Asset Prices and Liquidity with Market Power and Non-Gaussian Payoffs*

Sergei Glebkin, Semyon Malamud, and Alberto Teguia

January 15, 2020

Abstract

We consider an economy populated by CARA investors who trade, accounting for their price impact, multiple risky assets with arbitrary distributed payoffs. We propose a constructive solution method: finding the equilibrium reduces to solving a linear ordinary differential equation. With market power and non-Gaussian payoffs: (i) the equilibrium is nonlinear and the model can speak to key stylized facts regarding asymmetry and nonlinearity of price response to order imbalances, (ii) when risk aversion decreases, there are more liquidity providers and/or there is less uncertainty about future asset payoffs, liquidity can *decrease*, (iii) cross-section of returns is affected by endogenous illiquidity.

JEL Classification: D21, G31, G32, G35, L11

Keywords: Price Impact, Higher Cumulants, Strategic Trading, Liquidity, CAPM

*Sergei Glebkin (glebkin@insead.edu) is at INSEAD, Semyon Malamud (semyon.malamud@epfl.ch) is at EPFL, and Alberto Teguia (alberto.mokakteguia@sauder.ubc.ca) is at UBC Sauder. For valuable feedback we thank Bruno Biais, Svetlana Bryzgalova, Georgy Chabakauri, Jean-Edouard Colliard, Bernard Dumas, Daniel Ferreira, Thierry Foucault, Ravi Jaganathan, Dong Lou, John Kuong, Peter Kondor, Albert (Pete) Kyle, Jiasun Li, Igor Makarov, Ian Martin, Milan Martinovic, Konstantin Milbrandt, Rohit Rahi, Anna Obizhaeva, Olga Obizhaeva, Martin Oehmke, Joel Peress, Dimitri Vayanos, Gyuri Venter, Kathy Yuan, and Kostas Zachariadis. We are especially grateful to Marzena Rostek for her comments and suggestions.

1 Introduction

Illiquidity, or market's inability to accommodate large trades without a change in price, has important effects on trading and pricing of financial assets. The illiquidity is not negligible, even for such a developed market as US equities, and it is an important determinant of the cross-section of stock returns.[1] Illiquidity also limits the extent to which a particular investment strategy or anomaly-based trade can be scaled up while still generating alpha, determining the capacity and economic significance of a particular asset-pricing anomaly.[2] Traders do account for market illiquidity. Institutional investors, such as mutual and pension funds, often trade *strategically*, taking into account the fact that their trades can move prices. Some investors, such as J.P. Morgan or Citigroup, have in-house optimal execution desks that devise trading strategies to minimize price-impact costs. Other investors use the software and services provided by more specialized trading firms. Such strategic trading is in contrast to price-taking behavior commonly assumed in classical asset-pricing models.[3]

How are illiquidity and asset prices determined in equilibrium when investors take their price impact into account? Literature on strategic trading addresses this question, often adopting a CARA-normal framework for tractability: traders have negative exponential (CARA) utility functions, and asset payoffs are normally distributed.[4] Such a framework has a number of shortcomings. First, CARA-normal models feature linear equilibria, in which the price is a linear function of order size and purchases and sells have the same price impact. This is hard to

[1]For example, a recent paper by Koijen and Yogo (2019) estimates that the price impact of a 10% demand shock for the median US stock was consistently above 20% between 1980 and 2017. In the same paper, Koijen and Yogo (2019) document that most of the variation in the cross-section of stock returns is explained by demand shocks that are unrelated to changes in observed characteristics. In particular, they estimate that these shocks explain 81% of the cross-sectional variance of stock returns. The fact that such shocks would have no effect on returns in a perfectly liquid market underscores the importance of illiquidity for the cross-section of stock returns.

[2]See Frazzini, Israel, and Moskowitz (2012), Novy-Marx and Velikov (2015), and Landier, Simon, and Thesmar (2015) for estimates of the capacity of different asset-pricing anomalies. All of these estimates are based on authors' estimates of price impact (illiquidity).

[3]Indeed, in all models covered in a popular Cochrane (2009) textbook investors are price-takers.

[4]Settings with risk-neutral traders, such as Kyle (1985), are a particular case of a setting with CARA utility, with the coefficient of risk aversion equal to zero.

align with empirical evidence that documents that prices react to large orders in an asymmetric and nonlinear way: purchases typically have greater price impact compared to sells, and price response is a concave function of order size.[5] Second, normality implies that higher moments play no role, which is not true in practice.[6] Third, the strategic trading models are often cast in the domain of individual securities and are therefore silent on the determinants of cross-section of illiquidity and stock returns.[7]

To address the shortcomings mentioned above, we consider a tractable model of strategic trading that allows for multiple assets and general distribution of asset payoffs. Our main results are as follows. First, with non-Gaussian payoffs equilibrium becomes nonlinear and asymmetric between buys and sells and predictions of our model are in line with key stylized facts regarding asymmetry and nonlinearity of price response to order imbalances. Second, we demonstrate that some of the common wisdoms about illiquidity derived in CARA-normal models are not robust. In particular, we show that when risk aversion decreases, there are more liquidity providers and/or there is less uncertainty about future asset payoffs liquidity can *decrease*. Third, commonality in illiquidity arises naturally in our model, and we derive implications regarding the cross-section of illiquidity and asset returns, and show that higher moments affect both. The main technical challenge is that with non-Gaussian distribution, the traditional guess-and-verify approach is no longer applicable, as it is not clear what should be the guess. We propose a novel constructive solution method that allows overcoming this difficulty and solving the model in closed form for any distribution.

We assume that CARA traders, which we call *liquidity providers*, exchange multiple risky assets for a riskless asset over one period, accounting for their price impact. Liquidity

[5] Hausman, Lo, and MacKinlay (1992), Almgren, Thum, Hauptmann, and Li (2005), and Frazzini et al. (2012) find concave price reaction functions (absolute value of price change as a function of order size) for equities. Muravyev (2016) presents the evidence for options. Regarding asymmetry, Saar (2001) summarizes the evidence that shows bigger price impact of buy orders compared to sell orders.

[6] Harvey and Siddique (2000) present evidence that stocks co-skewness with the market is priced in the cross-section of stock returns. Martin (2017) points out that the significant difference between SVIX and VIX indices is inconsistent with normality of (log) returns.

[7] Two notable exceptions are Rostek and Weretka (2015a) and Malamud and Rostek (2017), who study multi-asset models in a CARA-normal framework.

providers have the same risk aversion coefficient and are symmetrically informed. The absence of asymmetry of information implies that the unique source of price impact in our setting is inventory risk.[8] Trading is structured as a uniform-price double auction: traders submit simultaneously demand functions, specifying an amount of assets they want to buy, depending on prices of all assets. All trades are executed at the prices that clear the market. Our main innovation relative to previous literature is to assume that the distribution of the risky assets payoffs is completely general, save for the restriction that it has bounded support.[9] Our main simplification is the absence of heterogeneity among liquidity providers (both in terms of information and preferences). In addition to liquidity providers, there are *liquidity demanders* who submit market orders. Trade occurs because liquidity providers compete to absorb part of the liquidity demanders' aggregate order, hence providing liquidity to them. We do not impose any restrictions on liquidity demanders except that their aggregate order is independent of the asset payoff, and so they are uninformed. We express all equilibrium quantities as functions of liquidity demanders' aggregate order.In equilibrium, traders determine their optimal demand function, knowing the demand functions of all other traders. We show that the optimization problem is equivalent to traders not knowing others' demand functions but knowing their own price impact matrix (i.e., how their trade moves prices of assets at the margin) for each order size. This is an intuitive representation of the problem: real-world traders typically have a market impact model that is an input in their optimal execution algorithm.[10] The equilibrium price impact matrix is pinned down by the requirement that it is consistent with the demand functions of the other traders. The consistency requirement yields a linear ordinary differential equation (ODE) in the single asset case and a system of partial differential equations (PDEs)

[8] We abstract from asymmetric information for tractability. However, our focus on inventory risk is justified by empirical findings that document it to be a dominant source of price impact in many markets. See for example, Muravyev (2016) for evidence in the options market. Extending our model to allow for asymmetric information component of price impact is left for future research.

[9] We can handle distributions with unbounded support as well. For example, in our benchmark case of Gaussian distribution, payoffs are unbounded. We analyze this case as a limit of our model with a truncated distribution as truncation bounds go to infinity. Our model can handle any distribution with unbounded support (for which the limit just described exists) in a similar way.

[10] Rostek and Weretka (2015a) was the first paper to derive such representation. Their model is cast in the CARA-normal framework. We generalize their result to non-Gaussian distributions.

in the multi-asset case. The single-asset ODE for equilibrium price function can be solved in closed-form for any probability distribution. We show that solving the system of PDEs in the multi-asset case can be reduced to solving a single-asset ODE. This ODE characterizes the price function in a single asset economy, where the single asset is an index, characterized by a vector of asset holdings q. The solution to PDEs can be obtained by differentiating the solution to the single-asset ODE with respect to q. Thus, we can characterize the equilibrium in a multi-asset economy in closed-form as well. We show equilibrium uniqueness in the class of equilibria with strictly decreasing demands and arbitrage-free equilibrium prices.

Using the closed-form solutions for the equilibrium price function, we examine the relationship between price and order size. We consider a tractable limit when liquidity providers' risk aversion is small, so that only the first few moments of payoff distribution are important for the properties of equilibrium. Our analysis highlights the importance of the third moment. When the payoff of the portfolio traded by the liquidity demanders is positively skewed, purchases of this portfolio have greater price impact compared to sells. The intuition for the result can be seen by contrasting with the benchmark case, in which the asset payoff is normally distributed and hence the skewness is zero. Consider first a sell of a portfolio by liquidity demanders. Liquidity suppliers who buy from them, receive a positively skewed profit, which they like. Intuitively, positive skewness implies that positive surprises to profits are more likely than negative ones. As a result, liquidity providers require a lower premium for providing liquidity and the price reaction to the order is smaller than in the benchmark case. For purchases, the trader's counterparties, who sell to him, receive a negatively skewed profit and require a greater premium. The price reaction is greater than in the benchmark case. Consequently, with positive skewness, purchases have greater price impact compared to sells. Similarly, when payoff of the portfolio traded by the liquidity demanders is negatively skewed, sells have greater price impact compared to purchases.[11] Since at the individual level stock returns are positively

[11]Similar intuition applies to the competitive version of our model, i.e., when liquidity providers take prices as given. Thus, even the competitive version of our model can speak to key stylized facts regarding asymmetry and nonlinearity of price response to order imbalances, which, to the best of our knowledge, is a new result. The

skewed, whereas at the level of indices the skewness is negative (e.g., Bakshi, Kapadia, and Madan (2003), and Albuquerque (2012)), our analysis implies that when liquidity demanders trade individual stocks, purchases move prices of more than sells, whereas the opposite is true when they trade indices. Both of these implications are consistent with empirical evidence.[12]

Our second second set of results examines equilibrium illiquidity (defined as the sensitivity of price to supply shocks) in a single-asset version of our model. We show that when liquidity providers' risk aversion is high, illiquidity can *decrease* when risk aversion increases, there are fewer liquidity providers and/or there is more uncertainty about future asset payoffs. In contrast, when risk aversion is low, the opposite, conventional results apply. The difference is rooted in the way a CARA trader's valuation of risky assets behaves with low versus high risk aversion: in the first case, a trader cares about mean and variance; in the second case he cares about worst-case scenario. Indeed, when the risk aversion is low, the valuation (or the marginal certainty equivalent utility, in technical terms) can be approximated well by a standard mean-variance trade-off and standard comparative statics apply. In contrast, when risk-aversion is high, the valuation is approximated by a worst case asset payoff (i.e., lowest payoff for long, and highest payoff for short position). Importantly, such valuation, and hence the price, is not sensitive to the quantity of asset, implying no illiquidity. Hence, as risk aversion increases to infinity, the illiquidity (which is generally positive) *decreases* to zero, resulting in the unconventional result that with higher risk aversion illiquidity can be smaller. Similar intuition applies to comparative statics with respect to the number of liquidity providers and uncertainty about future payoffs. These unconventional results imply that when risk aversion is high, i.e., in bad times, policy measures aimed at improving liquidity might have exactly the opposite effect. This is important, since policy interventions typically occur in bad times.

We also examine the determinants of the cross-section of illiquidity and asset returns in a

non-competitive part of the price is response is associated with short-term price reversals (as shown theoretically in Rostek and Weretka (2015a)). Our non-competitive model links the asymmetry in this part of price response to skewness.

[12]Saar (2001) summarizes the evidence that shows bigger price impact of buy orders compared to sell orders in individual stocks. Chordia, Roll, and Subrahmanyam (2002) document the opposite at the market level.

multi-asset version of our model. We show that despite the absence of arbitrage, the difference between an asset's expected return under risk-neutral measure and the risk-free rate is not zero. Instead, this difference reflects asset's illiquidity, defined as the incremental return on the asset of interest due to liquidation of an additional unit of liquidity demanders' portfolio. We argue that this difference is related to asset carry as defined in Koijen, Moskowitz, Pedersen, and Vrugt (2018). In line with our theoretical finding that carry is positively related to illiquidity Koijen et al. (2018) find that carry is a strong positive predictor of returns and that carry strategies are positively exposed to global liquidity shocks. We then consider the limit in which liquidity providers' risk aversion is small and show that cross-section of returns is driven by four factors: (i) return on the portfolio of liquidity providers, (ii) the square of this return, (iii) the return on the portfolio of liquidity suppliers, and (iv) the product of (i) and (iii). The last two factors also explain the cross-section of illiquidity.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 solves for equilibrium in the single-asset case and derives the implications concerning illiquidity and the shape of price response function. Section 4 considers equilibrium in the multi-asset case and derives implications regarding cross-section of illiquidity and asset returns. Section 5 summarizes the implications. Section 6 reviews related literature. Section 7 concludes. Technical details are relegated to appendices.

2 The model

There are two time periods $t \in \{0, 1\}$. A number $L > 2$ of strategic *liquidity providers* trade assets with *liquidity demanders* at $t = 0$ and consume at $t = 1$.[13] There are N risky assets (stocks) and a risk-free asset (a bond). The bond is in perfectly elastic supply and earns a gross

[13]Under more strict technical conditions on the distribution of δ than we impose below, an equilibrium with $L = 2$ exists in our model. However, with $L = 2$ the equilibrium does not exist in an important benchmark when the distribution of δ is Gaussian (as is well-known, Kyle (1989)). For this reason we restrict ourselves to the case $L > 2$. A demand function equilibrium with two traders is analyzed in Du and Zhu (2017).

return R_f. A stock k is a claim to a terminal dividend δ_k. We impose the following restrictions on the distribution dividends.

Assumption 1. *The random variables $(\delta_1, \delta_2, ..\delta_N)$ are linearly independent modulo constant. That is, there does not exist a non-trivial linear combination of $(\delta_1, \delta_2, ..\delta_N)$ that is almost surely a constant.*

Assumption 1 simply requires that there are no redundant securities.

Assumption 2. *The joint distribution of dividends has bounded support.*

Assumption 2 is natural. In reality, investors are protected by limited liability, which implies that dividends δ_i are nonnegative. Hence, there exists a lower bound. An upper bound is also natural, as the resources of any firm are limited, and hence an asset cannot have an infinite payoff. However, our benchmark case with Gaussian distribution does not satisfy Assumption 2. We will analyze this case as a limit of our model with δ distributed according to a truncated normal distribution as truncation bounds go to infinity. Our model can handle any distribution with unbounded support (for which the limit just described exists) in a similar way.

The joint distribution of dividends $\delta \equiv (\delta_1, \delta_2, ..\delta_N)$ is characterized by the *cumulant generating function (CGF)*

$$g(y) \equiv \log E[\exp(y^\top \delta)].$$

The CGF contains information on the moments of the distribution δ. In particular,

$$\frac{\partial g}{\partial y_i}(0) = E[\delta_i],$$

$$\frac{\partial^2 g}{\partial y_i \partial y_j}(0) = E[(\delta_i - E[\delta_i])(\delta_j - E[\delta_j])] \equiv \mathrm{cov}(\delta_i, \delta_j), \text{ and}$$

$$\frac{\partial^3 g}{\partial y_i \partial y_j \partial y_k}(0) = E[(\delta_i - E[\delta_i])(\delta_j - E[\delta_j])(\delta_k - E[\delta_k])] \equiv \mathrm{coskew}(\delta_i, \delta_j, \delta_k).$$

We follow the convention that skewness is the third central moment of the distribution. Therefore,

$$\text{skew}(\delta_i) = E\left[(\delta_i - E[\delta_i])^3\right] = \frac{\partial^3 g}{\partial y_i^3}(0) = \text{coskew}(\delta_i, \delta_i, \delta_i).$$

Most of our results do not require us to model liquidity demanders explicitly. Instead, we assume that their aggregate trade is characterized by the aggregate net supply $s \in \mathbb{R}^N$. We will express all equilibrium quantities as a function of the net supply s. We impose the following restrictions on the net supply.

Assumption 3. *The net supply s is independent of dividend δ.*

The assumption above implies that one cannot learn about δ from s. Thus, information is symmetric between liquidity suppliers and liquidity demanders.

Assumption 4. *The net supply s is uncertain and its distribution has full support.*

Supply uncertainty is important for ruling our the extreme multiplicity of equilibria, as in Klemperer and Meyer (1989) and Vayanos (1999). As in Klemperer and Meyer (1989), assumptions 3 and 4 imply that equilibrium quantities will depend on realization, but not the distribution of s. Therefore, we are not specifying a particular distribution for supply. Assumption 4 is not very restrictive: one can always achieve full support uncertainty about s by adding to liquidity demanders a small measure of noise traders whose random demands have full support and then taking a limit of the model as a measure of noise traders goes to zero.[14]

The large traders are identical and maximize the expected CARA utility from their terminal wealth W, accounting for their price impact. Suppose all traders except trader i submit demand functions $D(p)$. Then, optimal demand for a trader i, $D^i(p)$ solves the following

[14]As in robust Nash equilibrium of Rostek and Weretka (2015a).

problem;

$$\begin{aligned}
&\max_{D^i(p)} E[-\exp(-\gamma W)], \\
\text{s.t.: } &W = (\delta - R_f p(D^i(p), D(p)))^\top (D^i(p) + x_0), \text{ and} \\
&p(D^i(p), D(p)): \; D^i(p) + (L-1)D(p) = s.
\end{aligned} \tag{1}$$

Before moving to our equilibrium concept, we define below the set of arbitrage-free prices, which will be used in the definition of equilibrium.

Definition 1. *Let $\mathcal{A} \subset \mathbb{R}^N$ denote the set of arbitrage-free price vectors. Namely, for each $p \subset \mathcal{A}$ and each portfolio $q \subset \mathbb{R}^N$, $q^\top(\delta - p) < 0$ with positive probability.*

The equilibrium concept is an arbitrage-free symmetric Nash equilibrium with strictly decreasing demands (simply *equilibrium* in what follows) formally defined below.

Definition 2. *A function $D(p) : \mathbb{R}^N \to \mathbb{R}^N$ is an equilibrium demand if: (i) for any $i = 1, 2, ..L$, given that the traders $j \neq i$ submit demands $D^j(p) = D(p)$, it is optimal for the trader i to submit demand $D^i(p) = D(p)$, i.e., $D^i(p)$ solves the problem (1), (ii) $D(p)$ is strictly decreasing, i.e., $(D(p) - D(\hat{p}))^\top (p - \hat{p}) < 0$ for all $p \neq \hat{p}$, and (iii) $D(p)$ is continuously differentiable and the Jacobian ∇D is non-degenerate everywhere. Define $I(\cdot)$ as the inverse of $D(\cdot)$.[15] We also require that (iv) for any q, $I(q) \in \mathcal{A}$.*

Part *(i)* of the definition above is simply a Nash equilibrium requirement. Parts *(ii)* and *(iii)* are technical: they ensure that the inverse demand, which we will solve for when deriving the equilibrium, is well-defined. Part *(iv)* is important in ensuring the uniqueness of equilibrium: solving for equilibrium amounts to solving an ODE; the condition *(iv)* provides a transversality condition selecting a unique solution. The economic meaning of the condition *(iv)* is as follows. Suppose that in addition to strategic liquidity providers we have an arbitrarily small mass of competitive (price-taking) liquidity providers. Then, for prices that are not arbitrage-free,

[15]Part (ii) implies that D is bijective; (iii) combined with the inverse function theorem implies that the image of D is an open subset of $\mathbb{R}^N$ and that the inverse $I = D^{-1}$ of D is a continuously differentiable map.

the price-taking liquidity providers would submit infinite demands, and the market would not clear. Hence, there will be no equilibria with prices outside $\mathcal{A}$. Therefore, requirement *(iv)* selects among many potential equilibria the one that is robust to presence of a vanishingly small number of competitive liquidity providers.

Throughout the paper we use the following notation. The time $t = 0$ certainty equivalent of a position achieved after a trade $q \in \mathbb{R}^N$ in the risky assets, starting from a portfolio $x_0 \in \mathbb{R}^N$ is denoted by $f(q; x_0)$. By definition, $f(q; x_0)$ solves

$$\exp(-\gamma R_f f(q; x_0)) \equiv E_\delta[\exp(-\gamma(x_0 + q)^\top \delta)].$$

It is clear that the *certainty equivalent* $f(q)$ is related to the cumulant-generating function as follows:

$$f(q; x_0) = -\frac{1}{\gamma R_f} g(-\gamma(x_0 + q)).$$

We will often suppress the second argument ans simply write $f(q)$, whenever it does not cause confusion.

3 The case of a single risky asset

In this section we consider the case of a single risky asset and derive the implications concerning illiquidity and the shape of price response function. We derive the cross-sectional implications for stock returns in Section 4, while considering the multiple assets case.

To highlight some of the key economic forces at play in this section, it is instructive to consider approximations of marginal certainty equivalent for a CARA trader with large and small risk aversion γ. For simplicity we assume $x_0 = 0$ and $q > 0$ in expressions (2) and (3) below. When risk aversion is small one can write

$$f'(q) \approx \frac{1}{R_f}E[\delta] - \frac{\gamma q}{R_f}\text{var}[\delta] + \frac{\gamma^2 q^2}{2R_f}\text{skew}[\delta]. \tag{2}$$

The first two terms summarise the standard mean-variance trade-off. The third term illustrates that CARA traders like skewness. Intuitively, positive skewness implies that positive surprises to profits are more likely than negative ones. This preference for skewness will be important for the results in Section 3.4.

When risk aversion is large one can write

$$f'(q) \approx \text{ess inf}_\delta\, \delta^\top (x_0 + q). \tag{3}$$

When risk aversion is large, CARA traders evaluate the asset according to the worst-case payoff δ. The fact that CARA traders care about mean, variance and possibly higher moments when risk aversion is small, but care about the worst case scenario when risk-aversion is high, is important for the unconventional results about illiquidity in Section 3.3.

3.1 Characterization of equilibrium

We first derive the equilibrium characterization heuristically to show the intuition, and then justify the derivation in Theorem 1 below. Consider first a competitive (price-taking) trader and his equilibrium demand. He solves

$$\max_q f(q) - pq. \tag{4}$$

His inverse demand $p = I(q)$ is determined by the first-order condition in the problem above

$$f'(q) = p.$$

A strategic trader accounts for the fact that he can move prices, and hence his first-order condition has a new term that arises due to this ability, as follows:

$$f'(q) - \frac{\partial p}{\partial q} q = p. \tag{5}$$

Once the large trader knows the price sensitivity $\frac{\partial p}{\partial q}$, he is able to solve for his optimal demand from the first-order condition above.

Suppose that each trader has a *conjecture* $\Lambda(q) = \frac{\partial p}{\partial q}$ about how he can move prices in equilibrium. This function shows how much the trader of interest can move prices if he trades $q + dq$ instead of q. This conjecture, together with first-order condition (5), determines his optimal (inverse) demand,

$$f'(q) - \Lambda(q)q = p. \tag{6}$$

In a Nash equilibrium, the conjectures $\Lambda(q)$ have to be *consistent* with the inverse demands of other traders, i.e., $\Lambda(q)$ must be given by the slope of the inverse residual supply (see Kyle (1989)).

Consistency implies a relationship between $\Lambda(q)$ and $I'(q)$. In the symmetric equilibrium, there are $(L-1)$ identical demands contributing to the slope of the residual supply. The slope of the residual supply is thus $-(L-1)\frac{1}{I'(q)}$. The minus is to account for the fact that the residual supply is upward-sloping, while the demand is downward-sloping; we also exploit the fact that the slope of the demand is the reciprocal of the slope of its inverse. The slope of the inverse residual supply is therefore

$$\Lambda(q) = \frac{-1}{L-1} I'(q). \tag{7}$$

The equilibrium inverse demand should therefore satisfy two conditions: *optimality* (Equation (6)) and *consistency* (Equation (7)). Note that these two conditions result in a linear ODE,

$$f'(q) + \frac{1}{L-1} I'(q)\, q = I(q)\,. \tag{8}$$

The linearity of this ODE implies that it can be solved in closed-form using standard methods, which highlights the tractability of our approach.

In the derivation above, we also used the following condition implicitly: for the inverse bid and the inverse residual supply to be well-defined, the bids should be monotone. Finally, part *(iv)* of the equilibrium definition requires prices to be arbitrage-free, which translates in the no-free-lunch condition in the theorem below, which summarizes the above discussion.

Theorem 1. *(Equilibrium characterisation) A strictly decreasing function $I(q)$ is an equilibrium inverse demand if, and only if, it satisfies the following three conditions:*

(1) ***(Optimality)*** *The demand $I(q)$ is optimal given a conjecture about the price impact matrix $\Lambda(q)$,*

$$I(q) = f'(q) - \Lambda(q)q, \tag{9}$$

(2) ***(Consistency)*** *The conjecture about the price impact matrix $\Lambda(q)$ is consistent with the equilibrium demand $I(q)$*

$$\Lambda(q) = -\frac{1}{L-1} I'(q), \ \text{and} \tag{10}$$

(3) ***(No free lunch)***

$$I(q) \in \mathcal{A}, \ \forall q. \tag{11}$$

We follow Rostek and Weretka (2015a) in using the representation of equilibrium through optimality and consistency conditions because it captures nicely the decision-making of real traders. As in real life, traders have a market impact model and determine optimal bids (engage in optimal trade execution), given this model. In a Nash equilibrium, the market impact (price sensitivity) model is specified by the consistency condition. The main contribution, relative to Rostek and Weretka (2015a), is to derive the consistency condition when price impact is a function of order size. The fixed-point condition then results in an ODE, not an algebraic equation. The representation above is also very useful for solving the model. To find an equilibrium demand, one has to jointly solve (9) and (10), which reduces to a linear ODE, which can be solved using standard methods. We do this in the following section.

3.2 Closed-form solution

Theorem 1 implies that to find equilibrium demand $I(q)$, one needs a solution to (8) that is strictly decreasing and satisfies the no-arbitrage restriction (11).

Using the standard methods, one can write the solution to (8) for $q > 0$ as follows:

$$I(q) = (L-1)\int_1^\infty \xi^{-L} f'(\xi q) d\xi + cq^{L-1}, \tag{12}$$

for an arbitrary constant $c \le 0$.[16] The condition that $c \le 0$ is to ensure that the solution is downward sloping.

For (12) to be the equilibrium inverse demand, they must satisfy the no-free-lunch condition. It can easily be seen that solutions with $c < 0$ violate that condition. Indeed, these solutions are unbounded. Therefore, for q high enough $I(q)$ will be smaller than the minimal payoff to the asset, which means that for such q, $I(q) \notin \mathcal{A}$. In the proof of Proposition 1, we show that the solution with $c = 0$ does satisfy the no-free-lunch condition. Therefore, the unique equilibrium demand is given by (12), with $c = 0$. The proposition below summarises the discussion above.

Proposition 1. *(Closed-form solution) There exists unique equilibrium. The equilibrium in-*

[16] One way to obtain (12) is to multiply both sides of (8) by the integrating factor q^{-L} and to integrate both parts from q to ∞.

verse demand $I(q)$ and price impact $\Lambda(q)$ are given by

$$
\begin{aligned}
I(q) &= (L-1)\int_1^\infty \xi^{-L} f'(\xi q) d\xi && (13)\\
&= \frac{L-1}{R_f}\int_1^\infty \xi^{-L} g'(-\gamma\xi(q+x_0)) d\xi, && (14)\\
\Lambda(q) &= -\int_1^\infty \xi^{1-L} f''(\xi q) d\xi && (15)\\
&= \frac{\gamma}{R_f}\int_1^\infty \xi^{1-L} g''(-\gamma\xi(q+x_0)) d\xi. && (16)
\end{aligned}
$$

3.3 Equilibrium illiquidity

In this section we analyze equilibrium illiquidity, focusing on its comparative statics with respect to the parameters of the model. Our measure of illiquidity is the sensitivity of equilibrium price to supply s, as is standard in the literature (see, e.g., a review of Vayanos and Wang (2012)). A straightforward calculation yields that it is related to price impact of liquidity providers, as follows:

$$
\text{Illiquidity}(s) \equiv \left(\frac{\partial P(s)}{\partial s}\right)^{-1} = \frac{L-1}{L}\Lambda(s/L).
$$

To guide our analysis in the general case, we consider first the Gaussian benchmark in the corollary below.

Corollary 1. *Suppose that $\delta \sim N(\mu, \sigma^2)$. Then*

$$
\text{Illiquidity} = \frac{\gamma(L-1)\sigma^2}{L(L-2)}.
$$

Consequently, Illiquidity *is decreasing in L, and increasing in γ and σ^2.*

The corollary above is in line with the common wisdom about illiquidity. Illiquidity is higher when there are fewer liquidity providers, as they compete less to provide liquidity. It is also higher when risk of the uncertainty about asset's payoff is higher or when liquidity providers' risk aversion is higher, as they have less capacity to bear risk. We show in Proposition 2 that with general distribution the common wisdom holds for small risk aversion γ, but for large risk aversion the comparative statics of Corollary 1 are necessarily reversed.

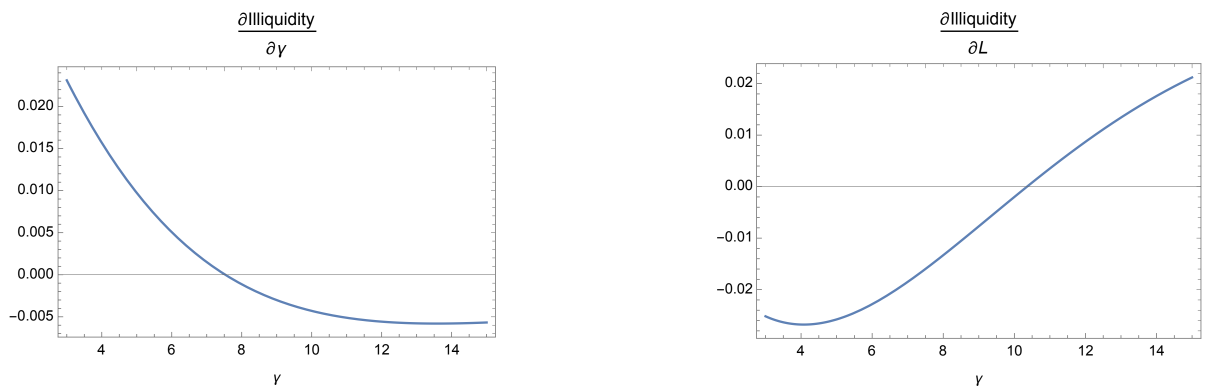

Figure 1: The comparative statics of Illiquidity with respect to γ and L: $\partial\text{Illiquidity}(s;\gamma,L)/\partial L$ and $\partial\text{Illiquidity}(s;\gamma,L)/\partial\gamma$ as functions of γ for $\delta \sim \text{Uniform}[0,1]$, $s=1$ and $L=4$.

To do the comparative statics with respect to the uncertainty about asset's payoff, we assume there is a public information release about δ. We assume that all traders start with information characterized by sigma-algebra $\mathcal{F}_1$, and, after information is released, their information set is characterized by $\mathcal{F}_2 \subset \mathcal{F}_1$. We impose only the following restriction on $\mathcal{F}_2$.

Assumption 5. *The support of distribution of δ, given $\mathcal{F}_2$, is the same as the support of δ, given $\mathcal{F}_1$.*

This assumption means that the information released allows for updating the probabilities but not the set of possible realizations of δ. This assumption is met, for example, when $\mathcal{F}_1$ changes to $\mathcal{F}_2$ as a result of announcement of a signal $\delta + u$, where the noise u is independent of δ and has full support.

Proposition 2. *Denote by* $\text{Illiquidity}(s;\gamma,L,\mathcal{F}_i)$ *the expectation (given $\mathcal{F}_1$) of illiquidity, given* *information $\mathcal{F}_i$. Then, for γ small enough we have*

$$\frac{\partial \text{Illiquidity}(s;\gamma,L,\mathcal{F}_1)}{\partial L} < 0, \tag{17}$$

$$\frac{\partial \text{Illiquidity}(s;\gamma,L,\mathcal{F}_1)}{\partial \gamma} > 0, \tag{18}$$

$$\textit{and} \text{Illiquidity}(s;\gamma,L,\mathcal{F}_1) > \text{Illiquidity}(s;\gamma,L,\mathcal{F}_2). \tag{19}$$

These inequalities must change to the opposites at least once, as γ changes from 0 to ∞.

Intuitively, when γ is small, higher moments do not matter much since they enter with higher powers of γ in the expansion of certainty equivalent around $\gamma = 0$. Therefore, for small γ, the model is close to Gaussian, and the comparative statics of Corollary 1 are preserved. However, when γ is large enough, the comparative statics must reverse, meaning that the Gaussian comparative statics are not robust.

Consider the comparative statics with respect to γ. There are two forces that contribute to the unconventional result. The first one was highlighted in Section 1: when risk aversion is high, $f'(q)$ becomes not sensitive to q. Hence, price is not sensitive to s and there is no illiquidity in $\gamma \to \infty$ limit. The second one is linked to strategic behavior. The illiquidity arises because liquidity providers are risk-averse and demand lower prices to hold more of the risky asset. Because liquidity providers are strategic, they additionally reduce their demands by the amount $q\Lambda(q)$ to account for price impact (cf. equation (9)). If price impact is increasing with γ for all γ, this demand reduction gets bigger and bigger, while prices are getting smaller and smaller as γ increases. This means that at some point, reducing the demand by $q\Lambda(q)$ will push the prices below the lower bound of payoff and create arbitrage opportunities, which cannot happen in equilibrium. Put differently, when γ is high, there is not much room for liquidity providers to exercise their market power. Therefore, their market power (reflected in Λ) will be decreasing in γ, when γ is large enough. Similar intuition applies to other comparative statics results. Figure 1 provides an illustration.

We conclude that the comparative statics derived in the Gaussian model are not robust to more general payoff specifications. In particular, they change to the exact opposites when the distribution of δ has bounded support and when risk aversion is high enough.

3.4 The shape of price response function

The purpose of this section is to study how the prices in the imperfectly competitive market may be affected by a block sell or buy order s, which is measured by the *price response function*

$$\pi(s) = |P(s) - P(0)|, \tag{20}$$

which is the absolute value of the difference between the equilibrium price when the supply is s and the equilibrium price when the supply is zero. It measures the total reaction of the equilibrium price to a block of size s. Positive values s correspond to block sales, whereas negative values of s correspond block purchases. Empirically, the function $\pi(s)$ is found to be asymmetric (i.e., $\pi(s) \neq \pi(-s)$) and nonlinear, typically concave, function of s.[17] Our analysis will therefore focus on convexity and asymmetry of the price response function $\pi(s)$.

We commence our analysis with the case of Gaussian distribution.

Corollary 2. *Suppose that $\delta \sim N(\mu, \sigma^2)$. Then*

$$\pi(s) = \frac{L-1}{L(L-2)}\gamma\sigma^2 \cdot |s|.$$

Consequently, $\pi(s)$ is a linear and symmetric function of s.

As it is clear from above, with Gaussian distribution, the price response function is linear and symmetric function of order size s, which contradicts the empirical evidence that we summarize in Section 5. We examine the general case in Proposition 3 below. We derive our

[17] See the summary of empirical facts in Section 5.

analytical results in the two limiting cases of small enough and large enough γ.

Proposition 3. *Fix s. For small enough γ we have*

(i) ***(Asymmetry of** $\pi(s)$**)**: for $s > 0$,*

$$\operatorname{sign}(\pi(s) - \pi(-s)) = -\operatorname{sign}(\operatorname{skew}(\delta)) \textit{ and}$$

(ii) ***(Convexity of** $\pi(s)$**)**:*

$$\operatorname{sign}(\pi''(s)) = -\operatorname{sign}(\operatorname{skew}(\delta)s).$$

Fix s. For large enough γ we have

*(iii)**(Convexity of** $\pi(s)$**)**: $\pi''(s) < 0$, provided that $g'''(x)$ does not change sign for $|x|$ large enough.*

Fix γ. For large enough s we have

*(iv)**(Convexity of** $\pi(s)$**)**: $\pi''(s) < 0$, provided that $g'''(x)$ does not change sign for $|x|$ large enough.*

The proposition above demonstrates that higher moments, in particular, skewness, are linked to the convexity and asymmetry of price response function when liquidity providers' risk aversion is low. For example, it follows that price response to a sell order is concave when the skewness of payoff δ is positive. To understand the intuition, consider a benchmark economy in which higher moments play no role. It is identical to the initial economy, except that the asset's payoff is normally distributed with mean and variance equal to those in the initial economy. In the benchmark economy, the price impact is linear in s (cf. Corollary 2). Moreover, for a very small values of γ, the role of higher moments is negligible; therefore, the price impact function in the benchmark economy should be arbitrarily close to that in the initial economy. Consequently, the straight dashed line representing the price response in the

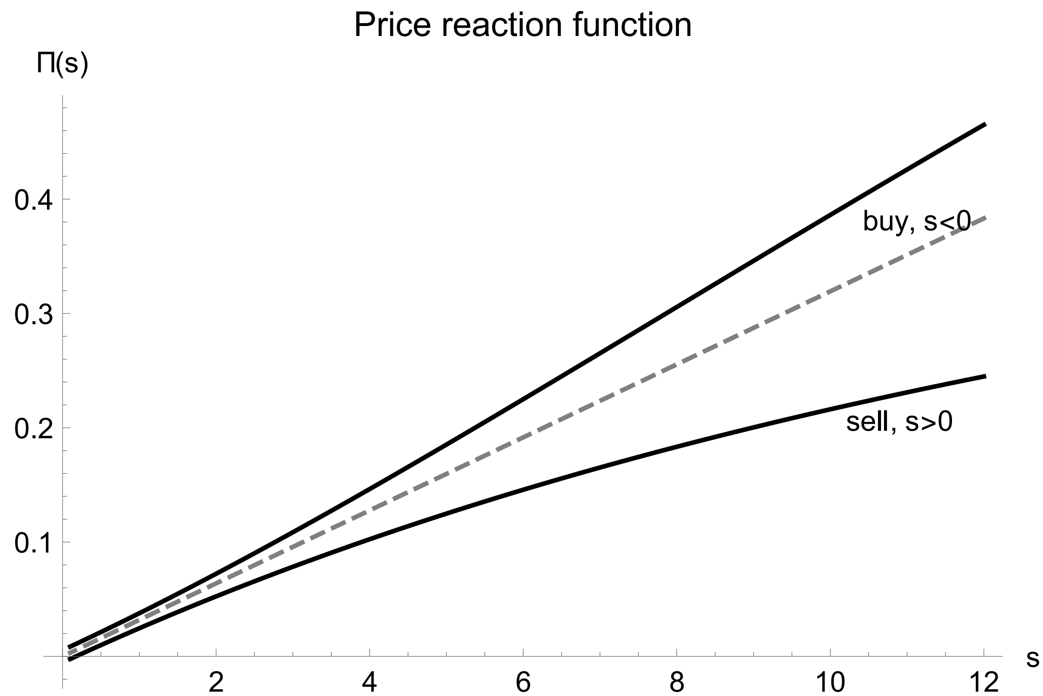

Figure 2: The figure represents the price response function for the case of a normal distribution $N(0, 1)$ truncated to the segment $[0, 2]$ (solid lines with buy and sell orders labeled accordingly). It compares the latter to the price response function for the case of a normal distribution (with mean 0.72 and variance 0.25 equal to that of the truncated normal distribution), represented by the dashed line. The other parameter values are $x_0 = .5$, $\gamma = 0.1$, and $L = 4$.

benchmark economy is tangent to the price response function in the initial economy (see Fig. 2). A concave function lies below its tangent line, therefore, it suffices to understand why the price impact in the benchmark economy is larger. The intuition is simple: with a positively skewed payoff, the trading profit of the investors accommodating the sale order is also positively skewed (i.e., they occasionally receive large positive surprises to their profits). Consequently, they require less price compensation relative to the case of zero skewness. Symmetric intuition implies that a price response to a buy order is convex when skewness is positive. Consequently, the price response to a sale order lies above the tangent line in Figure 2 and is therefore greater than that for a purchase. This explains the intuition for the asymmetry result.

The proposition above also demonstrates that under the mild condition, the price response function is concave for large enough s (part *(iv)*).[18] This is because of the role of payoff bounds. When s is large enough, the price is close to the payoff bounds. If price response function is convex, then prices are likely to be outside payoff bounds for larger supply shocks. This creates arbitrage opportunities that are not possible in equilibrium. Similar intuition applies to part (iii) of the proposition.

4 The case of multiple risky assets

In this section we consider the case of multiple risky assets and derive new implications regarding the cross-section of stock returns.

4.1 Characterization of equilibrium

We start with a heuristic derivation. Consider first a price-taking liquidity provider. His inverse demand $P = I(q)$ is determined by the first-order condition

$$\nabla f(q) = P.$$

A strategic trader accounts for the fact that he can move prices. With multiple assets, the price impact is a matrix, and its ij-th element measures impact of a trade in asset j on the price of asset i,

$$(\Lambda)_{ij} = \frac{\partial P_i}{\partial q_j}.$$

Suppose that each trader has a *conjecture* $\Lambda(q)$ about how he can move prices in equilibrium. The matrix $\Lambda(q)$ shows how much the trader of interest can move prices of different assets if he

[18] We could not find an example of a distribution with bounded support for which the condition that $g'''(s)$ does not change sign, for large enough $|x|$ does not hold.

trades a portfolio $q+dq$ instead of q. This conjecture determines his optimal (inverse) demand,

$$\nabla f(q) - \Lambda(q)q = P. \tag{21}$$

As in Section 3, the price impact $\Lambda(q)$ is pinned down by *consistency* condition. With multiple assets it means that price impact is a Jacobian of the price vector, which means that $\Lambda(q)$ is related to the Jacobian of $I(q)$, as follows:

$$\Lambda(q) = \frac{-1}{L-1}\nabla I(q). \tag{22}$$

Note the analogy between the equation above and (7).

The *optimality* (equation (21)) and *consistency* (equation (22)) conditions result in a system of PDEs,

$$\nabla f(q) + \frac{1}{L-1}\nabla I\left(q\right)q = I\left(q\right). \tag{23}$$

Relative to Section 3 we have an additional complication: the consistency and optimality conditions result in a system of PDEs, not a single ODE. However, in Section 4.2 we show that solving this system boils down to solving a linear ODE, similar to the one from Section 3. Thus, our approach remains tractable even in the case of multiple assets.

The following theorem summarizes the equilibrium characterization.

Theorem 2. *(Equilibrium characterization) A strictly decreasing function $I(q)$ is an equilibrium inverse demand if, and only if, it satisfies the following three conditions:*

(1) ***(Optimality)*** *The demand $I(q)$ is optimal given a conjecture about the price impact matrix $\Lambda(q)$,*

$$I(q) = \nabla f(q) - \Lambda(q)q, \tag{24}$$

(2) ***(Consistency)*** *The conjecture about the price impact matrix $\Lambda(q)$ is consistent with the*

equilibrium demand $I(q)$*,*

$$\Lambda(q) = -\frac{1}{L-1}\nabla I(q), \text{ and} \tag{25}$$

(3) ***(No free lunch)***

$$I(q) \in \mathcal{A}, \ \forall q. \tag{26}$$

4.2 Closed-form solution

Theorem 2 implies that to find equilibrium demand $I(q)$, one needs a solution to a system of PDEs (22) that is strictly decreasing and that satisfies the no-arbitrage restriction (26). In Proposition 4 below, we show that one can solve this system in two steps: (i) solve a single asset ODE similar to that from Section 3, with an asset being an index characterized by a vector of asset holdings q and (ii) differentiate this solution with respect to q. We provide a heuristic derivation for this two-step approach below.

Consider an economy in which all investors (including liquidity demanders) are restricted to trade a single index $q \in \mathbb{R}^N$. We refer to this economy as *restricted*, and to our initial as *unrestricted* economy. Denote inverse demand liquidity suppliers submit for t units of this portfolio by $\iota(t)$. The certainty-equivalent utility they derive from holding t units of this portfolio is $\phi(t) = f(tq)$. Our results for the case of a single risky asset (Section 3) apply in the restricted economy. Therefore, $\iota(t)$ has to satisfy the ODE (8), i.e.:

$$\frac{d}{dt}\phi(t) + \frac{t}{L-1}\frac{d\iota(t)}{dt} = \iota(t). \tag{27}$$

In the unrestricted economy, for supply realizations $s = tq$, $t \in \mathbb{R}$ the restriction to trade only portfolio q is non-binding in equilibrium.[19] Thus, what liquidity providers demand for t units of portfolio q in the unrestricted economy, i.e., $q^\top I(tq)$, should be an optimal demand in

[19]Indeed, in the symmetric equilibrium, it should be optimal for liquidity demanders to trade t/L units of portfolio q.

the restricted economy. It means that $q^\top I(tq) = \iota(t)$ should satisfy ODE (27).

We have established the first step to solving for equilibrium demand $I(q)$. One needs to solve (27) for $q^\top I(tq) = \iota(t)$. Note that $\iota(1) = q^\top I(q)$ is the expenditure $e(q)$ for portfolio q. Once $e(q)$ is known, we can derive the inverse demand by differentiating expenditure with respect to q. Indeed,

$$\nabla e(q) = I(q) + \nabla I(q) q.$$

Combining it with (22), we get

$$I(q) = 1/L \cdot \nabla e(q) + (1 - 1/L)\, \nabla f(q).$$

This establishes the second step in our two-step approach. We summarize in the proposition below.

Proposition 4. *(From PDE to ODE) The inverse demand $I(q)$ satisfies (23) if*

$$I(q) = 1/L \cdot \nabla e(q) + (1 - 1/L)\, \nabla f(q), \tag{28}$$

where $e(q) \equiv q^\top I(q)$ is trader's risky assets expenditure. The expenditure $e(q)$ can be found from $e(q) = \iota(1; q)$, where $\iota(t; q) \equiv q^\top I(tq)$ is the effective inverse demand for a portfolio q that satisfies the ODE

$$\iota(t; q) = \frac{d}{dt} f(tq) + \frac{t}{L-1} \frac{d\iota(t; q)}{dt}, \tag{29}$$

for every $t > 0$.

The first step in solving for equilibrium demand $I(q)$ is to solve the ODE (29). As in the previous section, there is only one solution to this ODE, such that $I(q) \in \mathcal{A}$. This solution is given by

$$\begin{aligned}\iota(t;q) &= (L-1)\int_1^\infty \xi^{-L}\phi'(t\xi)d\xi \\ &= q^\top\left((L-1)\int_1^\infty \xi^{-L}\nabla f(t\xi q)d\xi\right).\end{aligned}$$

Therefore, the expenditure is given by

$$e(q) = q^\top\left((L-1)\int_1^\infty \xi^{-L}\nabla f(\xi q)d\xi\right).$$

In the second step, one differentiates the equation above with respect to q and applies (28) to get the expression (30) for $I(q)$. The proposition below summarizes the closed-form solution.

Proposition 5. *(Closed-form solution) There exists unique equilibrium. The equilibrium inverse demand $I(q)$ and price impact matrix $\Lambda(q)$ are given by*

$$\begin{aligned}I(q) &= (L-1)\int_1^\infty \xi^{-L}\nabla f(\xi q)d\xi && (30)\\ &= \frac{L-1}{R_f}\int_1^\infty \xi^{-L}\nabla g(-\gamma\xi(x_0+q))d\xi && (31)\\ \Lambda(q) &= -\int_1^\infty \xi^{1-L}\nabla^2 f(\xi q)d\xi && (32)\\ &= \frac{\gamma}{R_f}\int_1^\infty \xi^{1-L}\nabla^2 g(-\gamma\xi(x_0+q))d\xi. && (33)\end{aligned}$$

4.3 Cross-section of returns and illiquidity

In this section we derive implications of our model for the cross-section of stock returns. The first-order condition (21) can be manipulated to produce the following illiquidity-adjusted consumption-CAPM.

Proposition 6. *Define* $Z = \frac{\exp\left(-\gamma(x_0+q)^\top\delta\right)}{E\left[\exp\left(-\gamma(x_0+q)^\top\delta\right)\right]}$. *The following is true:*

$$E[R_i] - R_f = -\text{cov}(Z, R_i) + R_f \lambda_{iq}^{\%}, \tag{34}$$

where $\lambda_{iq}^{\%} = \frac{1}{p_i} 1_i^\top \Lambda q$. *Moreover, one can write*

$$\lambda_{iq}^{\%} = -\text{cov}(Z_\lambda, R_i), \tag{35}$$

where

$$Z_\lambda = \frac{\gamma}{R_f} \int_1^\infty t^{1-L} E\left[\hat{Z}(t)\, q^\top \delta\right] \left(\hat{Z}(t) - \frac{\hat{Z}(t) q^\top \delta}{E\left[\hat{Z}(t)\, q^\top \delta\right]}\right) dt \tag{36}$$

and

$$\hat{Z}(t) = \frac{\exp\left(-\gamma t (x_0+q)^\top \delta\right)}{E\left[\exp\left(-\gamma t (x_0+q)^\top \delta\right)\right]}.$$

There exists a stochastic discount factor M *that prices all assets, i.e. for any asset* i*,*

$$1 = E[M R_i],$$

where

$$M = \frac{Z + Z_\lambda}{R_f E[Z + Z_\lambda]}. \tag{37}$$

The first term in (34) captures the consumption-CAPM expression in a frictionless market. It highlights the standard intuition that only the systematic risk, i.e., the part of return that covaries with the marginal utility of representative investor, affects stock returns. The second term, $\lambda_{iq}^{\%}$, captures the deviation from the consumption-CAPM due to market power of liquidity providers. The price impact $\lambda_{iq}^{\%}$ shows percentage change in price of asset i if liquidity providers trade one more unit of portfolio q. Equation (35) highlights that only systematic illiquidity, i.e., the part of return that covaries with Z_λ affects stock returns. It also highlights that commonality in illiquidity arises naturally in our model. Z_λ is the common factor driving the cross-section of illiquidity.

Equation (37) highlights that stochastic discount factor can be decomposed into two components. The first is proportional to Z, and hence is proportional to the SDF in a perfectly competitive economy. The second component, Z_λ arises due to illiquidity. As can be seen from (37), it is a transformation of a random variable $\hat{Z}(t)$, which is (up to a scalar multiplier) an SDF in a fictitious economy where liquidity providers have higher risk aversion ($t\gamma$, with $t > 1$, instead of γ). The intuition for this scaling up of risk aversion is as follows. The term Z_λ arises because liquidity providers exercise their market power. They do so by reducing their demands (see Ausubel, Cramton, Pycia, Rostek, and Weretka (2014)), similar to the way the demand of a more risk-averse trader is reduced compared to the demand of a less risk-averse trader. Overall, the stochastic discount factor M reflects the marginal utility of liquidity providers (in the "Z" component) and their demand-reduction incentives (in the "Z_λ" component).

To shed more light on the cross-section of expected returns $E[R_i]$ and illiquidity $\lambda_{iq}^{\%}$ we examine the equation (34) in the CARA-normal case.

Corollary 3. *Suppose that $\delta \sim N(\mu, \Sigma)$. Then*

$$E\left[R_i\right] - R_f = c_1 \text{cov}(R_i, R_{q+x_0}) + l_1 \text{cov}(R_i, R_q),$$

where

$$c_1 = \frac{\gamma}{p_{q+x_0}}, \qquad l_1 = \frac{\gamma}{(L-2)p_q}.$$

And $p_y = p^\top y$ denotes the price of portfolio y.

The first term, $c_1 \text{cov}(R_i, R_{q+x_0})$, is the standard CAPM term. Assets that covary positively with the $t = 1$ portfolio of liquidity providers, increase the variance of that portfolio and have to compensate for that with higher returns. The second term, $l_1 \text{cov}(R_i, R_q)$, is the illiquidity correction. A trade of portfolio q moves the price of asset i. In the competitive market, the price of asset i is driven by the covariance $\mathrm{cov}(\delta_i, \delta^\top(q+x_0))$. Strategic traders account for the impact of a marginal trade in the portfolio q on that covariance. This marginal effect is given by

$$\frac{\partial}{\partial t}\mathrm{cov}(\delta_i, \delta^\top(q(1+t)+x_0)) = \mathrm{cov}(\delta_i, \delta^\top q) = p_i\, p_q\, \mathrm{cov}(R_i, R_q).$$

Hence, the illiquidity correction term is proportional to the covariance between the R_i and R_q.[20] Since illiquidity is proportional to $\mathrm{cov}(R_i, R_q)$, the common factor driving the cross-section of illiquidity is R_q: assets that covary more with the return on the portfolio traded by liquidity demanders are more illiquid.

We now examine the equation (34) in the limit when γ is small.

Proposition 7. *We have*

$$\lambda_{iq}^{\%} = l_1 \mathrm{cov}(R_i, \hat{R}_q) - l_2 \mathrm{cov}(R_i, \hat{R}_q \cdot \hat{R}_{q+x_0}) + o\left(\gamma^2\right), \tag{38}$$

$$\mathrm{cov}(Z, R_i) = -c_1 \mathrm{cov}\left(R_i, \hat{R}_{q+x_0}\right) + c_2 \mathrm{cov}\left(R_i, \hat{R}^2_{q+x_0}\right) + o\left(\gamma^2\right), \tag{39}$$

Consequently,

$$E\left[R_i\right] - R_f = c_1 \mathrm{cov}(R_i, \hat{R}_{q+x_0}) - c_2 \mathrm{cov}\left(R_i, \hat{R}^2_{q+x_0}\right) + l_1 \mathrm{cov}(R_i, \hat{R}_q) - l_2 \mathrm{cov}(R_i, \hat{R}_q \cdot \hat{R}_{q+x_0}) + o\left(\gamma^2\right), \tag{40}$$

where

$$l_1 = \frac{\gamma}{(L-2)p_q}, \qquad l_2 = \frac{\gamma^2}{2(L-3)p_q \cdot p_{q+x_0}},$$

$$c_1 = \frac{\gamma}{p_{q+x_0}}, \qquad c_2 = \frac{\gamma^2}{2(p_{q+x_0})^2}.$$

The $\hat{R}_y$ denotes demeaned return, $\hat{R}_y = R_y - E[R_y]$.

[20] Rostek and Weretka (2015a) were the first to obtain this term.

There are two new terms in Equation (40). The first term, $c_2 \text{cov}\left(R_i, \hat{R}^2_{q+x_0}\right)$, is the correction for skewness. CARA investors have preference for skewness. The change in the skewness of payoff on portfolio $q + x_0$, if a marginal unit of asset i is added to it, is given by (assume for simplicity that all payoffs have zero mean)

$$\frac{\partial}{\partial t}\left(E[((q + x_0 + t1_i)^\top \delta)^3]\right)_{t=0} = 3E[((q + x_0)^\top \delta)^2 \delta_i] = 3p_i(p_{q+x_0})^2 \text{cov}(R_i, R^2_{q+x_0}).$$

Consequently, assets that have negative coskewness cov $\left(R_i, R^2_{q+x_0}\right)$ with the $t = 1$ portfolio decrease portfolio skewness and have to offer higher returns.

The second new term is the illiquidity correction $l_2 \text{cov}(R_i, \hat{R}_q \cdot \hat{R}_{q+x_0})$. It accounts for the fact that a trade in portfolio q changes the coskewness term just discussed, cov $\left(\delta_i, ((q + x_0)^\top \delta)^2\right)$. The marginal effect on this coskewness is given by (assuming for simplicity that all payoffs have zero mean)

$$\frac{\partial}{\partial t}\left(\text{cov}(\delta_i, ((q + x_0 + t1_i)^\top \delta)^2)\right)_{t=0} = 2\text{cov}(\delta_i, (q + x_0)^\top \delta \cdot q^\top \delta) = 2p_i p_q p_{q+x_0} \text{cov}(R_i, R_q R_{q+x_0}).$$

Thus, the marginal effect is proportional to $\text{cov}(R_i, R_q \cdot R_{q+x_0})$. Therefore, another common factor driving the illiquidity is $R_q \cdot R_{q+x_0}$.

5 Implications

Sections 3.3 and 3.4 provide implicaitons for equilibrium illiquidity and price response function. Section 4.3 provides implications for the cross-section of illiquidity and stock returns. We discuss them below.

Equilibrium illiquidity

The common wisdom about market illiquidity is that it is higher when liquidity providers are more risk-averse, the asset is riskier (given the information available), or liquidity suppliers are less numerous (see, e.g., a review paper of Vayanos and Wang (2012)). This wisdom is based on CARA-normal models commonly employed to analyze price impact. Our model with Gaussian distribution also produces the same result (Corollary 1).

However, with a general distribution, this wisdom is no longer robust. Proposition 2 shows that, while with small risk aversion γ, the conventional comparative statics hold, with γ large enough, they must necessarily reverse. This means that when risk aversion is high, i.e., in bad times, policy measures aimed at improving liquidity will have exactly the opposite effect. This is important, since the policy interventions are more likely to happen in bad times.

To test Proposition 2, one needs to have proxies for risk aversion and the competition between liquidity providers. The illiquidity can be measured as a regression coefficient of a price change on aggregate order imbalances (e.g., Sadka (2006)). Following Nagel (2012), one could use the VIX index as a proxy for liquidity providers' risk appetite in equity markets. Proposition 2 implies that when VIX is low, the changes in illiquidity are positively related to changes in VIX, but the opposite is possible when VIX is high.

The competition between liquidity providers can be measured by Herfindahl indices (e.g., Hasbrouck (2018)). Proposition 2 implies that when risk aversion (VIX) is low, the changes in illiquidity are negatively related to changes in Herfindahl index, but the opposite is possible when VIX is high.

Finally, Proposition 2 implies that when risk aversion (VIX) is low, the release of information about firm's fundamentals should lead to an improvement in its stock liquidity, but the opposite is possible when VIX is high. This can be tested by performing event studies around firms' announcements in high and low VIX periods.

The shape of price response function

We begin by summarizing the stylized facts concerning the shape of the price response function. These are as follows:

1. The price response function is an increasing and concave function of the order size.[21]

2. The price response function is asymmetric: the price impacts of sell and buy orders are different.[22]

Proposition 3 implies that for a large order size or large enough liquidity-providers' risk aversion, the price response function is concave, providing the explanation for the first stylized fact. The monotonicity of the two measures follows from Proposition 1. Proposition 3 also links the asymmetry of the price response function to the skewness of the distribution. The model predicts that with positive skewness (a natural property for stocks at the individual level (e.g., Chen, Hong, and Stein (2001)) price impact of purchases is greater than that of sells, consistent with evidence summarized by Saar (2001). While at the individual level stock returns are positively skewed, the skewness is negative at the aggregate level (e.g., Bakshi et al. (2003), Albuquerque (2012)). Our model implies that for indices the purchases move prices less than sells, consistent with the evidence in Chordia et al. (2002).

Chiyachantana, Jain, Jiang, and Wood (2004) find that the asymmetry of the permanent price impact is linked to the underlying market condition. Our model links the asymmetry to skewness, while Perez-Quiros and Timmermann (2001) present the evidence that skewness varies with the underlying market condition. The model predictions are therefore in line with these findings.

[21]Equities: Holthausen, Leftwich, and Mayers (1990), Keim and Madhavan (1996), and Almgren et al. (2005). Options: Muravyev (2016). For equities, a specification popular among practitioners is a linear permanent price impact and power-law temporary price impact, with the exponent being estimated as 1/2 (Plerou, Gopikrishnan, Gabaix, and Stanley (2002)) and 3/5 (Almgren et al. (2005)). Such a specification is employed in, e.g., Citigroup's BECS software (Almgren et al. (2005)). The price response function is a sum of temporary and permanent price impacts.

[22]Saar (2001) summarizes the evidence that shows bigger price impact of buy orders compared to sell orders.

For large order sizes, the bounds of the payoff represent another force affecting the curvature of price response function. One can examine the role of this force in the options market. For put options, the upside is limited, while this is not the case for calls. Consequently, for buy orders, the price response function should be more concave for puts relative to calls. Muravyev (2016) finds that the price response functions in options markets are concave, but, unfortunately, he did not estimate them separately for calls and puts.

Cross-section of illiquidity and stock returns

Proposition 7 shows that cross-section of returns is driven by four factors: (i) return on the portfolio of liquidity providers, R_{q+x_0}, (ii) the square of this return, $R^2_{q+x_0}$, (iii) the return on the portfolio of liquidity suppliers, R_q, and (iv) the product of (i) and (iii), $R_q R_{q+x_0}$. The last two factors are also explaining the cross-section of illiquidity. We discuss the potential empirical proxies for the two returns.

Consider the economy in which liquidity demanders are price-taking CARA traders who trade to share risks with liquidity providers. Suppose their risk aversion is very large. Then, the portfolio q will be equal to their initial endowment. The total endowment, or market portfolio in such economy is $q + x_0$. This motivates market return as a potential proxy for R_{q+x_0}.

Finding a proxy for the return R_q is challenging. Intuitively, R_q is a return on a portfolio that is proportional to aggregate order imbalances. Since the prices of stocks that liquidity demanders buy (and liquidity providers sell) typically go up and the prices of stocks that liquidity demanders sell (and liquidity providers buy) typically go down, the return R_q can be approximated by a return on reversal strategy of Nagel (2012) that buy stocks that went down over the prior days, and sell stocks that went up during the prior days.

Our analysis also helps to shed some light on the carry returns of Koijen et al. (2018). To see it, rewrite (34) as

$$E^Q[R_i] - R_f = R_f \lambda^{\%}_{iq}. \tag{41}$$

In the equation above $E^Q[\cdot]$ denotes the return under the risk-neutral measure (characterized by the Radon-Nikodym derivative Z). The left-hand side of (34) is related to carry return (the difference between the spot and forward prices normalized by the forward price) of Koijen et al. (2018). Indeed, rewriting their equation (8) in terms of notation of this paper yields

$$\text{carry} = \left(\underbrace{\frac{E^Q[\delta_i]}{P_i}}_{E^Q[R_i]} - R_f + 1 \right) \frac{P_i}{F_i}. \tag{42}$$

Moreover, Koijen et al. (2018) note that the last term, the ratio of spot and forward prices $\frac{P_i}{F_i}$ is close to one. Combining (41) and (42), we get

$$\text{carry} \approx R_f \lambda_{iq}^{\%} + 1.$$

That is, carry is positively related to asset illiquidity. In line with that finding, Koijen et al. (2018) find that carry is a strong positive predictor of returns and that carry strategies are positively exposed to global liquidity shocks.

6 Relation to the literature

Our paper is related to two broad strands of the literature: strategic trading and models of asset trading without normality. In our model, information is symmetric and price impact arises due to traders' limited risk-bearing capacity. We model trade using the classic double auction protocol (also known as the uniform price (divible) auction) in which traders submit price-contingent demand schedules. See, for example, Kyle (1989), Klemperer and Meyer (1989), Vayanos (1999), Wang and Zender (2002), Vives (2011), Rostek and Weretka (2012), Rostek and Weretka (2015b), Kyle, Obizhaeva, and Wang (2017), Ausubel et al. (2014), Bergemann, Heumann, and Morris (2015), Wang and Zender (2002), Du and Zhu (2017), and Lee and Kyle

(2018) for the single asset case, as well as Rostek and Weretka (2015a) and Malamud and Rostek (2017) for the multi-asset case.[23] Duffie and Zhu (2017) and Antill and Duffie (2017) consider models in which uniform price auction market is augmented with price-discovery sessions. All of these papers feature traders with marginal utilities that are linear in trade size (which is either assumed directly or follows from combination of CARA utility and normality of asset payoffs).[24] With exception of Du and Zhu (2017), they derive linear equilibria, whereby the slopes of the demand schedules are independent of the price level, and the equilibrium price impact (given by the inverse of the slope of the residual supply) is also constant, independent of the trade size. As noted before, such linear equilibria are hard to align with empirical evidence on nonlinearity and asymmetry of price response to order imbalances. Du and Zhu (2017) derive nonlinear equilibria in the two-agent case. With two agents, linear equilibria fail to exist. Du and Zhu (2017) show that there often exist nonlinear equilibria. This nonlinearity is not linked to higher moments, which is an important point of our paper; instead, in Du and Zhu (2017) it is linked to strategic behavior of traders. Overall, to the best of our knowledge, our paper is the first to derive closed-form solutions in a multi-asset double auction with nonlinear marginal utility and to link nonlinearities in equilibrium properties to higher moments of asset payoffs.[25]

There is a large body of literature on *competitive* trading with non-strategic liquidity providers beyond the CARA-normal setup. For example, several papers relax the assumption of normal payoff distributions but either maintain the CARA assumption or assume risk-neutrality. See, Gennotte and Leland (1990), Ausubel (1990b,a), Bhattacharya and Spiegel

[23] Sannikov and Skrzypacz (2016) develop an alternative trading protocol that they name "conditional double auction", where traders can condition their demand schedules on trading rates of other players.

[24] Bagnoli, Viswanathan, and Holden (2001) derive necessary and sufficient conditions for linear equilibrium in Kyle type models. They show that the linear equilibira are possible even without Gaussian distributions. They use a characteristic function approach. In contrast, we focus on nonlinear equilibria, the linearity is only possible in Gaussian case in our model and we employ a cumulant-generating function approach.

[25] Another class of strategic trading models assumes that strategic traders trade with market orders. See Kyle (1985), Bhattacharya and Spiegel (1991), Subrahmanyam (1991),Rochet and Vila (1994), Foster and Viswanathan (1996) and Vayanos (2001), among others. Rochet and Vila (1994) goes beyond the CARA-normal framework. They analyze a model a la Kyle (1985) without normality and prove uniqueness of equilibrium. They do not derive implications regarding cross-section of illiquidity and asset returns, price response asymmetry or comparative statics of illiquidity.

(1991), DeMarzo and Skiadas (1998, 1999), Yuan (2005), Breon-Drish (2015), Pálvölgyi and Venter (2015), Chabakauri, Yuan, and Zachariadis (2017), and Albagli, Hellwig, and Tsyvinski (2015). Peress (2003) and Malamud (2015) study noisy REE with non-CARA preferences. In all of these papers liquidity provision is competitive. In contrast, we assume that liquidity providers are strategic and demonstrate that it has important implications for the cross-section of stock returns.

A number of papers seek to explain the shape of the price impact. Roşu (2009) provides a model of the limit order book in which the key friction is costs associated with waiting for the execution of the limit orders. Keim and Madhavan (1996) explain concave price impact through a search friction in the upstairs market for block transactions. Saar (2001) provides an institutional explanation for the price impact asymmetry across buys and sells. Our paper adds to this literature by providing a unified treatment of the properties of the price reaction function and linking them to the shape of the probability distribution that describes asset payoffs.

Our paper is also related to the literature on transaction costs and asset prices. See Heaton and Lucas (1996), Vayanos (1998), Vayanos and Vila (1999), Lo, Mamaysky, and Wang (2004), Acharya and Pedersen (2005), and Buss and Dumas (2019). In contrast to these papers, transaction costs in our paper are endogenous. We also demonstrate that commonality in transaction costs (illiquidity) emerges endogenously in our model. Our paper also speaks to a literature on optimal dynamic execution algorithms under exogenous and non-constant price impact. See Bertsimas and Lo (1998), Almgren and Chriss (2001), Almgren et al. (2005), Huberman and Stanzl (2005), and Obizhaeva and Wang (2013). Our paper complements this literature by providing equilibrium foundations for nonlinear price functions.

We also mention a related strand of the literature that considers strategic liquidity provision and models trade using discriminatory price mechanisms. Notable examples are Biais, Martimort, and Rochet (2000) and Back and Baruch (2004), who also allow for non-Gaussian payoffs. An important difference between these papers and ours is that the liquidity providers in both Biais et al. (2000) and Back and Baruch (2004) are risk-neutral. Consequently, there is no inventory risk, which is the main focus of our model.

7 Conclusion

We present a tractable model of strategic trading in an economy populated by finite number of large, strategic CARA investors, and any finite number of assets with arbitrary (bounded) distribution of asset payoffs. We show that departing from the common (but unrealistic) assumption of normal payoffs has important economic implications for illiquidity and asset returns, both in the time-series and cross-sectionally. In particular, (i) the equilibrium price is nonlinear and the price-response function is asymmetric function of order size, (ii) liquidity is a non-monotonic function of risk aversion, of market power (number of large traders), and of uncertainty about future asset payoffs, and (iii) there is (endogenous) commonality in illiquidity that is affected by higher moments of asset payoffs. These results are consistent with extant empirical evidence about liquidity.

We develop a novel constructive approach to solve for the equilibrium in closed-form. In equilibrium, assets are priced according to the standard consumption Euler equation plus a correction term accounting for market liquidity (price impact). We show that solving for the equilibrium is equivalent to solving a linear ODE, which can be done using standard methods.

Our departure from the common CARA-normal assumption in strategic trading models can be extended along several important dimensions. We are currently working on exploring the equilibrium implications of wealth effects (that is, removing the CARA assumption) and heterogeneity in investors' wealth. Other interesting extensions include heterogeneity in investors' risk aversion to study risk sharing among strategic traders, strategic informed trading, and dynamic strategic trading.

Appendices

A A Summary of Notation

Notation	Explanation
	General mathematical notation
1_i	A vector with i-th element equal to 1 and all other elements being zero
$q^\top$	Transpose of a vector q
$\nabla f(q),$ where $f:\ \mathbb{R}^N \to \mathbb{R}$	Gradient of f, $(\nabla f)_l = \frac{\partial f}{\partial q_l}$
$\nabla^2 f(q),$ where $f:\ \mathbb{R}^N \to \mathbb{R}$	Hessian of f, $(\nabla^2 f)_{kl} = \frac{\partial^2 f}{\partial q_k \partial q_l}$
$\nabla I(q),$ where $I:\ \mathbb{R}^N \to \mathbb{R}^N$	Jacobian of I, $(\nabla f)_{ik} = \frac{\partial I^i}{\partial q_k}$
$a = \text{ess inf}(h(\delta))$	a is essential infimum of $h(\delta)$. Consider $h_l = \{\hat{a} \in \mathbb{R} : \hat{a} \leq\ h(\delta),\ \text{a.s.}\}$. Then $a = \sup h_l$ if $h_l \neq \emptyset$, and $a = -\infty$ otherwise.
$b = \text{ess sup}(h(\delta))$	b is essential supremum of $h(\delta)$. Consider $h_u = \{\hat{b} \in \mathbb{R} : \hat{b} \geq\ h(\delta),\ \text{a.s.}\}$. Then $b = \inf h_u$ if $h_u \neq \emptyset$, and $b = +\infty$ otherwise.
	Model variables

General note. Lowercase letters denote scalar-valued functions (e.g., $\iota(t;q)$ or $\lambda_{iq}(q)$) and uppercase letters denote vector- or matrix-valued functons (e.g., $I(q)$ or $\Lambda(q)$). We use subscripts to index assets/components of vector and superscripts to index traders (e.g., $I^i_k(q)$ is trader i's inverse demand for k-th asset, which is a k-th component of vector $I^i(q)$). The uppercase/lowercase distinction does not apply to arguments of functions (e.g., we use q, not Q for the argument of $I(q)$.)

Notation	Explanation
$I^i(q)$	Trader i's inverse demand. $I^i_k(q)$ is a price that a trader i bids for asset k, given that he gets allocation q.
$\iota^i(t;q)$	Trader i's effective inverse demand for a portfolio q, $\iota^i(t;q) = q^\top I^i(tq)$, is a price that a trader i bids for one unit of portfolio q, given that he gets allocation of t units of the portfolio q.
$P(s)$	Equilibrium price when the supply realization is s, $p(s) = I(s/L)$ in the symmetric equilibrium.

B Proofs

B.1 Proof of Theorem 1

The theorem follows from a more general Theorem 2.

B.2 Proof of Proposition 1

The proposition follows from a more general Proposition 5.

B.3 CARA Benchmark as a Limit

We analyze the single asset benchmark case with Gaussian distribution as the limit of our model with δ distributed according to a truncated normal distribution as the truncation bounds go to infinity. It suffices to show that Equations (13) and (15) converge as the truncation bounds go to infinity. The case of multiple assets follows along the same lines.

Assumption 6. *The random variable δ is a truncated normal random variable. That is, there exists $a, b \in \mathbb{R}$ with $a < b$ and a normal random variable X with mean μ and variance σ such that the random variable δ satisfies*

$$\delta \;\sim\; X \text{ conditional on } a < X < b\,.$$

For simplicity, set

$$R_f \;=\; 1\,.$$

It follows from Assumption 6 that

$$f(q) \;=\; (q + x_0)\mu \;-\; \frac{\gamma}{2}(q+x_0)^2\sigma^2 \;-\; \frac{1}{\gamma}\log\left[\frac{\operatorname{erf}\left(\frac{b-\mu+\gamma\sigma^2(q+x_0)}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{a-\mu+\gamma\sigma^2(q+x_0)}{\sqrt{2}\sigma}\right)}{\operatorname{erf}\left(\frac{\mu-a}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{\mu-b}{\sqrt{2}\sigma}\right)}\right].$$

Moreover

$$I(q) = \;\mu \;-\; \left[\frac{L-1}{L-2}q + x\right]\gamma\sigma^2 \;-\; \sqrt{\frac{2}{\pi}}\sigma(L-1)\int_1^\infty \xi^{-L}\frac{e^{-\frac{\left(b-\mu+\gamma\sigma^2(\xi q+x)\right)^2}{2\sigma^2}} - e^{-\frac{\left(a-\mu+\gamma\sigma^2(\xi q+x)\right)^2}{2\sigma^2}}}{\operatorname{erf}\left(\frac{b-\mu+\gamma\sigma^2(\xi q+x)}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{a-\mu+\gamma\sigma^2(\xi q+x)}{\sqrt{2}\sigma}\right)}d\xi$$

$$\Lambda(q) = \frac{\gamma\sigma^2}{L-2} \;-\; \frac{2}{\pi}\gamma\sigma^2\int_1^\infty \xi^{1-L}\left[\frac{e^{-\frac{\left(b-\mu+\gamma\sigma^2(\xi q+x)\right)^2}{2\sigma^2}} - e^{-\frac{\left(a-\mu+\gamma\sigma^2(\xi q+x)\right)^2}{2\sigma^2}}}{\operatorname{erf}\left(\frac{b-\mu+\gamma\sigma^2(\xi q+x)}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{a-\mu+\gamma\sigma^2(\xi q+x)}{\sqrt{2}\sigma}\right)}\right]^2 d\xi$$
$$+ \sqrt{\frac{2}{\pi}}\gamma\sigma\int_1^\infty \xi^{1-L}\frac{e^{-\frac{\left(a-\mu+\gamma\sigma^2(\xi q+x)\right)^2}{2\sigma^2}}(a-\mu+\gamma\sigma^2(\xi q+x)) - e^{-\frac{\left(b-\mu+\gamma\sigma^2(\xi q+x)\right)^2}{2\sigma^2}}(b-\mu+\gamma\sigma^2(\xi q+x))}{\operatorname{erf}\left(\frac{b-\mu+\gamma\sigma^2(\xi q+x)}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{a-\mu+\gamma\sigma^2(\xi q+x)}{\sqrt{2}\sigma}\right)}d\xi.$$

As the truncation bounds go to infinity, the Dominated Convergence Theorem, coupled with properties of the exponential function and error function (erf), imply that the equilibrium in our model converges to that in the benchmark case.

B.4 Proof of Proposition 2

Proof of Proposition 2.

For the first part of the proposition (i.e., when γ is small enough), differentiate with respect to corresponding parameter and take a limit as $\gamma \to 0$. Use the law of total variance to derive (19).

For the second part, note that

$$\Lambda(q;\gamma) = \frac{-\gamma}{(q+x_0)(L-1)} \frac{\partial}{\partial \gamma} I(q;\gamma). \tag{43}$$

We assume that $x_0 + q > 0$. The proof for the case $x_0 + q < 0$ is analogous.

One can write

$$\begin{aligned} I(s/L;0) - I(s/L;\infty) &= E[\delta] - a && (44)\\ &= (s/L + x_0) \int_0^\infty \frac{(L-1)}{L} \frac{\Lambda(s/L,\xi)}{\xi} d\xi && (45)\\ &= (s/L + x_0) \int_0^\infty \text{Illiquidity}(s,\xi)/\xi d\xi, && (46) \end{aligned}$$

where $a = \text{ess sup}(\delta)$. In (44) we used Lemma 6 to conclude that $I(s/L;\infty) = a$. In (45) we substituted (43). In (46) we substituted the definition of Illiquidity.

Step 1. $\frac{\partial \text{Illiquidity}}{\partial \gamma}$ changes sign at least once. Suppose not, i.e., Illiquidity always increases in γ. Then the right-hand side of (46) diverges whereas the left-hand side is bounded. A contradiction.

Step 2. $\frac{\partial \text{Illiquidity}}{\partial L}$ changes sign at least once. Differentiating both sides of (46) with respect to L, we get

$$0 = (s + x_0/L) \int_0^\infty \frac{\partial}{\partial L} \text{Illiquidity}(s;\xi,L)/\xi d\xi - \frac{x_0}{L^2} \int_0^\infty \text{Illiquidity}(s,\xi)/\xi d\xi.$$

Suppose not, i.e., $\frac{\partial}{\partial L}\text{Illiquidity}(s;\xi,L)$ is always less than zero. Then the right-hand side of the preceding displayed equation is negative, whereas the left-hand size is zero. A contradiction.

Step 3. The inequality $\text{Illiquidity}(\mathcal{F}_2) > \text{Illiquidity}(\mathcal{F}_1)$ *changes to the opposite at least once for* $\gamma \in (o,\infty)$ Write (46) for information set $\mathcal{F}_2$ and take expectation

$$\underbrace{E[E[\delta|\mathcal{F}_2]]}_{=E[\delta]} - a = (s + x_0/L)\int_0^\infty \text{Illiquidity}(s,\xi,\mathcal{F}_2)/\xi d\xi.$$

Subtract (46)

$$0 = (s + x_0/L)\int_0^\infty (\text{Illiquidity}(s,\xi,\mathcal{F}_2) - \text{Illiquidity}(s,\xi,\mathcal{F}_1))/\xi d\xi.$$

If $\text{Illiquidity}(s,\xi,\mathcal{F}_2) - \text{Illiquidity}(s,\xi,\mathcal{F}_1)$ is always negative, the right-hand side of the preceding displayed equation is negative, whereas the left-hand side is 0. A contradiction. ■

B.5 Proof of Proposition 3

Proof of Proposition 3.

Denote $\pi(s,\gamma)$ the equilibrium price in the economy with liquidity providers' risk aversion equal to γ, when the supply is s.

Part 1. $\text{sign}(\pi(s) - \pi(-s)) = -\text{sign}(\text{skew}(\delta))$. Recall that $p(s;\gamma) = I(s/L;\gamma)$. Differenting (14) with respect to γ we get

$$\begin{aligned}\frac{\partial}{\partial\gamma}\, p(s;\gamma)|_{\gamma=0} &= -\frac{L-1}{(L-2)R_f}(s/L + x_0)g''(0)\\ &= -\frac{L-1}{(L-2)R_f}(s/L + x_0)\text{var}(\delta) \text{ and}\end{aligned}$$

$$\frac{\partial^2}{\partial\gamma^2}\, p(s;\gamma)|_{\gamma=0} = \frac{L-1}{(L-2)(L-3)R_f}(s/L+x_0)^2 g'''(0)$$
$$= \frac{L-1}{(L-2)(L-3)R_f}(s/L+x_0)^2 \mathrm{skew}(\delta).$$

Applying Taylor's Theorem, we get

$$\pi(s;\gamma) = p(0;\gamma) - p(s;\gamma)$$
$$= g'(-\gamma x_0) - \frac{E[\delta]}{R_f} + \gamma\frac{(L-1)(s/L+x_0)}{(L-2)R_f}\mathrm{var}(\delta) - \frac{1}{2}\gamma^2\frac{(L-1)(s/L+x_0)^2}{(L-3)R_f}\mathrm{skew}(\delta) + o\left(\gamma^2\right).$$

Correspondingly,

$$\pi(s;\gamma) - \pi(-s;\gamma) = -\frac{1}{2}(x_0^2 + (s/L)^2)\gamma^2\frac{(L-1)}{(L-3)R_f}\mathrm{skew}(\delta) + o\left(\gamma^2\right),$$

from which the statement follows.

Part 2. $\mathrm{sign}(\pi''(s)) = -\mathrm{sign}(\mathrm{skew}(\delta)s)$.

Compute the second derivative of $\pi(s)$, for $s > 0$,

$$\pi''(s;\gamma) = -\gamma^2\frac{L-1}{R_f L^2}\int_1^\infty \xi^{2-L} g'''(-\gamma\xi(s/L+x_0))d\xi. \tag{47}$$

One can see that

$$\lim_{\gamma\to 0}\frac{\pi''(s;\gamma)}{\gamma^2} = -\frac{L-1}{R_f L^2(L-3)}\mathrm{skew}(\delta).$$

One can similarly get that for $s < 0$,

$$\lim_{\gamma\to 0}\frac{\pi''(s;\gamma)}{\gamma^2} = \frac{L-1}{R_f L^2(L-3)}\mathrm{skew}(\delta).$$

The claim follows.

Part 3. $\pi''(s) < 0$ *provided that* $g'''(x)$ *does not change sign for* $|x|$ *large enough.*

Note that $g'''(x)$ cannot be negative for $|x|$ large enough, because the function $g'(-x)$ is decreasing and bounded from below. Hence, we have that $g'''(x) > 0$ for $|x|$ large enough. From (47) we see that, provided that $g'''(x) > 0$, for γ large enough $\pi''(s, \gamma)$ is negative. ■

B.6 Proof of Theorem 2

Proof of Theorem 2. Given the equilibrium inverse demand $I(q)$, the inverse residual supply faced by trader i is given by $I\left(\frac{s-q^i}{L-1}\right)$, where q^i is the portfolio trader i would like to trade. Thus, trader's i ex-post optimization problem can be written as

$$\sup_{q^i} \left\{ f(q^i) - I\left(\frac{s-q^i}{L-1}\right)^\top q^i \right\}. \tag{$\mathcal{P}$}$$

The first-order condition yields

$$\nabla f\left(q^i\right) + \frac{1}{L-1} \nabla I\left(\frac{s-q^i}{L-1}\right) q^i = I\left(\frac{s-q^i}{L-1}\right). \tag{48}$$

In the symmetric equilibrium $q^i = s/L$ must be optimal for any s. Substituting $q^i = q = s/L$ to the above, we get the following system of PDEs:

$$\nabla f\,(q) + \frac{1}{L-1} \nabla I\,(q)\, q = I\,(q)\,. \tag{49}$$

The equilibrium inverse demand $I(q)$ must be a strictly decreasing solution to (49) such that $I(q) \in \mathcal{A}$. Lemma 5 states that there exists unique such solution $I(q)$ and provides a closed-form expression for $I(q)$. For such $I(q)$ Lemma 2 implies that there are only interior maxima in the problem ($\mathcal{P}$). Lemma 1 implies that the only such maximum is $q^i = s/L$. This implies that given $I(q)$ characterized in Lemma 5, the unique best response is $I(q)$. ■

Lemma 1. *Suppose that $I(q)$ is strictly decreasing and solves the system of PDEs (49), then $q = s/L$ is the unique solution to FOCs (48). Moreover, $q = s/L$ is a local maximum.*

Proof. Denote

$$\xi = \frac{s - q^i}{L - 1} \tag{50}$$

and rewrite (48) as follows:

$$\nabla f\left(q^i\right) + \frac{1}{L-1}\nabla I\left(\xi\right)q^i = I\left(\xi\right). \tag{51}$$

Instead of solving for $q^i(s)$ from (48), we will solve an equivalent system of equations (51) and (50).

Step 1. There is at most one solution to (51).

Indeed, suppose there are two solutions, q_1 and q_2. Then we can write

$$\nabla f\left(q_1\right) + \frac{1}{L-1}\nabla I\left(\xi\right)q_1 = I\left(\xi\right) \tag{52}$$

$$\nabla f\left(q_2\right) + \frac{1}{L-1}\nabla I\left(\xi\right)q_2 = I\left(\xi\right). \tag{53}$$

Multiply (52) and (53) by $(q_2 - q_1)^\top$ and subtract one equation from the other, as follows:

$$(q_2 - q_1)^\top(\nabla f\left(q_2\right) - \nabla f(q_1)) + \frac{1}{L-1}(q_2 - q_1)^\top \nabla I\left(\xi\right)\left(q_2 - q_1\right) = 0. \tag{54}$$

The first term in the preceding displayed equation, $(q_2 - q_1)^\top(\nabla f(q_2) - \nabla f(q_1))$, is negative. This is because $f(\cdot)$ is concave, hence ∇f is decreasing. The second term, $(q_2 - q_1)^\top \nabla I\left(\xi\right)\left(q_2 - q_1\right)$, is negative as well. This is because $I(\cdot)$ is decreasing, hence ∇I is negative-definite. Thus, we obtained a contradiction: the left-hand side of (54) is negative; the right-hand side is zero.

Step 2. The only solution to (51) *is* $q^i = \xi$.

Indeed, $q^i = \xi$ is a solution, since for such q^i, equation (51) becomes equation (49). By the previous step, there is at most one solution. Hence, $q^i = \xi$ is the only solution to (51).

Step 3. The only solution to (48) *is* $q^i = s/L$.

Indeed, (48) is equivalent to a system of equations (51) and (50). We know that the only solution to (51) is $q = \xi$. Therefore, the system of equations (51) and (50) becomes

$$q^i = \xi, \tag{55}$$

$$\xi = \frac{s - q^i}{L - 1}, \tag{56}$$

the unique solution to which is $q^i = s/L$.

Step 4. Portfolio $q^i = s/L$ is a local maximum.

We compute the hessian of of the investor's utility in ($\mathcal{P}$) and verify that it is negative-definite at $q^i = s/L$. Differentiating (48) and substituting $q^i = q^* \equiv s/L$, we get

$$\nabla^2 f(q^*) - \frac{1}{(L-1)^2} \nabla (\nabla I(q^*) x)|_{x=q^*} + \frac{2}{L-1} \nabla I(q^*),$$

where the partial derivatives in ∇ are taken with respect to the components of q^*. Differentiating (49), we get

$$\nabla^2 f(q^*) + \frac{1}{L-1} \nabla (\nabla I(q^*) x)|_{x=q^*} + \left(\frac{1}{L-1} - 1\right) \nabla I(q^*) = 0.$$

Combining the two preceding equations we get

$$\nabla^2 U = \left(\nabla^2 f(q^*) + \frac{\nabla I(q^*)}{L-1}\right) \frac{L}{L-1} < 0.$$

■

Lemma 2. *Given that $I(q)$ solves (49) and $I(q) \in \mathcal{A}$, there is no solution to problem ($\mathcal{P}$) at $q^i \to \infty$.*

Proof. Suppose not. Then there exists a sequence of portfolios $\{q_k\}_{k\in\mathbb{N}}$, such that $|q_k| \to \infty$ and the supremum in the problem ($\mathcal{P}$) is attained in the limit as $k \to \infty$. Let us rewrite q_k in the polar coordinates, so that $q_k = t_k\theta_k$, where $t_k = |q_k|$ and θ_k lives on the unit sphere in $\mathbb{R}^N$. Since the unit sphere is compact, the sequence $\{\theta_k\}_{k\in\mathbb{N}}$ contains a subsequence that converges to a point on a unit sphere. Thus, we can pass to such subsequence. By abuse of language, we call this subsequence θ_k and assume that it converges to a point θ_* on the unit sphere.

Denote

$$a \equiv \text{ess inf}(\delta^\top\theta_*) \qquad \text{and} \qquad b \equiv \text{ess sup}(\delta^\top\theta_*)\,.$$

This definition implies that $a \le b$. It follows from Assumption 1 that $a < b$ since equality holds if, and only if, $\delta^\top\theta_*$ is almost surely constant.

In Lemma 3 below we show that

$$\lim_{k\to\infty}\frac{1}{t_k}f(t_k\theta_k) = a.$$

In Lemma 6 we show that

$$\lim_{k\to\infty} I\left(\frac{s-t_k\theta_k}{L-1}\right)^\top \theta_k = b.$$

Therefore, the investor's utility in ($\mathcal{P}$) satisfies

$$\lim_{k\to\infty}\frac{U}{t_k} \;=\; \lim_{k\to\infty}\left(\frac{f(t_k\theta_k)}{t_k} - I\left(\frac{s-t_k\theta_k}{L-1}\right)^\top\theta_k\right) \;=\; a\;-\;b\;<\;0\,.$$

This inequality means that U goes to $-\infty$ as $t\to\infty$. A contradiciton. ∎

Lemma 3. *Suppose that $t_k \to \infty$ and $\theta_k \to \theta_*$. Then,*

$$\lim_{k\to\infty}\frac{1}{t_k}f(t_k\theta_k) \;=\; \text{ess inf}(\theta_*\cdot\delta)$$

Proof. For simplicity, we normalize $\gamma = 1$. We have

$$\frac{1}{t_k} f(t_k\theta_k) = -\frac{1}{t_k} \log E\left[e^{-t_k\left\{\text{ess inf}(\theta_k\delta) + \left(\theta_k\delta - \text{ess inf}(\theta_k\delta)\right)\right\}}\right]$$
$$= \text{ess inf}(\theta_k\delta) - \frac{1}{t_k} \log E\left[e^{-t_k\left(\theta_k\delta - \text{ess inf}(\theta_k\delta)\right)}\right].$$

Moreover, for any realization w, we have

$$\lim_{k\to\infty} e^{-t_k\left(\theta_k\delta - \text{ess inf}(\theta_k\delta)\right)}(w) \in \{0,1\} \qquad \text{and} \qquad \lim_{k\to\infty} \text{ess inf}(\theta_k\delta) = \text{ess inf}(\theta_*\delta).$$

The result then follows.

■

Lemma 4. *$p \in \mathcal{A}$ if, and only if, $\forall q$: $\text{ess inf}(q^\top\delta) < q^\top p$.*

Proof. Since

$$\text{ess inf}(q^\top\delta) < q^\top p$$

is equivalent to

$$\mathbb{P}((q^\top(\delta - p) < 0) > 0,$$

we have that $p \in \mathcal{A}$ if, and only if, $\forall q$: $\text{ess inf}(q^\top\delta) < q^\top p$. ■

Lemma 5. *The unique solution to (49) such that $I(q) \in \mathcal{A}$ is*

$$I(q) = (L-1)\int_1^\infty t^{-L}\nabla f(tq)dt. \tag{57}$$

Proof of Lemma 5. First note that by Lemma 4 $I(q) \in \mathcal{A}$ iff for any q

$$\text{ess inf}(q^\top\delta) < q^\top I(q). \tag{58}$$

Writing (58) for a portfolio tq as well as $-tq$, we also get

$$\text{ess inf}(q^\top \delta) < \iota(t;q) < \text{ess sup}(q^\top \delta), \tag{59}$$

which must hold for any t. According to Proposition 4, finding a solution to (49) amounts to solving linear ODE (29). This solution implies that $I(q) \in \mathcal{A}$ iff $\iota(t;q)$ is such that for any t, and any q (59) holds.

Step 1. Solving ODE (29).

We multiply both sides of (29) by the integrating factor t^{-L} so that the ODE becomes

$$\frac{d}{dt}\left(\frac{t^{1-L}}{1-L}\iota(t;q)\right) = t^{-L}\frac{d}{dt}f(tq).$$

Integrating the above from x to ∞ and noting that

$$\lim_{t\to\infty}\left(t^{1-L}\iota(t;q)\right) = 0,$$

which is true since (58) imply that $\iota(t;q)$ is bounded, we get a particular solution to (29)

$$\iota(x;q) = (L-1)\int_1^\infty \xi^{-L} q^\top \nabla f(x\xi q)d\xi.$$

The general solution is obtained by adding a general solution to the homogenous ODE $\frac{d}{dt}\left(\frac{t^{1-L}}{1-L}\iota(t;q)\right) = 0$, i.e., $\iota(t;q) = ct^{L-1}$. Thus, the general solution to (29) is given by

$$\iota(x;q) = (L-1)\int_1^\infty \xi^{-L} q^\top \nabla f(x\xi q)d\xi + ct^{L-1}, \tag{60}$$

for an arbitrary constant $c \in \mathbb{R}$.

Step 2. The solution (60) with $c=0$ implies $I(q) \in \mathcal{A}$.

It is easy to see that $\iota(x;q) = (L-1)\int_1^\infty \xi^{-L} q^\top \nabla f(x\xi q) d\xi$ is strictly decreasing in x and that ess $\inf(q^\top \delta) < \iota(0;q) = E\left[q^\top \delta\right] <$ ess $\sup(q^\top \delta)$. Therefore it suffices to prove that

$$\lim_{x\to\infty} \iota(x;q) \geq \text{ess } \inf(q^\top \delta).$$

Lemma (6) implies that $\lim_{x\to\infty} \iota(x;q) =$ ess $\inf(q^\top \delta)$, so that the last dispayed inequality holds.

Step 3. The solution (60) *with $c \neq 0$ implies $I(q) \notin \mathcal{A}$.*

A solution with $c \neq 0$ is unbounded as $t \to \infty$. For such a solution, (59) cannot hold.

Step 4. The solution (60) *with $c = 0$ implies $I(q) = (L-1)\int_1^\infty \xi^{-L} \nabla f(\xi q) d\xi$.*

Indeed, the solution (60) with $c = 0$ implies that

$$\begin{aligned}
e(q) &= (L-1)\, q^\top \int_1^\infty \xi^{-L} \nabla f(\xi q) d\xi \\
&= (L-1)\, \xi^{-L} f(\xi q)\big|_1^\infty + L(L-1) \int_1^\infty f(\xi q) \xi^{-L-1} d\xi \\
&= -\,(L-1)\, f(q) + L(L-1) \int_1^\infty f(\xi q) \xi^{-L-1} d\xi.
\end{aligned}$$

In the second line we noted that $q^\top \nabla f(\xi q) = \frac{d}{d\xi} f(\xi q)$ and integrated by parts. To get the third line, we noted that $\lim_{\xi\to\infty} \xi^{-L} f(\xi q) = 0$, which is true since Lemma (3) implies that $f(\xi q)$ grows slower than linear at infinity. We then applied (28) to get (57) ■

Lemma 6. *We have*

$$\lim_{k\to\infty} I\left(\frac{s - t_k \theta_k}{L-1}\right)^\top \theta_k \;=\; \text{ess } \sup(\theta_*^\top \delta)\,.$$

Proof.

$$I\left(\frac{s-t_k\theta_k}{L-1}\right)^\top \theta_k \tag{61}$$

$$= \; -L\int_1^\infty z^{-L-1}\nabla f\left(-zc\frac{s-t_k\theta_k}{L-1}\right)^\top \theta_k dz. \tag{62}$$

$$\tag{63}$$

We have

$$\nabla f(q)^\top q \;=\; \frac{E[(\delta^\top q)e^{-q^\top\delta}]}{E[e^{-q^\top\delta}]}.$$

Since δ has a bounded support, $f(q)$ is bounded, hence Lebesgue dominated convergence theorem implies that it suffices to prove the following lemma.

Lemma 7. *Suppose that $t_k \to +\infty$ and $\theta_k \to \theta_*$. Then,*

$$\lim_{k\to\infty}\frac{E[(\delta^\top\theta_k)e^{t_k\theta_k^\top\delta}]}{E[e^{t_k\theta_k^\top\delta}]} \;=\; \text{ess sup}(\theta_*^\top\delta).$$

Proof. First, let us pick a k large enough so that

$$\text{ess sup}(\delta^\top\theta_k) \;\le\; \epsilon \;+\; \text{ess sup}(\delta^\top\theta_*)\,.$$

Then, for all large k, we will have that

$$\frac{E[(\delta^\top\theta_k)e^{t_k\theta_k^\top\delta}]}{E[e^{t_k\theta_k^\top\delta}]} \;\le\; \epsilon \;+\; \text{ess sup}(\delta^\top\theta_*);$$

hence, since ϵ is arbitrary, we will always have that

$$\limsup_{k\to\infty}\frac{E[(\delta^\top\theta_k)e^{t_k\theta_k^\top\delta}]}{E[e^{t_k\theta_k^\top\delta}]} \;\le\; \text{ess sup}(\delta^\top\theta_*)$$

Now, let us pick an $\epsilon > 0$ and let K be large enough so that the subset

$$A_k \;=\; \{\delta :\; \theta_k^\top \delta \;\geq\; \text{ess sup}(\delta^\top \theta_k) \;-\; \epsilon\}$$

has a positive measure. Then,

$$E[(\delta^\top \theta_k) e^{t_k \theta_k^\top \delta}] \;>\; (b_k - \epsilon) E[e^{t_k \theta_k^\top \delta} \mathbf{1}_{A_k}],$$

where we have defined

$$b_k \;\equiv\; \text{ess sup}(\delta^\top \theta_k)\,.$$

Then,

$$E[e^{t_k \theta_k^\top \delta}] \;=\; E[e^{t_k \theta_k^\top \delta} \mathbf{1}_{A_k}] \;+\; E[e^{t_k \theta_k^\top \delta}(1 - \mathbf{1}_{A_k})] \;\leq\; E[e^{t_k \theta_k^\top \delta} \mathbf{1}_{A_k}] + e^{t_k (b_k - \epsilon)}\,. \tag{64}$$

Now, by the above, we know that

$$\lim_{k \to \infty} \frac{1}{t_k} \log E[e^{t_k \theta_k^\top \delta} \mathbf{1}_{A_k}] \;=\; b_*\,.$$

Pick a k large enough so that $b_k - \varepsilon < b_*$, and then pick k even larger so that $b_k - \epsilon < \frac{1}{t_k} \log E[e^{t_k \theta_k^\top \delta} \mathbf{1}_{A_k}] - \epsilon_1$ for some $\epsilon_1 > 0$. Then,

$$\frac{1}{t_k} \log \frac{e^{t_k (b_k - \epsilon)}}{E[e^{t_k \theta_k^\top \delta} \mathbf{1}_{A_k}]} \;<\; -\epsilon_1;$$

hence,

$$\frac{E[(\delta^\top \theta_k) e^{t_k \theta_k^\top \delta}]}{E[e^{t_k \theta_k^\top \delta}]} \;\geq\; \frac{E[(\delta^\top \theta_k) e^{t_k \theta_k^\top \delta} \mathbf{1}_{A_k}]}{E[e^{t_k \theta_k^\top \delta} \mathbf{1}_{A_k}] \;+\; e^{t_k (b_k - \epsilon)}}\,.$$

By the above, the right-hand side is asymptotically equivalent to

$$\frac{E[(\delta^\top \theta_k) e^{t_k \theta_k^\top \delta} \mathbf{1}_{A_k}]}{E[e^{t_k \theta_k^\top \delta} \mathbf{1}_{A_k}]} \;\geq\; b_k - \epsilon,$$

because on A_k we have $\delta^\top \theta_k > b_k - \epsilon$. ■ ■

B.7 Proof of Proposition 4

Proof of Proposition 4.

Step 1. PDE (23) implies ODE (29).

Note that $\frac{d}{dt}\iota(t;q) = q^\top \nabla I(tq) q$ and $\frac{d}{dt} f(tq) = q^\top \nabla f(tq)$. Then (29) can be rewritten as

$$q^\top I(tq) = q^\top \nabla f(tq) + \frac{1}{L-1} q^\top \nabla I(tq) tq,$$

which can be obtained from (23) by writing it for a portfolio tq and multiplying both sides of it by $q^\top$.

Step 2. Given an effective inverse demand for a portfolio q $\iota(t;q)$ the expenditure $e(q)$ can be found from $e(q) = \iota(1;q)$. Given the expenditure function $e(q)$, the inverse demand can be found from (28).

It follows from definitions of $e(q)$ and $\iota(t;q)$ that $e(q) = \iota(1;q)$. Adding $1/(L-1)I(q)$ to both parts of equation (23) and noting that $\frac{1}{L-1}\left(\nabla I\left(q\right) q + I(q)\right) = \nabla e(q)$ we get (28). ■

B.8 Proof of Proposition 5

Proof of Proposition 5. Equilibrium inverse demand is a solution to PDE (23), which is strictly decreasing and such that $I(q) \in \mathcal{A}$. Lemma 5 implies that there is unique such solution, given by (30) or, equivalently, (31). Expressions (32) and (33) are obtained by differentiating (30) and (31). ■

B.9 Proof of Proposition 6

Proof of Proposition 6. Take the first-order condition in the problem

$$\sup_q f(q) - p^\top q$$

with respect to q_l, l-th component of vector q. One gets

$$\frac{\partial}{\partial q_l} f(q) - p_l - \lambda_{lq} = 0. \tag{65}$$

Note that

$$\begin{aligned} \frac{\partial}{\partial q_l} f(q) &= \frac{E\left[\exp\left(-\gamma(x_0+q)^\top\delta\right)\delta_l/R_f\right]}{E\left[\exp\left(-\gamma(x_0+q)^\top\delta\right)\right]} \\ &= \frac{E^Q[\delta_l]}{R_f}. \end{aligned} \tag{66}$$

Combining (66) and (65), we get

$$\frac{E^Q[\delta_l]}{R_f} - p_l = \lambda_{lq}.$$

Rewriting the previous equation in terms of return R_l, we get

$$E^Q[R_l] - R_f = R_f \lambda_{lq}^{\%}. \tag{67}$$

Now note that

$$E\left[Z\delta_l/R_f\right] = \mathrm{cov}\left(Z, \delta_l/R_f\right) + E\left[\delta_l/R_f\right].$$

Thus,

$$E[R_l] - E^Q[R_l] = -\mathrm{cov}\left(Z, R_l\right). \tag{68}$$

Combining (68) and (67), we get (34).

To get (36), we start with (76) and use the Lemma 8 to get

$$
\begin{aligned}
\lambda_{lq} &= 1_l^\top \Lambda q \\
&= \frac{\gamma}{R_f} \int_1^\infty t^{1-L} 1_l \nabla^2 g\left(-\gamma t\left(x_0+q\right)\right) q dt \\
&= \frac{\gamma}{R_f} \int_1^\infty t^{1-L}\, \mathrm{rg}_{12}(0,0,-\gamma t; 1_l, q, q+x_0) dt \\
&= \frac{\gamma}{R_f} \int_1^\infty t^{1-L} \left(E\left[\hat{Z}(t)\, \delta_l\, q^\top \delta\right] - E\left[\hat{Z}(t)\, \delta_l\right] E\left[\hat{Z}(t)\, q^\top \delta\right]\right) dt \\
&= \frac{\gamma}{R_f} \int_1^\infty t^{1-L} E\left[\hat{Z}(t)\, q^\top \delta\right] \left(E\left[\frac{\hat{Z}(t) q^\top \delta}{E\left[\hat{Z}(t)\, q^\top \delta\right]}\, \delta_l\right] - E\left[\hat{Z}(t)\, \delta_l\right]\right) \\
&= \frac{\gamma}{R_f} \int_1^\infty t^{1-L} E\left[\hat{Z}(t)\, q^\top \delta\right] \left(E\left[\frac{\hat{Z}(t) q^\top \delta}{E\left[\hat{Z}(t)\, q^\top \delta\right]}\, \delta_l\right] - \operatorname{cov}\left(\hat{Z}(t), \delta_l\right) - E\left[\delta_l\right]\right) \\
&= \frac{\gamma}{R_f} \int_1^\infty t^{1-L} E\left[\hat{Z}(t)\, q^\top \delta\right] \left(\operatorname{cov}\left(\frac{\hat{Z}(t) q^\top \delta}{E\left[\hat{Z}(t)\, q^\top \delta\right]}, \delta_l\right) - \operatorname{cov}\left(\hat{Z}(t), \delta_l\right)\right) \\
&= \frac{\gamma}{R_f} \int_1^\infty t^{1-L} E\left[\hat{Z}(t)\, q^\top \delta\right] \operatorname{cov}\left(\frac{\hat{Z}(t) q^\top \delta}{E\left[\hat{Z}(t)\, q^\top \delta\right]} - \hat{Z}(t), \delta_l\right) \\
&= \operatorname{cov}\left(\frac{\gamma}{R_f} \int_1^\infty t^{1-L} E\left[\hat{Z}(t)\, q^\top \delta\right] \left(\frac{\hat{Z}(t) q^\top \delta}{E\left[\hat{Z}(t)\, q^\top \delta\right]} - \hat{Z}(t)\right), \delta_l\right).
\end{aligned}
$$

The fact that M is an SDF follows directly from the fact that Z_λ has mean 0 and Z has mean 1. ■

B.10 Proof of Proposition 7

We will use the following lemma.

Lemma 8. *Define*

$$\begin{aligned}\operatorname{rg}(t_x, t_y, t_z; x, y, z) &= g(t_x x + t_y y + t_z z)\\ &= \log E\left[\exp\left(\left(t_x x + t_y y + t_z z\right)^\top \delta\right)\right].\end{aligned}$$

Then, the following is true

$$\operatorname{rg}_1(0,0,0;x,y,z) = E\left[x^\top \delta\right],$$

$$\operatorname{rg}_{12}(0,0,0;x,y,z) = \operatorname{cov}\left(x^\top \delta, y^\top \delta\right),$$

$$\operatorname{rg}_{123}(0,0,0;x,y,z) = \operatorname{coskew}\left(x^\top \delta, y^\top \delta, z^\top \delta\right).$$

Denote $Z = \frac{E\left[\exp\left((t_x x+t_y y+t_z z)^\top \delta\right) x^\top \delta\right]}{E\left[\exp\left((t_x x+t_y y+t_z z)^\top \delta\right)\right]}$, *then,*

$$\operatorname{rg}_1(t_x, t_y, t_z; x, y, z) = x^\top \nabla g(t_x x + t_y y + t_z z) \tag{69}$$

$$= E\left[Z\, x^\top \delta\right]; \tag{70}$$

$$\operatorname{rg}_{12}(t_x, t_y, t_z; x, y, z) = x^\top \nabla^2 g(t_x x + t_y y + t_z z) y \tag{71}$$

$$= E\left[Z\, x^\top \delta\, y^\top \delta\right] - E\left[Z\, x^\top \delta\right] E\left[Z\, y^\top \delta\right]. \tag{72}$$

Proof of Lemma 8.

Note that

$$\operatorname{rg}_1(t_x, t_y, t_z; x, y, z) = x^\top \nabla g(t_x x + t_y y + t_z z), \tag{73}$$

$$= \frac{E\left[\exp\left(\left(t_x x + t_y y + t_z z\right)^\top \delta\right) x^\top \delta\right]}{E\left[\exp\left(\left(t_x x + t_y y + t_z z\right)^\top \delta\right)\right]} \tag{74}$$

$$= E\left[Z\, x^\top \delta\right], \tag{75}$$

where the subscript 1 indicates the derivative with respect to the first argument. It follows that

$$\mathrm{rg}_1(0,0,0;x,y,z) = E\left[x^\top \delta\right].$$

Note also that

$$\mathrm{rg}_{12}(t_x,t_y,t_z;x,y,z) = x^\top \nabla^2 g(t_x x + t_y y + t_z z) y.$$

It follows that

$$\mathrm{rg}_{12}(0,0,0;x,y,z) = \mathrm{cov}\left(x^\top \delta, y^\top \delta\right).$$

Differentiating (74), one can also get that

$$\mathrm{rg}_{12}(t_x,t_y,t_z;x,y,z) = E\left[Z\, x^\top \delta\, y^\top \delta\right] - E\left[Z\, x^\top \delta\right] E\left[Z\, y^\top \delta\right].$$

Similarly,

$$\mathrm{rg}_{123}(0,0,0;x,y,z) = \mathrm{coskew}\left(x^\top \delta, y^\top \delta, z^\top \delta\right).$$

■

Proof of Proposition 7.

Invoking Lemma 8, note that

$$\begin{aligned}
\lambda_{lq} &= 1_l^\top \Lambda q && (76)\\
&= \frac{\gamma}{R_f} \int_1^\infty t^{1-L} 1_l \nabla^2 g\left(-\gamma t\left(x_0 + q\right)\right) q dt && (77)\\
&= \frac{\gamma}{R_f} \int_1^\infty t^{1-L}\, \mathrm{rg}_{12}(0,0,-\gamma t; 1_l, q, q + x_0) dt. && (78)
\end{aligned}$$

Therefore,

$$\left.\left(R_f \frac{\lambda_{lq}}{\gamma}\right)\right|_{\gamma=0} = \frac{\mathrm{cov}(\delta_l, q^\top \delta)}{L-2}$$

$$\frac{d}{d\gamma}\left(R_f\frac{\lambda_{iq}}{\gamma}\right)\Big|_{\gamma=0} = -\left(\int_1^\infty t^{2-L}\,\mathrm{rg}_{123}(0,0,-\gamma t;1_l,q,q+x_0)dt\right)\Big|_{\gamma=0}$$
$$= -\frac{\mathrm{coskew}(\delta_l,\delta^\top q,\delta^\top(q+x_0))}{L-3}.$$

Thus,

$$\lambda_{lq} = \gamma\frac{\mathrm{cov}(\delta_l,q^\top\delta)}{L-2} - \frac{\gamma^2}{2}\frac{\mathrm{coskew}(\delta_l,\delta^\top q,\delta^\top(q+x_0))}{L-3} + o\left(\gamma^2\right).$$

For the second part, consider

$$\mathrm{cov}\left(\frac{\exp\left(-\gamma(x_0+q)^\top\delta\right)}{E\left[\exp\left(-\gamma(x_0+q)^\top\delta\right)\right]},\delta_l\right) = E\left[\frac{\exp\left(-\gamma(x_0+q)^\top\delta\right)}{E\left[\exp\left(-\gamma(x_0+q)^\top\delta\right)\right]}\delta_l\right] - E\left[\delta_l\right]$$
$$= \mathrm{rg}_1(0,0,-\gamma;1_l,q,q+x_0) - E\left[\delta_l\right]$$
$$= \mathrm{rg}(0,0,0;1_l,q,q+x_0) - E\left[\delta_l\right] -$$
$$-\gamma\,\mathrm{rg}_{13}(0,0,0;1_l,q,q+x_0)$$
$$+\frac{\gamma^2}{2}\mathrm{rg}_{133}(0,0,0;1_l,q,q+x_0) + o\left(\gamma^2\right)$$
$$= -\gamma\mathrm{cov}\left(\delta_l,(q+x_0)^\top\delta\right) + \frac{\gamma^2}{2}\mathrm{coskew}(\delta_l,(q+x_0)^\top\delta,(q+x_0)^\top\delta) + o\left(\gamma^2\right).$$

Noting that $\mathrm{coskew}(X,Y,Z) = \mathrm{cov}(X,\hat{Y}\hat{Z})$, and going from payoffs to returns, the proposition follows. ■

C Equilibrium price response asymmetry with many assets

We extend the analysis in Section 3.4 to the case of multiple assets. Consider a liquidity demander who purchases t units of a *portfolio* s. Due to illiquidity, a liquidity demander would have to pay an extra amount $|P_i(ts) - P_i(0)|$ for every unit of asset i in the portfolio s. The total extra cost, per unit of portfolio is thus

$$\pi(t;s) = |s^\top (P(ts) - P(0))|. \tag{79}$$

We call $\pi(t;s)$ the *price response function*. It measures the reaction of the price of a portfolio s to a purchase or sell of t units of this portfolio.

Our multi-asset model can speak to the empirical evidence concerning illiquidity and the shape of the price response function when the price-moving order is not concentrated in one stock. For example, our analysis applies to the case when s is an index.

In the proposition below, we show how the analysis in Section 3.4 extends to the case of multiple assets.

Proposition 8. *Fix s. For small enough γ, we have the following:*

(i) ***(Asymmetry of** $\pi(t;s)$**)**: for $t > 0$,*

$$\mathrm{sign}(\pi(t;s) - \pi(-t;s)) = -\mathrm{sign}(\mathrm{skew}(\delta^\top s)) \text{ and}$$

(ii) ***(Convexity of** $\pi(t;s)$**)**:*

$$\mathrm{sign}\left(\frac{\partial^2}{\partial t^2}\pi(t;s)\right) = -\mathrm{sign}(\mathrm{skew}(\delta^\top s)s).$$

Fix s. For large enough γ, we have the following:

(iii) ***(Convexity of** $\pi(t;s)$**)**: $\frac{\partial^2}{\partial t^2}\pi(t;s) < 0$, provided that $g'''(a(tq+x_0))$ does not change sign for $|a|$ large enough.*

Fix γ. For large enough s, we have the following:

(iv) ***(Convexity of** $\pi(t;s)$**)**: $\frac{\partial^2}{\partial t^2}\pi(t;s) < 0$, provided that $g'''(tq)$ does not change sign*

for $|t|$ large enough.

Proof of Proposition 8.

Define

$$\begin{aligned}\mathrm{rg}(t_x, t_y; x, y) &= g(t_x x + t_y y) \\ &= \log E\left[\exp\left((t_x x + t_y y)^\top \delta\right)\right],\end{aligned}$$

which is a CGF characterizing the joint distribution of $\left(x^\top \delta, y^\top \delta\right)$.

Note that

$$\begin{aligned}\mathrm{rg}_1(t_x, t_y; x, y) &= x^\top \nabla g(t_x x + t_y y), \\ &= \frac{E\left[\exp\left((t_x x + t_y y)^\top \delta\right) x^\top \delta\right]}{E\left[\exp\left((t_x x + t_y y)^\top \delta\right)\right]},\end{aligned}$$

where the subscript 1 indicates the derivative with respect to the first argument. It follows that

$$\mathrm{rg}_1(0, 0; x, y) = E\left[x^\top \delta\right].$$

Note also that

$$\mathrm{rg}_{12}(t_x, t_y; x, y) = x^\top \nabla^2 g(t_x x + t_y y) y.$$

It follows that

$$\mathrm{rg}_{12}(0, 0, 0; x, y, z) = \mathrm{cov}\left(x^\top \delta, y^\top \delta\right).$$

Similarly,

$$\mathrm{rg}_{122}(0, 0; x, y) = \mathrm{coskew}\left(x^\top \delta, y^\top \delta, y^\top \delta\right).$$

Part 1. $\mathrm{sign}(\pi(t; s) - \pi(-t; s)) = -\mathrm{sign}(\mathrm{skew}(\delta^\top s))$.

Note that

$$\begin{aligned}
\iota(t) &= q^\top I(tq) \\
&= \frac{L-1}{R_f} \int_1^\infty \xi^{-L} q^\top \nabla g(-\gamma \xi (x_0 + tq)) d\xi \\
&= \frac{L-1}{R_f} \int_1^\infty \xi^{-L} \operatorname{rg}_1(0, -\gamma\xi; q, tq + x_0) d\xi.
\end{aligned}$$

It follows that

$$\iota(t;0) = \frac{\operatorname{rg}_1(0,0;q,tq+x_0)}{R_f} = \frac{E[\delta^\top q]}{R_f},$$

$$\begin{aligned}
\frac{\partial}{\partial\gamma}\, \iota(t;\gamma)\big|_{\gamma=0} &= -\frac{L-1}{(L-2)R_f} \operatorname{rg}_{12}(0,0;q,tq+x_0) \\
&= -\frac{L-1}{(L-2)R_f} \operatorname{cov}(\delta^\top q, \delta^\top (tq + x_0)) \\
&= -\frac{L-1}{(L-2)R_f} \left(t\operatorname{var}(\delta^\top q) + \operatorname{cov}\left(\delta^\top q, \delta^\top x_0\right)\right), \text{ and}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial\gamma^2}\, \iota(t;\gamma)\big|_{\gamma=0} &= \frac{L-1}{(L-3)R_f} \operatorname{rg}_{122}(0,0;q,tq+x_0) \\
&= \frac{L-1}{(L-3)R_f} \operatorname{coskew}\left(q^\top\delta, (tq+x_0)^\top \delta, (tq+x_0)^\top \delta\right) \\
&= \frac{L-1}{(L-3)R_f} \operatorname{cov}\left(q^\top\delta, t^2\left(q^\top\left(\delta-\bar\delta\right)\right)^2 + 2tq^\top\left(\delta-\bar\delta\right)\cdot x_0^\top\left(\delta-\bar\delta\right) + \left(x_0^\top\left(\delta-\bar\delta\right)\right)^2\right).
\end{aligned}$$

Applying Taylor's Theorem we get, for $t > 0$,

$$
\begin{aligned}
\pi(t;\gamma) &= \iota(0;\gamma) - \iota(Lt;\gamma) \\
&= \gamma \frac{L-1}{(L-2)R_f} Lt\text{var}(\delta^\top q) - \\
&- \frac{\gamma^2}{2} \frac{L-1}{(L-3)R_f} \text{cov}\left(q^\top \delta, L^2 t^2 \left(q^\top \left(\delta - \bar{\delta}\right)\right)^2 + 2Ltq^\top \left(\delta - \bar{\delta}\right) \cdot x_0^\top \left(\delta - \bar{\delta}\right)\right) + o\left(\gamma^2\right).
\end{aligned}
$$

In the above we have used that $\iota(t;\gamma)$ decreases in t. Similarly,

$$
\begin{aligned}
\pi(-t;\gamma) &= \iota(-Lt;\gamma) - \iota(0;\gamma) \\
&= \gamma \frac{L-1}{(L-2)R_f} Lt\text{var}(\delta^\top q) + \\
&+ \frac{\gamma^2}{2} \frac{L-1}{(L-3)R_f} \text{cov}\left(q^\top \delta, L^2 t^2 \left(q^\top \left(\delta - \bar{\delta}\right)\right)^2 - 2Ltq^\top \left(\delta - \bar{\delta}\right) \cdot x_0^\top \left(\delta - \bar{\delta}\right)\right) + o\left(\gamma^2\right).
\end{aligned}
$$

Correspondingly,

$$
\pi(t;\gamma) - \pi(-t;\gamma) = -\frac{\gamma^2}{2} \frac{L-1}{(L-3)R_f} L^2 t^2 \text{skew}(\delta^\top q) + o\left(\gamma^2\right),
$$

from which the statement follows.

Part 2: $\text{sign}\left(\frac{\partial^2}{\partial t^2}\pi(t;s)\right) = -\text{sign}(\text{skew}(\delta^\top s)t)$.

Note that

$$
\begin{aligned}
\iota(t) &= q^\top I(tq) \\
&= \frac{L-1}{R_f} \int_1^\infty \xi^{-L} q^\top \nabla g(-\gamma\xi(x_0 + tq)) d\xi \\
&= \frac{L-1}{R_f} \int_1^\infty \xi^{-L}\, \text{rg}_1(-\gamma\xi t, -\gamma\xi; q, x_0) d\xi.
\end{aligned}
$$

Compute the second derivative of $\pi(t)$, for $t > 0$,

$$\pi''(t;\gamma) = -\frac{\partial^2}{\partial t^2}\iota(t/L;\gamma) \tag{80}$$
$$= -L^{-2}\gamma^2\frac{L-1}{R_f}\int_1^\infty \xi^{2-L}\,\mathrm{rg}_{111}(-\gamma\xi t, -\gamma\xi; q, x_0)d\xi. \tag{81}$$

One can see that

$$\lim_{\gamma\to 0}\frac{\pi''(s;\gamma)}{\gamma^2} = -\frac{L-1}{R_f L^2(L-3)}\mathrm{skew}(\delta^\top q).$$

One can similarly get that for $s < 0$,

$$\lim_{\gamma\to 0}\frac{\pi''(s;\gamma)}{\gamma^2} = \frac{L-1}{R_f L^2(L-3)}\mathrm{skew}(\delta^\top q).$$

The claim follows.

Part 3. $\frac{\partial^2}{\partial t^2}\pi(t;s) < 0$ *provided that* $g'''(a(tq + x_0))$ *does not change sign for* $|a|$ *large enough.*

Note that $g'''(tq)$ cannot be negative for $|t|$ large enough, because the function $g'(-tq)$ is decreasing and bounded from below. Hence, we have that $g'''(tq) > 0$ for $|t|$ large enough. From (80) we see that, provided that $g'''(-a(x_0 + tq)) > 0$, for γ large enough $\pi''(s,\gamma)$ is negative.

Part 4. $\frac{\partial^2}{\partial t^2}\pi(t;s) < 0$ *provided that* $g'''(tq)$ *does not change sign for* $|t|$ *large enough.*

As in part 3 we have that $g'''(tq) > 0$ for $|t|$ large enough. From (80) we see that, provided that $g'''(tq)) > 0$, for $|t|$ large enough $\pi''(s,\gamma)$ is negative.

■

References

Viral V Acharya and Lasse Heje Pedersen. Asset pricing with liquidity risk. Journal of financial Economics, 77(2):375–410, 2005.

Elias Albagli, Christian Hellwig, and Aleh Tsyvinski. A theory of asset prices based on heterogeneous information. 2015.

Rui Albuquerque. Skewness in stock returns: reconciling the evidence on firm versus aggregate returns. The Review of Financial Studies, 25(5):1630–1673, 2012.

Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. Journal of Risk, 3:5–40, 2001.

Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Direct estimation of equity market impact. Risk, 18(7):58–62, 2005.

Samuel Antill and Darrell Duffie. Augmenting markets with mechanisms. Technical report, National Bureau of Economic Research, 2017.

Lawrence M Ausubel. Insider trading in a rational expectations economy. The American Economic Review, pages 1022–1041, 1990a.

Lawrence M Ausubel. Partially-revealing rational expectations equilibrium in a competitive economy. Journal of Economic Theory, 50(1):93–126, 1990b.

Lawrence M Ausubel, Peter Cramton, Marek Pycia, Marzena Rostek, and Marek Weretka. Demand reduction and inefficiency in multi-unit auctions. The Review of Economic Studies, 81(4):1366–1400, 2014.

Kerry Back and Shmuel Baruch. Information in securities markets: Kyle meets glosten and milgrom. Econometrica, 72(2):433–465, 2004.

Mark Bagnoli, S Viswanathan, and Craig Holden. On the existence of linear equilibria in models of market making. Mathematical Finance, 11(1):1–31, 2001.

Gurdip Bakshi, Nikunj Kapadia, and Dilip Madan. Stock return characteristics, skew laws, and the differential pricing of individual equity options. The Review of Financial Studies, 16(1): 101–143, 2003.

Dirk Bergemann, Tibor Heumann, and Stephen Morris. Information and volatility. Journal of Economic Theory, 158:427–465, 2015.

Dimitris Bertsimas and Andrew W Lo. Optimal control of execution costs. Journal of Financial Markets, 1(1):1–50, 1998.

Utpal Bhattacharya and Matthew Spiegel. Insiders, outsiders, and market breakdowns. The Review of Financial Studies, 4(2):255–282, 1991.

Bruno Biais, David Martimort, and Jean-Charles Rochet. Competing mechanisms in a common value environment. Econometrica, 68(4):799–837, 2000.

Bradyn Breon-Drish. On existence and uniqueness of equilibrium in a class of noisy rational expectations models. The Review of Economic Studies, 82(3):868–921, 2015.

Adrian Buss and Bernard Dumas. The dynamic properties of financial-market equilibrium with trading fees. The Journal of Finance, 74(2):795–844, 2019.

Georgy Chabakauri, Kathy Yuan, and Konstantinos E Zachariadis. Multi-asset noisy rational expectations equilibrium with contingent claims. LSE working paper, 2017.

Joseph Chen, Harrison Hong, and Jeremy C Stein. Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. Journal of financial Economics, 61(3): 345–381, 2001.

Chiraphol N Chiyachantana, Pankaj K Jain, Christine Jiang, and Robert A Wood. International evidence on institutional trading behavior and price impact. The Journal of Finance, 59(2): 869–898, 2004.

Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Order imbalance, liquidity, and market returns. Journal of Financial economics, 65(1):111–130, 2002.

John H Cochrane. Asset pricing: Revised edition. Princeton university press, 2009.

Peter DeMarzo and Costis Skiadas. Aggregation, determinacy, and informational efficiency for a class of economies with asymmetric information. journal of economic theory, 80(1):123–152, 1998.

Peter DeMarzo and Costis Skiadas. On the uniqueness of fully informative rational expectations equilibria. Economic Theory, 13(1):1–24, 1999.

Songzi Du and Haoxiang Zhu. Bilateral trading in divisible double auctions. Journal of Economic Theory, 167:285–311, 2017.

Darrell Duffie and Haoxiang Zhu. Size discovery. The Review of Financial Studies, 30(4): 1095–1150, 2017.

F Douglas Foster and S Viswanathan. Strategic trading when agents forecast the forecasts of others. The Journal of Finance, 51(4):1437–1478, 1996.

Andrea Frazzini, Ronen Israel, and Tobias J Moskowitz. Trading costs of asset pricing anomalies. Fama-Miller working paper, pages 14–05, 2012.

Gerard Gennotte and Hayne Leland. Market liquidity, hedging, and crashes. The American Economic Review, pages 999–1021, 1990.

Campbell R Harvey and Akhtar Siddique. Conditional skewness in asset pricing tests. The Journal of Finance, 55(3):1263–1295, 2000.

Joel Hasbrouck. High-frequency quoting: Short-term volatility in bids and offers. Journal of Financial and Quantitative Analysis, 53(2):613–641, 2018.

Jerry A Hausman, Andrew W Lo, and A Craig MacKinlay. An ordered probit analysis of transaction stock prices. Journal of financial economics, 31(3):319–379, 1992.

John Heaton and Deborah J Lucas. Evaluating the effects of incomplete markets on risk sharing and asset pricing. Journal of political Economy, 104(3):443–487, 1996.

Robert W Holthausen, Richard W Leftwich, and David Mayers. Large-block transactions, the speed of response, and temporary and permanent stock-price effects. Journal of Financial Economics, 26(1):71–95, 1990.

Gur Huberman and Werner Stanzl. Optimal liquidity trading. Review of Finance, 9(2):165–200, 2005.

Donald B Keim and Ananth Madhavan. The upstairs market for large-block transactions: Analysis and measurement of price effects. The Review of Financial Studies, 9(1):1–36, 1996.

Paul D Klemperer and Margaret A Meyer. Supply function equilibria in oligopoly under uncertainty. Econometrica: Journal of the Econometric Society, pages 1243–1277, 1989.

Ralph S. J. Koijen and Motohiro Yogo. A demand system approach to asset pricing. Journal of Political Economy, 127(4):1475–1515, 2019. doi: 10.1086/701683. URL https://doi.org/10.1086/701683.

Ralph SJ Koijen, Tobias J Moskowitz, Lasse Heje Pedersen, and Evert B Vrugt. Carry. Journal of Financial Economics, 127(2):197–225, 2018.

Albert S Kyle. Continuous auctions and insider trading. Econometrica: Journal of the Econometric Society, pages 1315–1335, 1985.

Albert S Kyle. Informed speculation with imperfect competition. The Review of Economic Studies, 56(3):317–355, 1989.

Albert S Kyle, Anna A Obizhaeva, and Yajun Wang. Smooth trading with overconfidence and market power. The Review of Economic Studies, 85(1):611–662, 2017.

Augustin Landier, Guillaume Simon, and David Thesmar. The capacity of trading strategies. HEC Paris Research Paper No. FIN-2015-1089, 2015.

Jeongmin Lee and Albert S Kyle. When are financial markets perfectly competitive? 2018.

Andrew W Lo, Harry Mamaysky, and Jiang Wang. Asset prices and trading volume under fixed transactions costs. Journal of Political Economy, 112(5):1054–1090, 2004.

Semyon Malamud. Noisy arrow-debreu equilibria. Available at SSRN 2572881, 2015.

Semyon Malamud and Marzena Rostek. Decentralized exchange. American Economic Review, 107(11):3320–62, 2017.

Ian Martin. What is the expected return on the market? The Quarterly Journal of Economics, 132(1):367–433, 2017.

Dmitriy Muravyev. Order flow and expected option returns. The Journal of Finance, 71(2): 673–708, 2016.

Stefan Nagel. Evaporating liquidity. The Review of Financial Studies, 25(7):2005–2039, 2012.

Robert Novy-Marx and Mihail Velikov. A taxonomy of anomalies and their trading costs. The Review of Financial Studies, 29(1):104–147, 2015.

Anna A Obizhaeva and Jiang Wang. Optimal trading strategy and supply/demand dynamics. Journal of Financial Markets, 16(1):1–32, 2013.

Dömötör Pálvölgyi and Gyuri Venter. Multiple equilibria in noisy rational expectations economies. Available at SSRN 2524105, 2015.

Joel Peress. Wealth, information acquisition, and portfolio choice. The Review of Financial Studies, 17(3):879–914, 2003.

Gabriel Perez-Quiros and Allan Timmermann. Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities. Journal of Econometrics, 103 (1-2):259–306, 2001.

Vasiliki Plerou, Parameswaran Gopikrishnan, Xavier Gabaix, and H Eugene Stanley. Quantifying stock-price response to demand fluctuations. Physical Review E, 66(2):027104, 2002.

Jean-Charles Rochet and Jean-Luc Vila. Insider trading without normality. The review of economic studies, 61(1):131–152, 1994.

Marzena Rostek and Marek Weretka. Price inference in small markets. Econometrica, 80(2): 687–711, 2012.

Marzena Rostek and Marek Weretka. Dynamic thin markets. The Review of Financial Studies, 28(10):2946–2992, 2015a.

Marzena Rostek and Marek Weretka. Information and strategic behavior. Journal of Economic Theory, 158:536–557, 2015b.

Ioanid Roşu. A dynamic model of the limit order book. The Review of Financial Studies, 22 (11):4601–4641, 2009.

Gideon Saar. Price impact asymmetry of block trades: An institutional trading explanation. The Review of Financial Studies, 14(4):1153–1181, 2001.

Ronnie Sadka. Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. Journal of Financial Economics, 80(2):309–349, 2006.

Yuliy Sannikov and Andrzej Skrzypacz. Dynamic trading: Price inertia and front-running. Preprint, 2016.

Avanidhar Subrahmanyam. Risk aversion, market liquidity, and price efficiency. The Review of Financial Studies, 4(3):417–441, 1991.

Dimitri Vayanos. Transaction costs and asset prices: A dynamic equilibrium model. The Review of Financial Studies, 11(1):1–58, 1998.

Dimitri Vayanos. Strategic trading and welfare in a dynamic market. The Review of Economic Studies, 66(2):219–254, 1999.

Dimitri Vayanos. Strategic trading in a dynamic noisy market. The Journal of Finance, 56(1): 131–171, 2001.

Dimitri Vayanos and Jean-Luc Vila. Equilibrium interest rate and liquidity premium with transaction costs. Economic theory, 13(3):509–539, 1999.

Dimitri Vayanos and Jiang Wang. Theories of liquidity. Foundations and Trends® in Finance, 6(4):221–317, 2012.

Xavier Vives. Strategic supply function competition with private information. Econometrica, 79(6):1919–1966, 2011.

James JD Wang and Jaime F Zender. Auctioning divisible goods. Economic theory, 19(4): 673–705, 2002.

Kathy Yuan. Asymmetric price movements and borrowing constraints: A rational expectations equilibrium model of crises, contagion, and confusion. The Journal of Finance, 60(1):379–411, 2005.