal{M}}_t = \bar{\mathcal{M}}_t$ is a fixed point of $\mathcal{F}_t$, by (48), it defines consistent price impacts.

Step 2. ($\Delta\bar{\theta}_t^i(\cdot, \bar{\mathcal{M}}_t^i)$, defined in (45), is a robust Nash equilibrium in demands in t.) The functional form of $\Delta\bar{\theta}_t^i(\cdot, \bar{\mathcal{M}}_t^i)$ is as in Lemma 1, and (48) implies that consistent price impact matrices $\bar{\mathcal{M}}_t$ define a fixed point of $\mathcal{F}_t$. Consequently, Lemma 1 implies that $\Delta\bar{\theta}_t^i(\cdot, \bar{\mathcal{M}}_t^i)$ is a robust Nash equilibrium in t. Observe that, by construction, the derived Nash equilibrium is associated with $(\bar{p}, \Delta\bar{\theta}, \bar{\mathcal{M}})$.

Step 3. (The value functions in robust Nash and noncompetitive equilibrium coincide.) In the last period, the value function corresponds to the mean-variance preferences in (1). If, for a degenerate distribution of ε, both equilibria are equivalent from $t+1$ to T, the value functions in t are the same, and they are given by (41) (see proof at the beginning of Appendix I). ■