Dynamic Thin Markets[1]

Marzena Rostek[2] and Marek Weretka[3]

August 11, 2009

Abstract. Extensive empirical research has shown that, in many markets, institutional investors have a significant impact on prices and mitigate its adverse effects through 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 the equilibrium properties of prices and arbitrage opportunities as well as dynamic trading strategies, compared to existing competitive or non-competitive models. Our predictions match a number of empirical facts that are hard to reconcile with prior modeling approaches.

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

Keywords: Thin Markets, Price Impact, Inventory Effects, Overshooting, Nash in Demands

1 Introduction

The assumption of price-taking behavior underlies many central results in asset pricing, including the no-arbitrage principle and full diversification of risk. Since trade-level data first became available two decades ago, it has become well understood that transactions by institutional investors exert an economically significant price impact in many financial markets.[4] Trading costs associated with price impact are not only significant but, in fact, dominate explicit trading costs, including commission fees, brokerage fees and 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 available in the Market Impact Models offered by Citigroup, EQ International, ITG, MCI Barra, and OptiMark, among others. In financial jargon, markets in which individual trades are large relative to average daily volume

[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, Bruno Strulovici, Elu Von Thaden, Jan Werner, and especially John Geanakoplos and Dimitri Vayanos for very helpful comments. We are also grateful for the suggestions offered by seminar participants at Chicago, Johns Hopkins, LSE (Mathematics), Mannheim (Finance), Northwestern, Oxford (Nuffield), Penn State, Princeton, Queen's, Southampton, UCL, Warwick (CRETA) and Wisconsin-Madison, and participants at the 2009 Minnesota Workshop in Macroeconomic Theory, the 2009 SAET Meeting at Ischia, the 2009 Caress-Cowles Conference at Yale, 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).

and, hence, impact prices, are known as *thin markets*. The goal of this paper is to understand the implications of market thinness for the equilibrium behavior of prices and trades, arbitrage opportunities, and asset valuation.

A voluminous body of research has emerged to explore why and how price impact affects individual portfolio choice and equilibrium in financial markets. The modeling approaches in the literature can be grouped around two theoretical mechanisms: asymmetric information[5] or inventory effects[6]. Our model belongs to the strand of research—reviewed in more detail below—that examines the traders' primitive decreasing marginal utility (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., are risk averse), selling or buying requires price concessions in order for other traders to be willing to absorb the investor's order. In this paper, market thinness does not result from asymmetric information, as we seek a source of price impact that would also be present in perfect foresight environments.

We take the view that the equilibrium properties of prices and trading strategies in a given market are determined by relatively few large, dynamically optimizing investors who continually monitor prices, are ready to provide liquidity for the market at any time and to take advantage of any price differentials. Trades by small competitive traders are seen as exogenous shocks in demand or supply, which we incorporate in the second part of the paper. The key feature of our model is that *all* liquidity providers are large relative to the market size. Thus, our market structure is one of bilateral oligopoly. We show that bilateral price impact fundamentally changes the properties of equilibrium compared to the competitive or existing non-competitive models with one-sided market power, in a way that allows us to explain a number of robust empirical phenomena observed in thin markets that are hard to reconcile with the previous approaches, and that arise naturally on the equilibrium path with price-making liquidity providers.

The idea of modeling financial markets as a bilaterally oligopolistic trading environment is not new. The market structure of bilateral oligopoly underlies the workhorse model of the finance literature based on the game of Nash in which traders submit demand schedules that are defined by limit and stop orders (e.g., the seminal work of Kyle (1989) and Vayanos (1999)).[7] Nevertheless, as such, the model is not well developed in that only a very special case with one good/asset and

[5] 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. (See Section 5.) 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, price impacts in downstairs markets do not differ significantly from those in upstairs markets, which are more transparent and less susceptible to informational asymmetries. This suggests that, in both types of markets, price impact is not mainly driven by the asymmetry of information.

[6] 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).

[7] In industrial organization, the game of Nash in demands (or supplies) was introduced for an oligopolistic industry by Grossman (1981) and further developed by Klemperer and Meyer (1989).

one-period market has been solved—with a notable exception by Vayanos (1999), who solved a stationary game. Rostek and Weretka (2008) solved an abstract, one-good dynamic model with bilateral market power. The modeling contribution of the present paper is building a non-stationary model with many assets—which thus allows study of asset pricing—that also permits analysis of demand and supply shocks. We provide a closed-form characterization of equilibrium trades, prices, and investor price impacts in dynamic trading environments with bilateral market power. While in this paper, equilibrium properties are derived in the context of asset pricing, they naturally extend to other dynamic bilaterally oligopolistic markets with multi-unit demands.

In order to delineate how the mere presence of bilateral price impact affects equilibrium, we consider the description of preferences and assets from 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 allow trade to potentially occur more frequently than dividend payments, which typically occur semiannually. In a competitive model without shocks and discounting, allowing for multiple trading opportunities would not have any effect on asset allocation at maturity and price behavior; in a non-competitive setting, new mechanisms arise. We next describe the main implications of market thinness.

Common practice among institutional investors involves breaking up orders into *blocks*, which are then traded sequentially. 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. If traded at once, a typical institutional package would represent over 60% of the average daily trading volume (Chan and Lakonishok (1995)). In the competitive model, Pareto efficiency after the first round of trade precludes order breakup as an equilibrium prediction. Equilibrium in TM-CAPM features order break-up as the optimal strategy for handling large orders in thin markets.

Another large body of evidence in finance concerns price behavior that is often interpreted as temporary departures of asset prices from their fundamental values. Such price behavior is observed as the reaction of prices in thin markets to one-time exogenous supply or demand shocks as well as to block trades.[8] A typical price pattern features a significant price change followed by a partial reversal of the price in subsequent periods. Thus, apart from permanent effect, the resulting price adjustment also has a temporary, overshooting component. These two effects were first empirically discovered by Kraus and Stoll (1972) and subsequently confirmed by numerous studies for various securities. Crucially, in the data, temporary price change occurs on the date of the shock, even if the shock was pre-announced.[9] The evidence on market reaction to supply and demand shocks is

[8]Shocks examined in the empirical literature include forced liquidations, issuance of new debt, selling Initial Public Offerings (IPOs), inclusions of new stocks into stock market indices, such as the S&P 500, or changes of index weights. Index funds invest a constant fraction of wealth in the companies included in an index, regardless of performance of an asset; therefore, a change of index weights induces a demand shock that is not associated with new information about the fundamental value of an asset. Rather, weight changes are bureaucratic decisions and data on ownership used for reweighting are publicly available prior to events.

[9]Indeed, pre-announced weight changes in stock market indices have a significant price effect on the day of

striking because it demonstrates that trade announcements and trade-induced price effects can be separated in time and that *anticipated* price changes can be observed in markets. In the standard competitive model, all of these features of price behavior are ruled out simply by no-arbitrage.

TM-CAPM predicts such temporary departures of equilibrium prices from the fundamental values as the equilibrium reaction of thin markets to shocks. We establish that *any* exogenous demand or supply shocks in thin markets have two effects on prices—*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 portfolio in the market. This effect would be observed also in markets with price-taking liquidity providers. The fundamental effect is amplified by a temporary liquidity effect, which results from the noncompetitiveness of both buying and selling traders. The permanent effect always occurs immediately after investors learn about the shock. Consistent with the data, the temporary effect attains its maximum only at the moment of trade, whether or not the shock is anticipated. Thus, in our model, overshooting is not a market friction, but rather an equilibrium phenomenon that is consistent with dynamic optimization and robust to arbitrage. Once one allows agents to place arbitrarily large orders—an implicit assumption in arbitrage—these orders then have price impact and limits to arbitrage arise endogenously.

We show that endogenously determined market depth (price impact) is not constant, but evolves over time, even in the absence of shocks or information revelation. Market depth changes because 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 investor willingness to absorb orders placed by other traders and translates into different price concessions. We demonstrate that the endogenous non-stationarity of price impact combined with the existence of the liquidity effect of shocks, generate a number of empirically documented phenomena. These include changes in price volatility that are unrelated to changes in fundamentals and volatility clustering. Additionally, by allowing for many assets, the model indicates that simultaneous shocks in different markets increase overall price volatility via cross-market price impact effects. Moreover, the model can rationalize the empirical evidence on the shape of the price impact—i.e., why the permanent price impact function is typically estimated as a linear function of block size, while its temporary counterpart appears to be a concave function of the size of a block.[10]

Our model has normative implications for asset valuation. Apprising assets in thin markets is challenging because the current market price of an asset does not reflect the cash value that could be obtained by liquidating a portfolio. To account for the adverse effect of market thinness on market value, valuation specialists apply the so-called blockage discount, which is recognized by

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).

[10] The following academic and practitioner 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).

the IRS since 1937. TM-CAPM can directly be used to derive blockage discounts, only heuristic formalizations of which are previously available, in an equilibrium asset pricing model. That the model allows for many risky assets is particularly important when determining blockage-discount formulas for the entire portfolio.

To explain further the bite of our approach, let us describe how it differs from existing models with price impact based on inventory effects. Within the inventory literature, incorporating price impact has been typically accomplished by building a Monopoly/Cournot-type model with $I \geq 1$ large investors trading with a fringe of price-taking traders. We argue that applying non-competitive models of the Cournot type to financial markets leads to the following difficulty. The Cournot model's market microstructure does not capture accurately how institutional trading occurs. Roughly speaking, the Cournot model assumes that demand for large strategic investors is defined by small, price-taking traders who provide them with liquidity, whereas a more accurate description of institutional trading is that liquidity providers are large and they provide liquidity for, and effectively trade with, one another. This is important because, as we show, the modeling assumption of who provides liquidity is key to explaining price and trading behaviors in thin markets. In particular, the equilibrium implications of an asset pricing model in which liquidity providers recognize their price impact can accommodate some of the empirically most robust features of thin markets that are associated with price impact on the equilibrium path.[11]

2 Set-up

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 also denotes the set of traders. Investment opportunities include N risky assets and one riskless asset (e.g., a bond). Investors can potentially 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 the 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 providers end trading period t with holdings of $\theta^i_t = \theta^i_{t-1} + \Delta\theta^i_t$, and $\theta^i_{b,t} = \theta^i_{b,t-1} + \Delta\theta^i_{b,t}$, with which they enter $t+1$. Trades $\Delta\theta^i_t$ and $\Delta\theta^i_{b,t}$ denote net demands in period t. $(\theta^i_0, \theta^i_{b,0})$ denotes the exogenously given initial portfolio of trader i. Investors choose their trades to maximize expected CARA utility functions. Using the standard argument, such assumptions are jointly equivalent to

[11] For detailed comparative analysis of the equilibrium implications of market structures with one-sided versus two-sided market power in perfect foresight models, see Rostek and Weretka (2008).

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,

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

To measure market size, it is useful to define *market participation rate* as an increasing function of the number of traders

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

In hindsight, the closer γ is to one, the more competitive are market interactions.[12] Let θ^{Av} denote the *average portfolio*, which is defined as the asset-by-asset average of the risky part of the initial holdings of all liquidity providers,

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

The fundamental value profile $v \in \mathbb{R}^N$ is defined as the average marginal utility from risky assets,

$$v \equiv A - \alpha \mathcal{V} \theta^{Av}. \tag{4}$$

It is straightforward to show that the fundamental values coincide with the vector of prices from the competitive CAPM. Throughout, a bar '$^-$' indicates equilibrium.

2.2 Equilibrium and Interpretation of Trading Behavior

We model spot markets as a Walrasian auction. In each period, traders submit downward-sloping (net) demand schedules $\Delta\theta_t^i(p)$, formed, e.g., by limit and stop orders. The market maker aggregates such bids and finds a price $\bar{p}_t$ that clears the market $\sum_{i \in I} \Delta\theta_t^i(\bar{p}_t) = 0$. In the spot market at t, investor i trades $\Delta\bar{\theta}_t^i \equiv \Delta\theta_t^i(\bar{p}_t)$ risky assets for $\bar{p} \cdot \Delta\bar{\theta}_t^i$ in terms of bonds. Similar to Kyle (1989) or Vayanos (1999), we study the Nash equilibrium in linear demands; we do not restrict the strategies to linear bids, but rather analyze equilibrium in which it is optimal for a trader to bid linearly given that others do. It is well known that there exists a continuum of equilibria (in terms of outcomes) in a deterministic Walrasian auction. As is standard in the literature, we focus on the unique linear equilibrium that is robust to perturbations in demands. We solve for a dynamic equilibrium using backward induction.

We suggest that the classical Walrasian auction, or the Nash demand game, admit an alternative and revealing interpretation. Consider a market for one risky asset. From the perspective of trader i, in each period t, the demand schedules of other traders and the market clearing condition define a linear residual supply faced by trader i with a deterministic slope $\mathcal{M}_t^i$. The slope of the residual supply against which investor i trades measures how a marginal change in quantity affects the price;

[12] With only two traders, γ is equal to zero, in which case equilibrium does not exist. The non-existence of equilibrium with two traders is also present in closely related models by Kyle (1989) and Klemperer and Meyer (1989) with a vertical demand.

in short, his *price impact*. It is revealing to rewrite the (robust) Nash equilibrium as a profile of demand schedules such that (i) given his assumed price impact $\mathcal{M}_t^i$, each trader optimizes (i.e., equalizes his marginal utility with his marginal revenue) for every price, and (ii) his assumed price impact is correct (i.e., it is equal to the slope of the residual supply that results from aggregating the bids of others). Let $V_t^i\left(\Delta\theta_t^i\right)$ be the value function at time t, which, in the last period T, coincides with utility function (1).

Lemma 1 (Characterization of Robust Nash Equilibrium) *A profile of demands* $\left\{\Delta\theta_t^i(\cdot)\right\}_{i\in I}$ *constitutes a symmetric robust Nash equilibrium at* t *if, and only if, for all* $i \in I$, *(i)* $D_{\Delta\theta_t^i}V_t^i(\cdot) = p + \mathcal{M}_t^i\Delta\theta_t^i$ *for all* p, *where* $\mathcal{M}_t^i$ *is such that (ii)* $\mathcal{M}_t^i = (1-\gamma)\left(D_p\Delta\theta_t^i(\cdot)\right)^{-1}$.

The characterization makes explicit that, in the model presented in this paper, the price impacts of each trader are not exogenous, but determined in equilibrium jointly with trades and prices. Strikingly, the only information any investor i needs in order to respond optimally to *all* prices—not just the equilibrium price—and to *arbitrary* orders of others—and not just the equilibrium orders—is, apart from his own preference, his price impact $\mathcal{M}_t^i$. In particular, no information about the number—let alone the utility functions, identities, or trading strategies of his trading partners—is required; nor does a trader need to know the equilibrium price in order to trade optimally.[13] Thus, our model fits particularly well anonymous markets in which each trader's price impact summarizes all the payoff-relevant information about the residual market against which he trades.

Writing the (robust) Nash demand equilibrium using Lemma (1) allows us to relate the non-competitive equilibrium in a Walrasian auction to the competitive equilibrium: As is apparent from condition (ii), in our model, traders are slope-takers rather than price-takers. They act assuming that they don't affect the slope of their residual supply by changing trading position; they might affect prices and they will as long as that slope is not zero. If the endogenously derived price impacts are equal to zero, the competitive equilibrium is obtained. For markets that can be described by interior parameter values, however, the predictions of TM-CAPM will differ from those in the competitive model. Price impact $\mathcal{M}_t^i$, which in our model is endogenously identical for all traders, serves as a measure of market thinness in period t.

Weretka (2006) develops a static general-equilibrium model with price impact and, using the strategic representation (1) from the present paper, Weretka (2009) shows outcome equivalence between Nash in demands and the general-equilibrium representation for the quadratic (CARA-Normal) setting. Thus, we can directly compare the results from TM-CAPM with the standard competitive theory of asset pricing.

The idea of an equilibrium with slope-takers fits well the description of institutional traders who use Market Impact Models and is consistent with practitioner views. Incidentally, apart from the models with price impact based on asymmetric information or inventory effects, which we review in the Introduction, several non-equilibrium models with price impact have been proposed, also by

[13] This property is particularly attractive in a one-period market; knowing one's preference in dynamic trading requires knowledge of one's value function.

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 endow traders with exogenously given price impact functions, the shape of which is motivated empirically and which are then used to analyze prices and allocations in thin markets. Our model derives such price impact functions endogenously, as part of equilibrium.

For many asset markets, 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 that results 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 not zero, the asset demands on which investor i operates are not perfectly elastic. We derive symmetric equilibrium in a closed form.[14] By Lemma 1, equilibrium in every period t is fully described by the profile of prices, trades and price impacts for all traders, $\left\{\left(\bar{p}_t, \Delta\bar{\theta}_t^i, \bar{\mathcal{M}}_t^i\right)\right\}_{i \in I}$.

3 Model Predictions

This section presents the model's predictions about trading strategies, prices and price impacts in the symmetric equilibrium of dynamic thin markets. We begin with a description of the basic mechanisms operating in a one-period thin market (Section 3.1). We then describe the main predictions for dynamic thin markets (Section 3.2). Throughout, we highlight the differences with the competitive benchmark. In the Appendix, we derive equilibrium for an arbitrary T.

3.1 Last Period

Consider the last trading period T. The equilibrium price impact of investor i is equal to

$$\bar{\mathcal{M}}_T^i = \frac{1-\gamma}{\gamma} \alpha \mathcal{V}. \tag{5}$$

TM-CAPM predicts that price impact is strictly positive and, in particular, competitive equilibrium does not satisfy equilibrium conditions. Given the finite number of traders ($\gamma < 1$) and strictly decreasing marginal utility ($\alpha > 0$), each trader faces an upward-sloping residual supply, and hence, his orders impact prices. Where does market power come from? When traders are risk averse and the shares are risky, purchasing or selling shares changes the traders' marginal utility. Therefore, investors require price concessions when trading. Thus, TM-CAPM predicts that the essential determinant of price impact of trader i is risk aversion (α) and return riskiness amplified or weakened by cross-market impact, as specified by the variance-covariance matrix ($\mathcal{V}$). What is less apparent in a symmetric solution is that it is other investors' risk aversion that enters directly into trader's i price impact; more risk averse trading partners are more reluctant to increase their holdings of risky assets, which implies larger price concessions in trading. Price impact strictly

[14] For one asset, asymmetric equilibria do not exist. We conjecture that this is also true for many assets, but do not offer a formal argument.

decreases in the number of liquidity providers, captured by γ. When the number of traders increases, the effect that the 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. Predictions approach the competitive outcome when investors are approximately risk neutral ($\alpha \sim 0$) or when the number of traders is large ($\gamma \sim 1$).

In response to their market power, each trader reduces his order relative to the competitive bid—for any given price, he buys or sells less,

$$\begin{aligned} \Delta\bar{\theta}^i_T(\cdot) &= (\mathcal{M}^i_T + \alpha\mathcal{V})^{-1}(A - \alpha\mathcal{V}\cdot\bar{\theta}^i_{T-1} - p_T) = \\ &= \gamma\underbrace{(\alpha\mathcal{V})^{-1}(A - \alpha\mathcal{V}\cdot\bar{\theta}^i_{T-1} - p_T)}_{\text{Competitive bid}}. \end{aligned} \tag{6}$$

Hence, compared to the competitive market, inverse demands become steeper by a factor of γ. The equilibrium asset prices coincide with the competitive prices

$$\bar{p}_T = v \equiv A - \alpha\mathcal{V}\theta^{Av}. \tag{7}$$

We postpone discussion of this result to Section 3.2.

In a perfectly competitive CAPM with symmetric bidders in equilibrium, each investor sells his initial holdings at T, $\bar{\theta}^i_{T-1}$, and replaces them with the average portfolio θ^{Av}. In a thin market, the trader sells a fraction γ of the portfolio with which he entered the trading round and replaces it with γ of the average portfolio

$$\Delta\bar{\theta}^i_T = \gamma(\theta^{Av} - \bar{\theta}^i_{T-1}). \tag{8}$$

Interestingly, how much of the initial holdings is rebalanced is determined solely by the market participation rate γ; in particular, it is independent of risk aversion α.

It is worth noting that multiple trading opportunities introduce a new source of price impact. Market power in a static market (or, in the last-period) results solely from the liquidity providers' risk aversion, which makes them willing to absorb the risky assets of other traders only at price concessions. The non-competitiveness of dynamic markets is, in turn, governed by a dynamic mechanism: Future market thinness begets present market thinness. Intuitively, the inability to diversify risk without price concessions at T exposes traders to risk at maturity and endogenously induces risk aversion at $T-1$ (and, by a recursive argument, in all previous periods). This can be seen by substituting the policy function in utility (1), which gives the indirect utility function of trader i at $T-1$ as a function of trade in this period. Crucially, the presence of period-$(T-1)$ trade $\Delta\theta^i_{T-1}$ in the final portfolio implies that the value function has a quadratic term (and is of the mean-variance form) and is strictly concave, which makes the investor effectively risk averse at $T-1$. The coefficient of effective risk aversion, $(1-\gamma)^2\alpha$, depends on the market participation rate γ, because only $1-\gamma$ of the $(T-1)$-period trade survives till maturity, while the remaining fraction γ is liquidated at T. The derivative of the value function at $T-1$ is equal to

$$D_{\Delta\theta^i_{T-1}} V^i_{T-1}(\cdot) = v - (1-\gamma)^2\, \alpha \mathcal{V} \left(\bar{\theta}^i_{T-2} + \Delta\bar{\theta}^i_{T-1} - \theta^{Av} \right). \tag{9}$$

Higher market competitiveness in the last period improves the hedging possibility in that period, which weakens the impact of trade $\Delta\theta^i_{T-1}$ on final holdings $\bar{\theta}^i_T$. Thus, it is the thinness of the future-period market that makes investors averse to absorbing risky shares in earlier periods and induces them to require price concessions in trading. On the other hand, with a trading opportunity in the future, liquidity providers are more willing to absorb risky assets from other traders, knowing that they will be able to partially diversify their positions in subsequent rounds. As a result, in earlier trading rounds, the value function is less concave, the required price concessions are smaller, and so is the price impact. Since, in response to their market power at $T-1$, traders reduce their orders in that period, the outcome is not Pareto efficient, and it follows that there are gains to trade at T. Hence, investors trade in both periods.

3.2 Dynamic Thin Markets

We now characterize the equilibrium in a market in which investors can potentially trade for T periods after which assets mature. There are no shocks throughout.

Optimal trading strategies. Proposition 1 characterizes how market thinness affects the optimal execution of trade. In the competitive CAPM, investors instantaneously sell their initial inventories and rebalance their holdings within one period; 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, $\Delta\bar{\theta}^i_t = \gamma(\theta^{Av} - \bar{\theta}^i_{t-1})$, and place them on the market sequentially.[15]

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 fraction of (the remaining part of) their initial portfolios to invest between the average portfolio and the riskless asset.

Proposition 1 (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 portfolios, θ^i_0 and θ^{Av}, and the weight assigned to θ^{Av} monotonically increases over time:*

$$\bar{\theta}^i_t = (1-\gamma)^t \theta^i_0 + (1 - (1-\gamma)^t)\theta^{Av}. \tag{10}$$

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

[15] The number of blocks in the optimal trading strategy corresponds to the number of trading opportunities T. That number can be endogenized by introducing fixed transaction costs.

Order break-up is a common practice among large investors. Table 1 presents typical figures from the NYSE. Only 20% of the total volume of all institutional purchases and sales is completed within one day, while more than 30% of the orders takes at least six 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%

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))

An average package in the smallest firms amounts to two or three times the daily volume of trade, and even in the largest firms, an average package takes up 25% of the daily volume.

Partial Diversification of Risk. When the number of trading periods is sufficiently large, the portfolios in TM-CAPM converge to the competitive holdings, but for any finite number of periods, the portfolios are distinct. Consequently, at any point in time, idiosyncratic risk is not perfectly hedged. On the other hand, 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 interactions less competitive and reducing the trade. In a quadratic symmetric model, these 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.

Non-stationarity of Price Impact. The noncompetitive CAPM predicts that price impact is not constant across the trading periods, but instead increases as 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 (35) in the Appendix), which becomes more and more concave over time to reflect the traders' increasing effective risk aversion, and their decreasing willingness to buy risky assets at given price concessions. To the extent that market depth can proxy the level of market competitiveness, according to TM-CAPM, markets are least competitive just prior to maturity.

Proposition 2 (Time Structure of Price Impact) *In TM-CAPM, price impact exponentially decreases with time-to-maturity $T-t$,*

$$\bar{\mathcal{M}}_t^i = \frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\alpha\mathcal{V}. \tag{11}$$

Notably, 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, characterization (11) shows that there are cross-market price-impact effects when the payoffs of two stocks are correlated. Then, the sale of one asset inflicts 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 determinants of the concavity of investors' preferences in TM-CAPM on price impact.

One implication of the non-stationarity of price impact is that liquidity is correlated across assets, even if their returns are independent.[16] 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)).[17] The presence of non-stationarity, *per se*, in the model is not surprising given the finite horizon and end-time effect. That the market depth is non-stationary has, however, striking implications for how thin markets react to exogenous shocks in supply or demand, which we discuss in Section 4.3.

The derived monotone time structure of price impact hints at a new channel through which systemic risk might destabilize markets, as it suggests that market depth can endogenously become higher prior to news announcements.[18] Similarly, Proposition 2 implies that, in thin markets, asset maturity might become an active instrument in stabilizing markets; 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.

[16] 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.

[17] It might seem that, apart from predicting commonality in liquidity, Propositions 1 and 2 imply that trade volume and liquidity (measured by the inverse of price impact) should be positively correlated in time series data. Yet, although empirical support for the cross-sectional relation of liquidity across markets is strong, the evidence on a dynamic relation is mixed (see, e.g., Johnson (2008)). 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 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.

[18] In our model, at the end of period T assets mature, which can be viewed as full resolution of uncertainty. Our analysis implies that if the disclosure of information is introduced to the model, T would alternatively be interpreted as a period at the end of which partial information about dividends is revealed. Our predictions are consistent with the extensively documented fact that liquidity tends to increase prior to scheduled news announcements and decrease on the announcement day as uncertainty is resolved by market participants.

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 significantly harder in the noncompetitive model, as asset prices no longer coincide with the marginal utilities of traders, and, moreover, marginal utilities typically differ across agents.

Nevertheless, under our assumptions, equilibrium asset prices coincide with the fundamental values of assets and are, therefore, identical to the competitive prices in every period. The price result might be somewhat surprising, but it arises from the following mechanism. With symmetric price impacts, market power of buyers and sellers is balanced, and buyers and sellers reduce their demands and supplies for each asset by the same factor γ. Consequently, thin markets clear at the competitive prices, even though the trades are not competitive. The price result, however, relies on the joint assumption of: (i) CARA-Normal (quadratic) and (ii) homogenous utility functions, and (iii) the deterministic structure of the model. Moving away from these assumptions would introduce price effects.

One implication of the price result is 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 return and the return on the average portfolio. Let R^{Av} denote the expected return of an average portfolio, 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 3 (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,*

$$R_n - R = \beta_n(R^{Av} - R). \tag{12}$$

We should stress here that, in our model, the average portfolio is defined as a *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 CAPM predictions, based on the whole market portfolio, should not be applied in this instance. To empirically test for the Security Market Line in TM-CAPM, one should first properly identify a thin market—a group of institutional investors who trade a given collection of risky assets.

One lesson from Proposition 3 is that price impact, *per se*, does not necessarily distort asset returns and in order to explain liquidity premia within CAPM, one needs other distortions (on that, see Section 6).

4 Price Effects in Thin Markets

TM-CAPM suggests that, in the absence of shocks and with a 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 competitive markets to exogenous shocks in asset supply and demand, or when investors need to quickly liquidate their portfolios. We show that accounting for market thinness may explain a number of empirical phenomena that are hard to reconcile within the competitive models; these include: asset price overshooting (Section 4.1); the existence of instruments for asset valuation that account for price impact (Section 4.2); and stylized facts about return volatility (Section 4.3). Typical supply or demand shocks that have been examined in the empirical finance literature include forced liquidation, issuance of new debt, selling Initial Public Offerings (IPOs), an inclusion of an asset into the S&P Index, or a change of index weights. Alternatively, exogenous shocks can 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 an unanticipated, as well as an anticipated, exogenous sale of a large block of shares by a trader other than I liquidity providers.

4.1 Fundamental and Liquidity Effects

As evidenced from the empirical literature, the exogenous shocks in asset supply or demand result in price response that is commonly interpreted as temporary departures of price from the fundamental value, often referred to as "mispricing" or "asset price *overshooting*." In the data, an unanticipated supply shock results in an immediate and significant price drop followed by a partial reversal of the price change in subsequent periods. Notably, even if the shock is pre-announced, the temporary price drop below the long-run level occurs on the actual event date and not on the date of the announcement, and attains the long-run value only in subsequent periods. This phenomenon cannot occur in the standard competitive model. 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. What should be observed, instead, is that the price adjusts to the new fundamental value immediately following the shock announcement and remains there until maturity. The central difficulty is that the presence of price-taking liquidity providers—traders who are ready to respond to price differentials at any time—ties the equilibrium price to the fundamental value. The overshooting price pattern was first empirically documented by Kraus and Stoll (1972) and subsequently confirmed by numerous 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); Newman and Rierson (2004); and Greenwood (2005)).

Our model predicts such temporary departures of prices from the fundamental values as the equilibrium reaction of thin markets to shocks. We show that, in a thin market, the price change resulting from any exogenous supply or demand shock has two effects on prices: a *fundamental effect*, which is permanent, and a *liquidity effect*, which is temporary. Proposition 4 demonstrates that price overshooting is a direct consequence of the existence of the liquidity effect of shocks. As we proceed to explain, these 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. (Section 5 provides analysis for anticipated shocks.) In period t^*, a portfolio $\hat{\theta} > 0$ is being liquidated along with the trade by liquidity providers. The fundamental effect represents the change in the fundamental value. The supply shock permanently increases the average holdings of risky assets by liquidity providers to $\theta^{Av*} \equiv \theta^{Av} + \hat{\theta}/I$, which, given investors' decreasing marginal utility, lowers the average marginal valuation to $v^* \equiv A - \alpha\mathcal{V}\theta^{Av*}$ and, thus, the fundamental value drops. Since all liquidity providers learn about the shock in period t^*, the post-shock fundamental value changes at t^* by

$$\Delta^F \equiv -\alpha\mathcal{V}\frac{\hat{\theta}}{I}. \tag{13}$$

The fundamental effect would also be observed in a model with price-taking liquidity providers, so 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 traders demand price concessions beyond the drop in the fundamental value? On the equilibrium path, with or without the shock, in any period t, each trader equalizes his period-t marginal utility and his marginal revenue. This also holds in averages,

$$v^* = \frac{1}{I}\sum_{i\in I} D_{\Delta\theta_t^i} V_t^i(\cdot) = \bar{p}_t + \bar{\mathcal{M}}_t^i \frac{1}{I}\sum_{i\in I}\Delta\theta_t^i. \tag{14}$$

Without the shock, the net trade $(1/I)\sum_{i\in I}\Delta\theta_t^i$ is equal to zero by market clearing, and the price equals the fundamental value, as explained above in the model predictions. 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 $\bar{\mathcal{M}}_{t^*}^i\hat{\theta}/I > 0$. It then follows from optimality that the price is below the average marginal utility, $\bar{p}_{t^*} = v - \bar{\mathcal{M}}_t^i\hat{\theta}/I$. The liquidity effect of the equilibrium price is equal to

$$\Delta^L \equiv -\frac{(1-\gamma)^{2(T-t^*)+1}}{\gamma}\alpha\mathcal{V}\frac{\hat{\theta}}{I}. \tag{15}$$

Proposition 4 describes the price behavior on the equilibrium path in response to a supply shock.

Proposition 4 (Asset Price Overshooting) *Following an unanticipated liquidity shock, $\hat{\theta}$, in period t^*, equilibrium prices adjust by $\Delta^F + \Delta^L$. In period $t^* + 1$, prices revert by Δ^L to their post-shock fundamental value v^* and remain at this level in all subsequent periods.*

Why does the liquidity effect not persist as does the fundamental effect? Since no other shocks occur after t^*, in all periods following t^*, the net trade of liquidity providers becomes equal to zero by market clearing, and the price attains the new fundamental value $\bar{p}_t = v^*$.

The analysis reveals another insight from TM-CAPM; namely that, by contrast to competitive markets, thin markets react differently to endowment shocks compared to shocks in demand or supply. Specifically, if the assets $\hat{\theta}$ were an increase in the endowments of the liquidity providers, then only a fundamental effect but no liquidity effect would be observed.

The overshooting effect in thin markets can be conceptualized using the notion of market demand. Empirical studies on the reaction of markets to shocks make a distinction between *short-* and *long-run market demands.* The short-run (inverse) demand corresponds to period-t^* price reaction to trade in period t^*, while the long-run demand specifies the price for period-t^* trade after all the price adjustments occur. Our model provides a microfoundation for such a representation of thin markets. In TM-CAPM, the long-run demand for $\hat{\theta}$ is given by a horizontal sum of marginal utilities of the liquidity providers, $A - \alpha\mathcal{V}\left(\theta^{Av} + \hat{\theta}\right)$. The short-run demand corresponds to the horizontal sum of bid schedules submitted by liquidity providers in period t^*, the inverse of $(1/I)\sum_{i\in I}\Delta\bar{\theta}^i_{t^*}(p)$. Due to order reduction, the short-run market demand is steeper than the long-run demand and, because of endogenously varying market thinness, its slope changes over time. The difference between the short-and long-run demand results in the liquidity effect. (See Figure 2.) On the day of the shock, the price moves along the short-run demand, but after the period in which the shock occurs ends, only the fundamental effect is observed. Lastly, note that while the long-run market demand depends solely on the primitive preferences of liquidity providers, the short-run demand is also shaped by time-to-maturity and, as explained in Section 5, on whether or not trade $\hat{\theta}$ is anticipated.

Whereas the magnitude of the fundamental effect Δ^F does not depend on the timing of the shock t^*, due to the non-stationarity of price impact, the magnitude of overshooting Δ^L is greater when time to maturity is shorter.

As observed in the data, TM-CAPM predicts that when asset returns are correlated, overshooting in one market spills over to other markets along with the fundamental effect. Apart from the permanent adjustment of the fundamental value in substitute or complement asset markets,

$$\theta \cdot \Delta^F = \frac{\alpha}{I} Cov(A\hat{\theta}, A\theta), \tag{16}$$

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

$$\theta \cdot \Delta^L = \frac{(1-\gamma)^{2(T-t^*)+1}}{\gamma} \frac{\alpha}{I} Cov(A\hat{\theta}, A\theta). \tag{17}$$

All of these features of price behavior have been documented in the data. An immediate reversal of the price change on the trade subsequent to a large transaction is a major finding in event studies (e.g., Holthausen, Leftwich and Mayers (1990)). The predictions are also strongly supported by methodology recently implemented by Citigroup to estimate price impact.[19] Figure 3 depicts both

[19] In this program, 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 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)).

effects for an exogenous shock $\hat{\theta}$ in period t^*. Panel A shows the path for the trade of an asset, while Panel B depicts the price of an asset $\bar{p}_t$.

The influential paper by Brunnermeier and Pedersen (2005) explained price overshooting in a Cournot-based model by "predatory trading". When an investor 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 traders who define a downward-sloping demand, buying assets when they are expensive and selling when assets are cheap. These traders, by assumption, do not take advantage of 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. Our explanation of overshooting is complementary in that predatory trading does not occur in TM-CAPM, since all traders maximize their preferences. In addition, while predatory trading can rationalize price overshooting as a response to unanticipated shocks, our model can also explain delayed overshooting.

Limits to Arbitrage. When traders' individual orders can influence prices, an important question becomes how price impact affects the arbitrage possibilities in a market. Indeed, the key to understanding why the liquidity effect occurs concerns arbitrage. Since the price path is deterministic and 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, as they would in the competitive models? We now show that another qualitative change in thin markets involves endogenously arising limits to arbitrage.

Suppose, for the sake of simplicity, that investors' holdings of risky assets are fully diversified so that their only trade is from shock absorption. Suppose that in period t^* when the shock occurs, an investor increases his trade by ε and sells the same amount in the next period. Such trade would have a first-order benefit equal to ε times the price differential induced by the shock, $\varepsilon \times \bar{\mathcal{M}}^i_{t^*}\hat{\theta}/I$. At the same time, however, the additional demand created by arbitrage increases the price in t^* by $\bar{\mathcal{M}}^i_{t^*} \times \varepsilon$, which adversely affects the terms of trade of $\hat{\theta}/I$. Hence, on the equilibrium path, the marginal benefit from arbitrage is exactly offset by the marginal externality of the price increase on the remaining units being traded.

An implicit assumption maintained so far is that trade takes place among a fixed number of liquidity providers. 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. To complete the argument behind the coexistence of anticipated price differentials in equilibrium and limits to arbitrage in thin markets, we need to consider potential entrants. Relative to the liquidity providers, the additional difficulty is that, for potential entrants, arbitrage has no externality cost on their existing trades. Nonetheless, by contrast to the competitive markets, in thin markets, potential profits from entering the market are not infinite, but bounded due to price impact. This then limits the benefits from arbitrage and reduces incentives to enter the market. To see this, suppose that, at t^*, an (outside) entrant purchases a block of assets to be sold in the next period. Taking an unbounded position at t^*, or buying a few more shares than the amount of the shock $\hat{\theta}$ results in a strictly negative profit, as the purchase drives the price above the fundamental value in this period. Similarly, selling the shares in period t^*+1 also has an adverse effect on the price, which is 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 potential profits from entry, however, unlike in a competitive model, fixed entry costs might prevent outsiders from arbitraging the price overshooting. In practice, entry costs include explicit trading costs, such as transactions costs, but also costs associated with learning and monitoring the characteristics of particular stocks. Mitchell, Pedersen and Pulvino (2007) document that it may take months for outside capital to bid prices back to fundamental values.[21] This slow entry is attributed by the authors to information barriers and the costs of maintaining dormant financial and human capital in a state of readiness when arbitrage opportunities arise.

Once one acknowledges that arbitrageurs who can place large orders have price impact, limits to arbitrage naturally arise. Note finally that the argument behind no-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, profits from arbitrage are bounded for any round-trip trade.

Price Manipulation. Given that traders can affect prices, another critical question is whether investors have incentive to manipulate prices. More precisely, so far, we have examined whether it is profitable for an investor to arbitrage the price differential created by an exogenous liquidity shock. We now ask whether an investor would choose to destabilize the market and generate such shocks himself 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 entire sequence of the shock, and thus, the price path, such asymmetry in information might lead to a positive profit from price manipulation.

Price manipulation can be formalized 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 5 establishes that, even though it is possible for investors to affect prices in thin markets, doing so can never be profitable.

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

Proposition 5 (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* $\left\{\hat{\theta}_t\right\}_t$*.*

The key feature of the model underlying the robustness of thin markets to price manipulation is the 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.

Taking a view of a single agent trading against a market demand, Huberman and Stanzl (2004) also decompose an (exogenous) price impact into the permanent and temporary component to identify conditions on the two price impact functions under which price manipulation is not feasible. Proposition 5 provides an equilibrium-based alternative to their argument; our setting is less general in that the permanent price impact is constant over time (albeit endogenously) and is more general in that it allows for an arbitrary number of assets. Our model offers justification for time-independence of the permanent effect as resulting from price changes along the long-run aggregate demand.

4.2 Market Value and Blockage Discount

When markets are thin, assessing the value of a portfolio is nontrivial. In a perfectly competitive market, the cash value of a block of shares, $\hat{\theta}$, is simply equal to the corresponding prices currently observed on the market times the quantity of shares, $\bar{p} \cdot \hat{\theta}$. When markets are thin, selling a large block of shares exerts 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 apprising assets traded in thin markets has been recognized by valuation specialists, who apply an instrument called *blockage discount*. 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).[22] 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.[23] According to Federal Tax Regulations, the burden of demonstrating that a blockage discount is justified lies on the taxpayer. Yet, there are no equilibrium-based guidelines about 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

[22] 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.

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

(e.g., Estabrook (1999, 2001)), and these methods have been adopted in appraisal businesses and valuation consulting.

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 the case of a divorce. It is in the interest of 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 reasoning corresponds to how price impact is modeled in TM-CAPM. Our results can directly be applied to formally address asset valuation in thin markets and, thus, derive blockage discounts. 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 corresponding 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 4. Consequently, the blockage discount becomes

$$BD = \underbrace{\frac{\alpha}{I} Var(A \cdot \hat{\theta})}_{\hat{\theta} \cdot \Delta^F} + \underbrace{\frac{(1-\gamma)^{2(T-t^*)+1}}{\gamma} \frac{\alpha}{I} Var(A \cdot \hat{\theta})}_{\hat{\theta} \cdot \Delta^L}. \tag{18}$$

In the derivation of formula (18), 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 those of the considered block. In practice, traders break up large amounts into smaller blocks and sell them over time to mitigate the adverse effects that result from market thinness. Therefore, formula (18) is likely to overestimate the value of a blockage discount and should be interpreted as the upper bound on the discount. The lower bound for the blockage discount is the fundamental effect as this effect is present even if the trade is spread over time. Additionally, if a trader has other assets that are *not* included in $\hat{\theta}$, and whose payoffs are positively (negatively) correlated with the liquidated portfolio, then liquidation also affects the values of these assets via cross-market liquidity and permanent effects. The blockage discount (18) should then be adjusted upwards (downwards) accordingly, applying (16) and (17).

4.3 Price Volatility

This section demonstrates that two implications of market thinness, the existence of the liquidity effect of shocks combined with the endogenous behavior of price impact, shed light on empirical findings about 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 the unconditional distribution of asset returns has heavy tails.

To study volatility, suppose that instead of a once-and-for-all shock $\hat{\theta}$, the market features i.i.d. mean-zero shocks in asset supply $\hat{\theta}_t$ (generated, for example, by noise traders), which we treat as unanticipated by the traders.[24] Price volatility depends on two independent factors: the market participation rate and whether or not traders account for 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 price volatility above of the volatility of the fundamental value. Thus, excess volatility is a direct consequence of bilateral market power in every trading period. In addition, overall price volatility is further magnified by cross-market effects: As long as 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. 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 thin markets.

Because the equilibrium price impact varies over time, price volatility changes independently of changes in fundamentals. If, in addition, the monotone time structure of the equilibrium price impact is taken into account, high (low) price volatility is clustered at the end (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 three.

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 4.1, 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: 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. Unlike the sequences of prices taken 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 the data: two trading rounds correspond to the shortest time for which prices are martingales.

[24] Proper formulation of the problem would require agents to take randomness of 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 and demand shocks on prices.

4.4 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 block size. It is now standard to distinguish between permanent and temporary price impact functions, which are based on long-run and short-run price changes, respectively. In the data, permanent impact function is typically estimated as a linear function of blocks (e.g., Almgren et al. (2005)), whereas temporary price impact appears to be a concave function of 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)).

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 block size (cf. (13) and (15)). Much of the empirical evidence, however, has been established assuming that price impact is *time-independent*. In TM-CAPM, the permanent price effect is stationary—the fundamental price effect depends on the magnitude of shock realizations, but shocks per share are constant across all trading periods. If, as predicted by TM-CAPM, price impact indeed evolves over the trading horizon and large blocks are liquidated when price impact is small, while small blocks 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 make this mechanism more precise in TM-CAPM.

In order to capture the negative correlation between block size and price impact, we depart from the i.i.d. assumption by considering liquidation shocks $\hat{\theta}_t$—a less extreme (and more realistic) version of the stochastic noise trade from Section 4.3—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

$$u_t\left(\hat{\theta},\hat{\theta}_b\right)=\hat{\theta}_{b,t}+A\cdot\left(\theta^o_t+\hat{\theta}_t\right)-\alpha(\theta^o_t+\hat{\theta}_t)\cdot\mathcal{V}(\theta^o_t+\hat{\theta}_t), \tag{19}$$

where θ^o_t 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 normalize their number to one in any trading period without loss of generality. The optimal trade of the occasional trader is equal to

$$\hat{\theta}_t(\cdot)=(\bar{\mathcal{M}}^o_t+\alpha\mathcal{V})^{-1}(A-\bar{p}-\alpha\mathcal{V}\theta^o_t), \tag{20}$$

where $\bar{\mathcal{M}}^o_t$ stands for the price impact of the occasional trader in t.[25] Although stochastic, the

[25] The price impact induced by the occasional investor's trade need not coincide with the price impact of liquidity providers. The occasional trader faces I (rather than $I-1$) liquidity providers, which lowers his price impact. On

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

Given that θ_t^o is i.i.d. with $E(\theta_t^o) = \theta^{Av}$, the expected unconditional price is constant and equal to the fundamental value, and block size $\hat{\theta}_t$ is negatively correlated with price impact $\bar{\mathcal{M}}_t^o$. Consequently, large blocks, which tend to be observed in earlier trading periods when price impact is smaller, typically have a weaker temporary price impact per share than small blocks—observed toward 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.

5 Anticipated Shocks

So far, we have considered the supply shocks that were announced on the day of the shock. A large body of research documents price behavior for shocks that are announced long before they actually occur, such as changes of the weights in a stock market index and inclusions of a new stock into an index. The evidence shows that such pre-announced shocks have price effects not only on the day of the announcement, but also on the actual day of the shock; that the severity of the effect is enhanced by asset riskiness and correlation; and that price is additionally depressed (increased) between the day of the announcement and the shock day when the transitory effect peaks.[26] Thus, the observed price effects cannot be attributed to any revelation of information about the fundamental value, which should be incorporated on the day of the announcement. None of these effects are observed in a competitive model—pre-announcing the shock does not alter the price adjustment, which amounts to a change of the fundamental value at the announcement. In this section, we study whether, and if so, how price reaction to shocks in TM-CAPM is 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 Shock Announcement and Price Effect. To examine the effect of pre-announcing a shock on price adjustment, suppose that, in period t^*, liquidity providers learn that an extra supply of assets $\hat{\theta}$ will be available in period t^{**}. Since the fundamental effect Δ^F, given by (13), is defined with respect to the expectation in t about the average holdings at the end of period T, the effect occurs at the moment of the announcement *and not* at the moment when the shock is realized. The fundamental value instantaneously adjusts to its post-shock level v^* in the announcement period t^* and remains there until T, as no new information about the shock is revealed. The additional liquidity effect of shocks Δ^L in TM-CAPM, given by (15) evaluated at t^{**}, takes place only in the actual period of the shock, t^{**}, whether or not the shock is pre-announced. This holds because it is in that period that the net trade is positive and during which traders need

the other hand, by changing $\hat{\theta}^{Av}$, the trader also induces a fundamental effect, which increases his price impact.

[26] These effects were documented, for example, by Newman and Rierson (2004). In particular, the study found that new bond issuance in the European telecommunication sector increased the yield spreads of other firms in the sector. The effect was transitory, significant, and peaked on the day of issuance, not on the announcement day.

to absorb it (cf. (14)). The liquidity effect is not driven by information disclosure, but rather by the effect that absorption of the extra assets has on the average marginal payment. Pre-announcing a shock in a thin market introduces a time separation of fundamental and liquidity effects.

Another new feature of price behavior for pre-announced shocks is that, in addition to fundamental and liquidity price effects, we observe a 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 liquidity providers anticipate the drop in price in t^{**}, which lessens their willingness to buy the assets in all periods prior to t^{**}. The formalization of these three effects is a special case of Proposition 6. The path of price response to an announced once-and-for-all shock is depicted in Figure 4.

Multiple Anticipated Blocks. In practice, portfolios are often liquidated in blocks. In Section 4.3, 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 a 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 and 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. By contrast, in thin markets, 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 6 (Anticipated Multiple Blocks) *Consider a sequence of anticipated sales* $\left\{\hat{\theta}_t\right\}_t$. *In any trading period* t*, price is equal to*

$$\bar{p}_t = v^* + \Delta_t^L + \gamma \sum_{l=1}^{T-t} \Delta_{t+l}^L, \tag{21}$$

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

$$\Delta_t^L = \frac{(1-\gamma)^{2(T-t)+1}}{\gamma} \alpha\mathcal{V}\hat{\theta}_t/I. \tag{22}$$

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. While the current liquidity effect in (21) results directly from non-competitive trading, as explained in Section 4.1, the cumulative effect reflects the resulting anticipation of depressed prices in the future, which lowers the price by weakening incentives to buy and strengthening incentives to sell today. The cumulative effect generalizes the third effect discussed above in the context of a single anticipated shock. 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 results from 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. Using 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 liquidity providers have no incentive to arbitrage.

In sum, just as when the sequence of sales is not anticipated, 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. Nevertheless, the price path in a thin market is sensitive to how the portfolio is partitioned. This occurs because future sales depress price during the whole period between the announcement and the shock, and the effects on prices are cumulative. As a result, if sales are anticipated (and credible), the liquidator can be expected to concentrate most of the trade in the first period.

Proposition 6 (equation (21)) 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. Further, ability to spread the current liquidity effect is reduced, as the anticipation of future liquidity effects adversely affects 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 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.

6 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, empirical evidence on the return volatility, the shape of permanent and temporary price impact functions, the existence of valuation instruments, such as blockage discount, etc. The empirical literature on illiquid markets has demonstrated that expected asset returns are higher for illiquid assets, as 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 (2009) 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 traders are price-takers, introducing the liquidity concern has no effects on prices and allocations, as the cash value from liquidation coincides with the value at maturity. It is the interaction of price impact and 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 price impact.

Appendix

Proof. (LEMMA 1: CHARACTERIZATION OF ROBUST NASH EQUILIBRIUM) Symmetry of robust Nash equilibrium is not required in the lemma and the proof we provide does not assume it.

("if") Let $\left\{\Delta\theta_t^j(\cdot)\right\}_{j\neq i}$ be the equilibrium strategies of traders other than i. With additive perturbations (e.g., exogenous noise) in demand, the residual supply faced by trader i has a stochastic intercept and a deterministic slope, given by $\bar{\mathcal{M}}_t^i = (1-\gamma)\mathcal{H}\left(\left(D_p\Delta\theta_t^j(\cdot)\right)^{-1}|j\neq i\right)$, where $\mathcal{H}(\cdot)$ is the harmonic average operator. (In a symmetric equilibrium, bid slopes $D_p\Delta\theta_t^j(\cdot)$ are the same for all traders and, hence, the slope of the residual supply of trader i is given by $\bar{\mathcal{M}}_t^i = (1-\gamma)\left(D_p\Delta\theta_t^i(\cdot)\right)^{-1}$.) By condition (i), $\Delta\theta_t^i(\cdot)$ equalizes the t-period marginal utility (marginal value function) with marginal revenue for any possible realization of the residual supply (or noise); it is thus a best response to $\left\{\Delta\theta_t^j(\cdot)\right\}_{j\neq i}$. Since this is true for any i, the profile of demands $\left\{\Delta\theta_t^i(\cdot)\right\}_{i\in I}$ is a Nash equilibrium with an arbitrary additive noise and, hence, it is a robust Nash.

("only if") Let $\left\{\Delta\theta_t^i(\cdot)\right\}_{i\in I}$ be demand functions such that the conditions (i) and (ii) in the Lemma are not satisfied. By the linearity of demands, for almost all $\bar{p}$, $D_{\Delta\theta_t^i}V_t^i(\cdot)\neq\bar{p}+\mathcal{M}_t^i\Delta\theta_t^i$, where $\mathcal{M}_t^i$ is the slope of the residual supply. With any additive noise that puts positive mass on prices for which $D_{\Delta\theta_t^i}V_t^i(\cdot)\neq\bar{p}+\mathcal{M}_t^i\Delta\theta_t^i$, the bid that equalizes marginal utility with marginal revenue, $D_{\Delta\theta_t^i}V_t^i(\cdot)=\bar{p}+\mathcal{M}_t^i\Delta\theta_t^i$, gives a strictly higher expected utility. It follows that $\Delta\theta_t^i(\cdot)$ is not a best response and that noise exists for which $\left\{\Delta\theta_t^i(\cdot)\right\}_{i\in I}$ is not a Nash equilibrium. Therefore, $\left\{\Delta\theta_t^i(\cdot)\right\}_{i\in I}$ is not a robust Nash equilibrium. ■

Since many of the results in the paper can be derived from formulas of the general model with a sequence of anticipated shocks, we first derive equilibrium in the general model (Proposition 6).

Proof. (PROPOSITION 6: ANTICIPATED MULTIPLE BLOCKS) Recall that $\theta_t^{Av}=(1/I)\sum_{i\in I}\bar{\theta}_{t-1}^i$ is the average portfolio at the beginning of trading period t and v^* is the average marginal utility (i.e., the fundamental value) after the last period of trade, $v^*=A-\alpha\mathcal{V}\theta_{T+1}^{Av}$. Let $(\hat{\theta}_1,...,\hat{\theta}_T)$ be a sequence of anticipated shocks. We show that the equilibrium price is given by

$$\bar{p}_t = v^* + \Delta_t^L + \gamma\sum_{l=1}^{T-t}\Delta_{t+l}^L, \tag{23}$$

where the liquidity effect in period t is equal to

$$\Delta_t^L = \frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\alpha\mathcal{V}\frac{\hat{\theta}_t}{I}, \tag{24}$$

and the equilibrium portfolio is given by

$$\bar{\theta}_t^i = (1-\gamma)^t\theta_0^i + (1-(1-\gamma)^t)\theta^{Av} + \sum_{l=1}^{t}\frac{\hat{\theta}_t}{I}. \tag{25}$$

We proceed by induction. In the last trading period, T, the value function is

$$V_T^i(\Delta\theta_T^i, \Delta\theta_{b,T}^i) = \Delta\theta_{b,T}^i + \bar{\theta}_{b,T-1}^i + A\cdot(\bar{\theta}_{T-1}^i + \Delta\theta_T^i) - \frac{\alpha}{2}(\bar{\theta}_{T-1}^i + \Delta\theta_T^i)\cdot\mathcal{V}(\bar{\theta}_{T-1}^i + \Delta\theta_T^i). \tag{26}$$

In the robust Nash equilibrium, trader i equalizes his marginal utility with marginal revenue (expenditure), given his equilibrium price impact $\bar{\mathcal{M}}_T^i$, which yields

$$A - \alpha\mathcal{V}(\bar{\theta}_{T-1}^i + \Delta\bar{\theta}_T^i) = \bar{p}_T + \bar{\mathcal{M}}_T^i\Delta\bar{\theta}_T^i. \tag{27}$$

Solving for trade $\Delta\bar{\theta}_T^i(\bar{p}_T, \bar{\mathcal{M}}_T^i)$ from (27) 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 $\bar{\mathcal{M}}_T^i = (1-\gamma)(\alpha\mathcal{V}+\bar{\mathcal{M}}_T^j)$; Applying Lemma 1, in a symmetric equilibrium, $\bar{\mathcal{M}}_T^i = \alpha\mathcal{V}(1-\gamma)/\gamma$. Averaging the F.O.C. (27) across all trades, using $\sum_{i\in I}\Delta\bar{\theta}_T^i = \hat{\theta}_T$ and substituting for price impact gives

$$A - \alpha\mathcal{V}(\theta_T^{Av} + \frac{\hat{\theta}_T}{I}) = \bar{p}_T + \frac{1-\gamma}{\gamma}\alpha\mathcal{V}\frac{\hat{\theta}_T}{I}, \tag{28}$$

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

$$\bar{p}_T = v^* - \frac{1-\gamma}{\gamma}\alpha\mathcal{V}\frac{\hat{\theta}_T}{I} = v + \Delta_T^L. \tag{29}$$

Substituting the derived prices (29) back into the F.O.C. (27), we obtain policy functions for the equilibrium trades,

$$\Delta\bar{\theta}_T^i = \gamma(\theta_T^{Av} - \bar{\theta}_{T-1}^i) + \frac{\hat{\theta}_T}{I}. \tag{30}$$

Equations (29) and (30) characterize the equilibrium outcome in T in terms of exogenous parameters, and for T the formulas coincide with (23) and (25). Suppose now that the formulas hold in all periods between $t+1$ and T. We show that they must hold in t. The ultimate portfolio, as a function of trades in period t, is given by

$$\bar{\theta}_T\left(\Delta\theta_t^i, \Delta\theta_{b,t}^i\right) = (1-(1-\gamma)^{T-t})\theta_{t+1}^{Av} + (1-\gamma)^{T-t}(\bar{\theta}_{t-1}^i + \Delta\theta_t^i) + \sum_{l=1}^{T-t}\frac{\hat{\theta}_{t+l}}{I}. \tag{31}$$

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

$$\Delta\theta_{t+l}^{i} = -\gamma\left(1-\gamma\right)^{l-1}\left(\bar{\theta}_{t-1}^{i} + \Delta\theta_{t}^{i}\right) + c_{1}^{i}, \tag{32}$$

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

$$\bar{\theta}_{b,T}^{i}\left(\Delta\theta_{t}^{i},\Delta\theta_{b,t}^{i}\right) = \bar{\theta}_{b,t-1}^{i} + \Delta\theta_{b,t}^{i} + \sum_{l=1}^{T-t}\Delta\theta_{t+l}^{i}\cdot\bar{p}_{t+l} = \bar{\theta}_{b,t-1}^{i} + \Delta\theta_{b,t}^{i} + \left(\bar{\theta}_{t-1}^{i} + \Delta\theta_{t}^{i}\right)\sum_{l=1}^{T-t}\bar{p}_{t+l}\gamma\left(1-\gamma\right)^{l-1} + c_{2}^{i}, \tag{33}$$

where c_2^i is independent of $\Delta\theta_t$ and $\Delta\theta_{b,t}$. Applying (31) and (33) to (1), we observe that the value function is linear in the trade of bonds $\Delta\theta_{b,t}^i$ and quadratic in $\Delta\theta_t^i$. With λ_t defined as

$$\lambda_t \equiv \left(1-\gamma\right)^{T-t}, \tag{34}$$

the derivative of the value function with respect to $\Delta\theta_t^i$ equals

$$D_{\Delta\theta_t^i}V_t^i\left(\cdot\right) = \lambda_t\left[A - \alpha\mathcal{V}\left((1-\lambda_t)\theta_{t+1}^{Av} + \lambda_t(\bar{\theta}_{t-1}^{i} + \Delta\theta_t^i) + \sum_{l=1}^{T-t}\frac{\hat{\theta}_{t+l}}{I}\right)\right] + \sum_{l=1}^{T-t}\bar{p}_{t+l}\gamma\left(1-\gamma\right)^{l-1}. \tag{35}$$

Using the quasilinearity of the value function, the first-order (necessary and sufficient) optimality condition is $D_{\Delta\theta_t^i}V_t^i\left(\cdot\right) = \bar{p}_t + \bar{\mathcal{M}}_t^i\Delta\theta_t^i$, 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 that i's price impact is 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,

$$\bar{\mathcal{M}}_t^i = \frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\alpha\mathcal{V}, \tag{36}$$

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

$$\underbrace{\lambda_t\left(A + \alpha\mathcal{V}\left(\theta_{t+1}^{Av} + \sum_{l=1}^{T-t}\frac{\hat{\theta}_{t+l}}{I}\right)\right)}_{\text{Term 1}} + \underbrace{\sum_{l=1}^{T-t}\bar{p}_{t+l}\gamma\left(1-\gamma\right)^{l-1}}_{\text{Term 2}} = \bar{p}_t - \Delta_t^L. \tag{37}$$

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 prices in all periods in Term 2, we find the coefficient that multiplies v^*

$$v^*\sum_{l=1}^{T-t}\gamma\left(1-\gamma\right)^{l-1} = \gamma\left(\frac{1-\left(1-\gamma\right)^{T-t}}{1-1-\gamma}\right) = v^*\left(1-\left(1-\gamma\right)^{T-t}\right) = v^*\left(1-\lambda_t\right). \tag{38}$$

For any $k = t+1,...,T$, 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 prices following period k (see (21)). This allows us to find the sum of all components that contain Δ_k^L in Term 2 as

$$\gamma\Delta_k^L\left(\gamma\sum_{l=1}^{k-1-t}(1-\gamma)^{l-1}+(1-\gamma)^{k-t-1}\right)=\gamma\Delta_k^L\left(\gamma\frac{1-(1-\gamma)^{k-t-1}}{\gamma}+(1-\gamma)^{k-t-1}\right)=\gamma\Delta_k^L. \tag{39}$$

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

$$\text{Term 2}=v^*(1-\lambda_t)+\gamma\sum_{l=1}^{T-t}\Delta_{t+l}^L. \tag{40}$$

Therefore, the averaged F.O.C. (37) simplifies to

$$\underbrace{\lambda_t v^*}_{\text{Term 1}}+\underbrace{v^*(1-\lambda_t)+\gamma\sum_{l=1}^{T-t}\Delta_{t+l}^L}_{\text{Term 2}}=v^*+\gamma\sum_{l=1}^{T-t}\Delta_{t+l}^L=\bar{p}_t+\Delta_t^L. \tag{41}$$

Solving for the price

$$\bar{p}_t=v^*+\Delta_t^L-\gamma\sum_{l=1}^{T-t}\Delta_{t+l}^L \tag{42}$$

establishes that the equilibrium price in t is as asserted by (23). 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. becomes

$$\begin{aligned}&\lambda_t\left(A-\alpha\mathcal{V}\left((1-\lambda_t)\theta_{t+1}^{Av}+\lambda_t\left(\bar{\theta}_{t-1}^i+\Delta\theta_t^i\right)+\sum_{l=1}^{T-t}\frac{\hat{\theta}_{t+l}}{I}\right)\right)+\sum_{l=1}^{T-t}\bar{p}_{t+l}\gamma(1-\gamma)^{l-1}\\ =\;&\bar{p}_t+\frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\alpha\mathcal{V}\Delta\theta_t^i.\end{aligned} \tag{43}$$

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

$$\Delta\theta_t^i=\gamma\left(\theta_t^{Av}-\bar{\theta}_{t-1}^i\right)+\frac{\hat{\theta}_t}{I}, \tag{44}$$

which gives trades (25). ■

Proof. (Proposition 1: Three-Fund Separation) The result is implied by (25) with $(\hat{\theta}_1,...,\hat{\theta}_T)=0$. ■

Proof. (Proposition 2: Time Structure of Price Impact) The result is derived in (36). ■

Proof. (Proposition 3: Security Market Line) The result follows from two observations:

first, by (42) with $(\hat{\theta}_1, ..., \hat{\theta}_T) = 0$ so that in every period t the liquidity effect $\Delta_t^L = 0$, the prices, and hence the asset returns, are as in the competitive model; second, formula (12) holds in the competitive model. ■

Proof. (Proposition 4: Asset Price Overshooting) In Proposition 6, normalize $t^* = 1$ (w.l.o.g.) and take the sequence of shocks equal to $\{\hat{\theta}_1, 0, \ldots, 0\}$. ■

Proof. (Proposition 5: Price Manipulation) Consider an unanticipated round-trip trade of N risky assets, i.e., a non-zero sequence of trades $(\hat{\theta}_1, ..., \hat{\theta}_T)$, such that $\sum_{t=1}^T \hat{\theta}_t = 0$. The price vector in each period is given by

$$\bar{p}_t = v - \frac{\alpha}{I}\left(\frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\mathcal{V}\hat{\theta}_t + \sum_{k\leq t}\mathcal{V}\hat{\theta}_k\right), \tag{45}$$

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 that are induced by the round trip up to t. The cash obtained from the round-trip trade is equal to

$$\begin{aligned}\sum_{t=1}^T \bar{p}_t \cdot \hat{\theta}_t &= \sum_{t=1}^T \left[v - \left(\frac{\alpha}{I}\frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\mathcal{V}\hat{\theta}_t + \sum_{k\leq t}\mathcal{V}\hat{\theta}_k\right)\right]\cdot\hat{\theta}_t = \\ &= -\frac{\alpha}{I}\sum_{t=1}^T \frac{(1-\gamma)^{2(T-t)+1}}{\gamma}\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_t - \frac{\alpha}{I}\sum_{t=1}^T\left(\sum_{k\leq t}\mathcal{V}\hat{\theta}_k\right)\cdot\hat{\theta}_t, \end{aligned}\tag{46}$$

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

$$\sum_{t=1}^T\left(\sum_{k\leq t}\mathcal{V}\hat{\theta}_k\right)\cdot\hat{\theta}_t = \sum_{t=2}^T\sum_{k<t}\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_k + \sum_t \hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_t. \tag{47}$$

The sum $\sum_{t=2}^T\sum_{k<t}\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_k$ comprises 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

$$\sum_{t=2}^T\sum_{k<t}\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_k + \frac{1}{2}\sum_t\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_t = \frac{1}{2}\sum_{k=1}^T\hat{\theta}_k\cdot\mathcal{V}\sum_{t=1}^T\hat{\theta}_t = 0, \tag{48}$$

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

$$\sum_{t=1}^T\bar{p}_t\cdot\hat{\theta}_t = -\frac{\alpha}{I}\sum_{t=1}^T\left(\frac{(1-\gamma)^{2(T-t)+1}}{\gamma}+\frac{1}{2}\right)\hat{\theta}_t\cdot\mathcal{V}\hat{\theta}_t. \tag{49}$$

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. ■

References

[1] Almgren, R. (2003): "Optimal Execution with Non-linear Impact Functions and Trading Enhanced Risk," *Applied Mathematical Finance*, 10: 1-18.

[2] Almgren, R. and N. Chriss (2000): "Optimal Execution of Portfolio Transactions," *Journal of Risk*, 3: 5-39.

[3] Almgren, R., C. Thum, E. Hauptmann, and H. Li (2005): "Equity Market Impact," *Risk*: 57-62.

[4] Amihud, Y., H. Mendelson, and L. Pedersen (2005): "Liquidity and Asset Prices," *Foundations and Trends in Finance*, 1: 269-364.

[5] Attari, M., A. Mello, and M. Ruckes (2005): "Arbitraging Arbitrageurs," *Journal of Finance*, 60, 5: 2471-2511.

[6] Back, K. (1992): "Insider Trading in Continuous Time," *Review of Financial Studies*, 5, 3: 387-409.

[7] Beneish, M. and R. Whaley (1996): "An Anatomy of the 'S&P Game, The Effects of Changing the Rules," *Journal of Finance*, 5: 1909-1930.

[8] Bertsimas, D. and A. Lo (1998): "Optimal Control of Execution Costs," *Journal of Financial Markets*, 1: 1-50.

[9] Brunnermeier, M. and L. Pedersen (2005): "Predatory Trading," *Journal of Finance*, 4: 1825-1863.

[10] Chan, L. and J. Lakonishok (1993): "Institutional Traders and Intraday Stock Price Behavior," *Journal of Financial Economics*, 33: 173-199.

[11] Chan, L. and J. Lakonishok (1995): "The Behavior of Stock Price Around Institutional Trades," *Journal of Finance*, 50: 1147-1174.

[12] Chordia, T., Roll R., and A. Subrahmanyam (2002): "Commonality in Liquidity," *Journal of Financial Economics*, 56: 3-28.

[13] DeMarzo, P. M. and B. Urošević (2006): "Ownership Dynamics and Asset Pricing with a Large Shareholder," *Journal of Political Economy*, 4: 145-174.

[14] Domowitz, I., O. Hansch, and X. Wang (2005): "Liquidity Commonality and Return Comovement," *Journal of Financial Markets*, 8, 4: 351-376.

[15] Dubil, R. (2002): "Optimal Liquidation of Venture Capital Stakes," Journal of Entrepreneurial Finance and Business Ventures, 7, 2: 56-81.

[16] Easley, D. and M. O'Hara (1987): "Price, Trade Size, and Information in Securities Markets," *Journal of Financial Economics*, 19: 69-90.

[17] Estabrook, J. (1999): "Blockage Discounts," in: *Handbook of Advanced Business Valuation*, R. F. Reilly and R. P. Schweihs (eds.), McGraw-Hill.

[18] Estabrook, J. (2001): "Blockage Discounts: What are they and How do they Affect the Fair Market Value of Publicly Traded Stocks, Real Estate, and Other Assets," American Society of Appraisers International Appraisal Conference.

[19] Foster, F. D. and S. Viswanathan (1996): "Strategic Trading When Agents Forecast the Forecasts of Others," *Journal of Finance*, 51, 4: 1437-78.

[20] Glosten, L. and P. Milgrom (1985): "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 13: 71-100.

[21] Greenwood, R. (2005): "Short- and Long-term Demand Curves for Stocks: Theory and Evidence on the Dynamics of Arbitrage," *Journal of Financial Economics*, 75, 3: 607-649.

[22] Grossman, S. J. (1981): "Nash Equilibrium and the Industrial Organization of Markets with Large Fixed Costs," *Econometrica*, 49, 5: 1149-72.

[23] Grossman, S. J. and M. H. Miller (1988): "Liquidity and Market Structure," *Journal of Finance*, 43, 3: 617-637.

[24] Harris, L. and E. Gurel (1986): "Price and Volume Effects Associated with Changes in the S&P 500: New Evidence for the Existence of Price Pressures," *Journal of Finance*, 41: 815-829.

[25] Hasbrouck, J. and D. Seppi (2001): "Common Factors in Prices, Order Flows and Liquidity," *Journal of Financial Economics*, 59: 383-411.

[26] Hau, H. and H. Rey (2004): "Can portfolio Rebalancing Explain the Dynamics of Equity Returns, Equity Flows, and Exchange Rates?," *American Economic Review*, 94: 126-133.

[27] Hau, H., M. Massa, and J. Peress (2005): "Do Demand Curves for Curriencies Slope Down? Evidence from the MSCI Global Index Change," CERP Discussion Papers, 4862.

[28] Ho, T. and H. R. Stoll (1981): "Optimal Dealer Pricing under Transactions and Return Uncertainty," *Journal of Financial Economics*, 9: 47-73.

[29] Holden, C. W. and A. Subrahmanyam (1996): "Risk Aversion, Liquidity, and Endogenous Short Horizons," *Review of Financial Studies*, 9, 2: 691-722.

[30] Holthausen, R., R. Leftwich, and D. Mayers (1987): "The Effects of Large Block Transactions on Security Prices: A Cross Sectional Analyses," *Journal of Financial Economics*, 19: 237-267.

[31] Holthausen, R., R. Leftwich, and D. Mayers (1990): "Large Block Transactions The Speed of Response and Temporary and Permanent Stock-Price Effects," *Journal of Financial Economics*, 26: 71-95.

[32] Huberman, G. and D. Halka (2001): "Systematic Liquidity," *Journal of Financial Research*, 24: 161-178.

[33] Huberman, G. and W. Stanzl (2004): "Price Manipulation and Quasi-Arbitrage," *Econometrica*, 74, 4: 1247-1276.

[34] Johnson, T. (2008): "Volume, Liquidity, and Liquidity Risk," *Journal of Financial Economics*, 87: 388-417.

[35] Kaul, A., V. Mehrotra, and R. Morck (2000): "Demand Curves for Stocks Do Slope Down: New Evidence from an Index Weights Adjustment," *Journal of Finance*, 55, 2: 893-912.

[36] Keim, D. and A. Madhavan (1995): "The Anatomy of the Trading Process: Empirical Evidence on the Behavior of Institutional Traders," *Journal of Financial Economics*, 37: 371-398.

[37] Keim, D. and A. Madhavan (1996): "The Upstairs Markets for Large-Block Transactions. Analyses and Measurement of Price Effects," *Review of Financial Studies*, 9: 1-39.

[38] Keim, D. and A. Madhavan (1998): "Execution Costs and Investment Performance: An Empirical Analysis of Institutional Equity Trades," Rodney L. White Center for Financial Research Working Paper, 09-95.

[39] Kempf, A. and O. Korn (1999): "Market Depth and Order Size," *Journal of Financial Markets*, 2: 29-48.

[40] Klemperer, P. and M. Meyer (1989): "Supply Function Equilibria in Oligopoly under Uncertainty," *Econometrica*, 57: 1243-1277.

[41] Kraus, A. and H. Stoll (1972): "Price Impact on Block Trading on the New York Stock Exchange," *Journal of Finance*, 27: 569-588.

[42] Kyle, A. S. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53: 1315-1336.

[43] Kyle, A. S. (1989): "Informed Speculation and Imperfect Competition," *Review of Economic Studies*, 56: 517-556.

[44] Loderer, C., J. Cooney, and L. Van Drunen (1991): "The Price Elasticity of Demand for Common Stock," *Journal of Finance*, 46: 621–651.

[45] Lynch A. and R. Mendenhall (1997): "New Evidence on Stock Price Effects Associated with Changes in the S&P Index," *Journal of Business*, 70: 351-383.

[46] Madhavan, A. and M. Cheng (1997): "The Upstairs Markets for Large-Block Transactions: Analysis and Measurement of Price Effects," *Review of Financial Studies*, 9: 1-36.

[47] Mitchell, M., L. H. Pedersen, and T. Pulvino (2007): "Slow Moving Capital," *American Economic Review, P&P*, 97, 2: 215-220.

[48] Newman, Y. and M. Rierson (2004): "Illiquidity Spillovers: Theory and Evidence from European Telecom Bond Issuance," working paper.

[49] Obizhaeva, A., and J. Wang (2005): "Optimal Trading Strategy and Supply/Demand Dynamics," working paper.

[50] Plerou, V., P. Gopikrishnan, X. Gabaix, and H. Stanley (2002): "Qunatifying Stock-price Response to Demand Fluctuactions," *Phisical Review E*, 66, 027104: 1-4.

[51] Pratt, S. (2001): *Business Valuation Discounts and Premiums*, John Wiley & Sons Inc., New York.

[52] Pritsker, M. (2005): "Large Investors: Implications for Equilibrium Asset, Returns, Shock Absorbtion, and Liquidity," Fincance and Economic Discussion Series 2005-36, Board of Governors of the Federal Reserve System.

[53] Rostek, M. and M. Weretka (2008): "Dynamic Bilateral Oligopoly," working paper, University of Wisconsin-Madison.

[54] Rostek, M. and M. Weretka (2009): "Liquidity Concerns in Thin Financial Markets," working paper, University of Wisconsin-Madison.

[55] Stoll, H. (1978): "The Pricing of Security Dealers Service: An Empirical Study of NASDAQ Stocks," *Journal of Finance*, 33: 1153-1172.

[56] Subramanian, A. and R. Jarrow (2001): "The Liquidity Discount," *Mathematical Finance*, 11: 447-474.

[57] Urošević, B. (2005): "Moral Hazard and Dynamics of Insider Ownership Stakes," Universitat Pompeu Fabra, Economics Working Papers, 787.

[58] Vayanos D. (1999): "Strategic Trading and Welfare in a Dynamic Market," *Review of Economic Studies*, 66: 219-54.

[59] Vayanos, D. (2001): "Strategic Trading in a Dynamic Noisy Market," *Journal of Finance*, 56, 1: 131-171.

[60] Weretka, M. (2006): "Endogenous Market Power," working paper, University of Wisconsin-Madison.

[61] Weretka, M. (2009): "Slope Takers in Anonymous Markets," working paper, University of Wisconsin-Madison.

FIGURE 1: EVOLUTION OF EQUILIBRIUM PORTFOLIOS (A) AND PRICE IMPACTS (B)

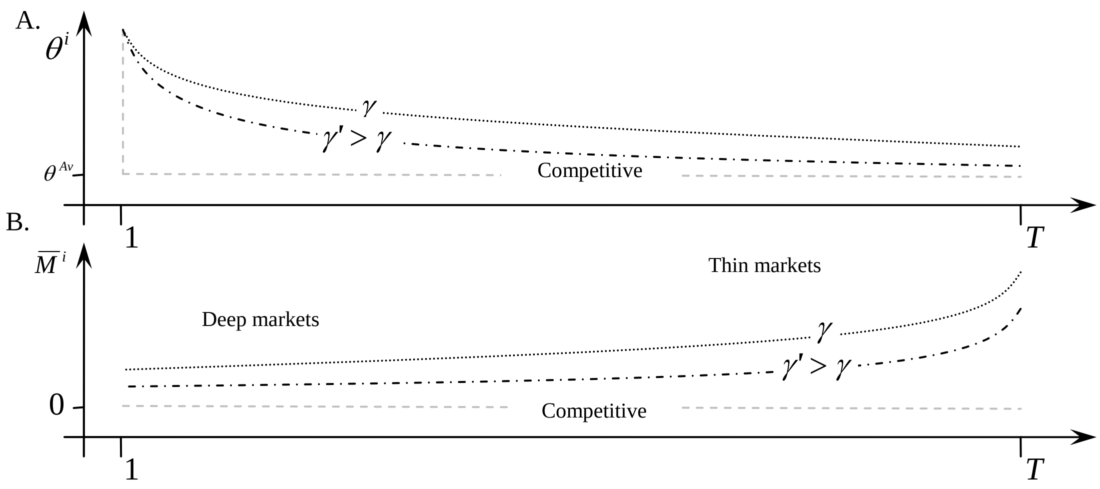

FIGURE 2: SHORT AND LONG-RUN DEMAND

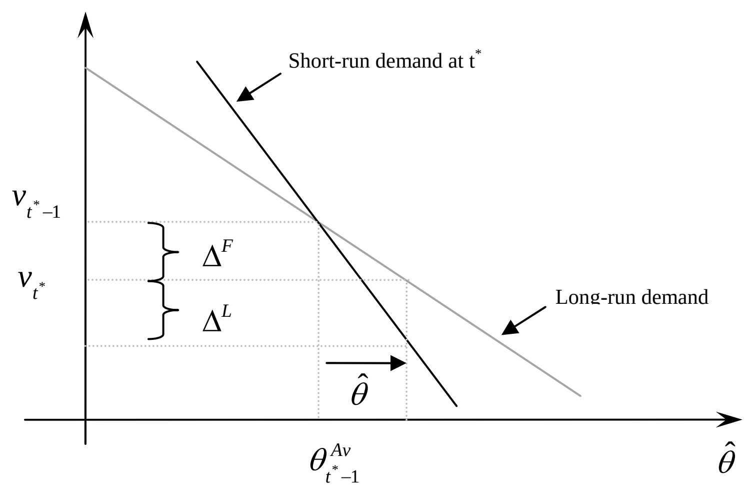

FIGURE 3: RESPONSE OF PORTFOLIOS (A) AND PRICES (B) TO AN UNANTICIPATED SUPPLY SHOCK

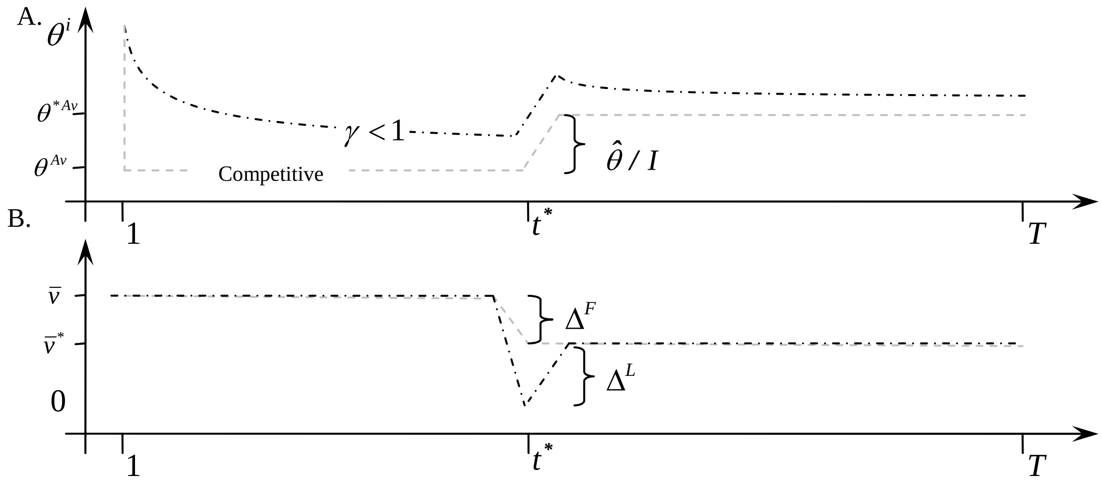

FIGURE 4: RESPONSE OF PORTFOLIOS (A) AND PRICES (B) TO AN ANTICIPATED SUPPLY SHOCK