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

Sergei Glebkin, Semyon Malamud, and Alberto Teguia

September 14, 2020

Abstract

We consider an economy populated by strategic CARA investors who trade multiple risky assets with arbitrarily distributed payoffs. Our solution method reduces finding the equilibrium to solving a linear ordinary differential equation. With non-Gaussian payoffs: (i) asymmetry and nonlinearity of the price response to order imbalances are linked to higher moments of returns, in line with stylized facts; (ii) liquidity may be reduced when risk aversion or uncertainty decreases; (iii) market illiquidity is proportional to its risk-neutral variance; and (iv) illiquidity of individual assets is proportional to the risk-neutral covariance between returns earned by liquidity providers and asset returns.

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, Craig Holden, Yunzhi Hu, Ravi Jaganathan, Mina Lee, Dong Lou, Peter Kondor, John Kuong, Albert (Pete) Kyle, Jiasun Li, Igor Makarov, Ian Martin, Milan Martinovic, Konstantin Milbrandt, Anna Obizhaeva, Olga Obizhaeva, Martin Oehmke, Joel Peress, Cameron Pfiffer, Rohit Rahi, 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 the market's inability to accommodate large trades without a change in price, has a large impact on the trading and pricing of financial assets. This illiquidity is not negligible, even for a market as developed as US equities, and it is a primary 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, thereby 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*—that is, while accounting for how their trades can move prices. Some investors (e.g., J.P. Morgan, 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 the 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? The literature on strategic trading addresses this question, often adopting a *CARA-normal* framework for tractability: traders have constant absolute risk aversion (CARA) utility functions, and asset payoffs are assumed to be normally distributed.[4] Such a framework has a number of shortcomings. First, CARA-normal models feature linear equilibria, where price is a linear function of order size and where purchases and sales have the same price impact. These predictions are hard to align with empirical evidence documenting

[1]For example, Koijen and Yogo (2019) estimate that—for the median US stock—the price impact of a 10% demand shock was consistently greater than 20% between 1980 and 2017. These authors also 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. They estimate that these shocks explain 81% of the cross-sectional variance of stock returns. That such shocks would not affect 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), Landier, Simon, and Thesmar (2015), and Novy-Marx and Velikov (2015) for estimates of the capacity of different asset pricing anomalies. All of these are based on the authors' estimates of price impact (illiquidity).

[3]Indeed, investors are price takers in all the models covered by Cochrane's (2009) popular textbook.

[4]Settings with risk-neutral traders (as in Kyle 1985) amount to a particular case of CARA utility characterized by a zero coefficient of risk aversion.

that prices react to large orders in an asymmetric and nonlinear way: purchases typically have a greater effect on prices than do sales, and the 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, most strategic trading models are restricted to the domain of individual securities and so are silent about what determines the cross section of illiquidity and stock returns.[7]

To address these shortcomings, we consider a tractable model of strategic trading that allows for multiple assets and a 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 conventional wisdom about illiquidity derived from CARA-normal models is not robust. More specifically, we show that liquidity can *decrease* when risk aversion decreases or when there is less uncertainty about future asset payoffs. Third, we establish that market illiquidity is proportional to its risk-neutral variance and also that illiquidity of individual assets is proportional to the risk-neutral covariance between asset returns and the returns earned by liquidity providers. Thus, equity illiquidity can be measured from options data. The main technical challenge is that, with a non-Gaussian distribution of payoffs, the traditional guess-and-verify approach no longer applies because it is not clear what the guess should be. We propose a novel constructive solution method that overcomes this difficulty and enables us to solve the model, in closed form, for any distribution.

We assume that CARA traders, whom we refer to as *liquidity providers*, exchange multiple risky assets for a riskless asset over one period—while accounting for the price impact.

[5]Hausman, Lo, and MacKinlay (1992), Almgren, Thum, Hauptmann, and Li (2005), and Frazzini et al. (2012) find concave price response functions (absolute value of price change as a function of order size) for equities; Muravyev (2016) presents the evidence for options. With regard to asymmetry, Saar (2001) summarizes the evidence documenting that buy orders have a greater price impact than do 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 VIX and SVIX indices is inconsistent with the 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.

Liquidity providers all have the same risk aversion coefficient and are symmetrically informed. The absence of information asymmetry implies that, in our setting, the unique source of price impact is inventory risk.[8] Trading is organized as a uniform-price double auction: traders simultaneously submit demand functions specifying the number of assets they want to buy as a function of the prices of all assets. All trades are executed at prices that clear the market. Our main innovation (as compared with previous research) is to allow for an arbitrary distribution of the risky asset payoffs under the sole restriction of bounded support.[9] In addition to liquidity providers, there are *liquidity demanders* who submit market orders. The only restriction we impose on liquidity demanders is that their aggregate order be independent of the asset payoff, from which it follows that they are uninformed. We express all equilibrium quantities as functions of aggregate liquidity demand.

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 a trader not knowing others' demand functions yet knowing, for each order size, their price impact matrix (i.e., how their trades move prices of all assets at the margin). This is an intuitive representation of the problem. Real-world traders typically have a market impact model that serves as an input to their optimal execution algorithm.[10] The equilibrium price impact matrix is pinned down by the requirement that it be consistent with the demand functions of all other traders. This consistency requirement yields a linear ordinary differential equation (ODE) in the single-asset case and a system of partial differential equations (PDEs) in the multi-asset case. It turns out that the single-asset ODE for the equilibrium price function can be solved in closed form for any probability distribution. A surprising result is that solving the system of PDEs

[8]For the sake of tractability, we abstract from asymmetric information. However, our focus on inventory risk is justified by empirical results documenting that risk is a dominant source of price impact in some markets (see e.g. Muravyev 2016 for evidence from the options market). Extending our model to allow price impact to have an asymmetric information component is left for future research.

[9]We can also handle distributions with unbounded support. In the benchmark case of a Gaussian distribution, for example, payoffs are naturally 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) were the first to derive such a representation, and their model is cast in the CARA-normal framework. We generalize their result to non-Gaussian distributions.

in the multi-asset case can be reduced to solving a single-asset ODE. This ODE characterizes the price function in an economy whose single asset is an index characterized by a vector q of asset holdings. The PDEs can be solved by differentiating the single-asset ODE's solution with respect to q. Hence we can also characterize the equilibrium in a multi-asset economy in closed form. We establish 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. We show that when the payoff of the portfolio traded by the liquidity demanders is positively skewed, purchases of this portfolio have a greater price impact compared to sales. It is intuitive that, when liquidity providers absorb a *sell* order, they receive a positively skewed profit—which they prefer because positive skewness implies that positive profit surprises are more likely than negative ones. Yet liquidity providers who absorb a *buy* order receive a negatively skewed profit, which (conversely) they do not like. As a result, they require a greater premium when absorbing the buy order. So with positive skewness, then, purchases have more of a price impact than do sales. Similarly, if skewness is negative, then sales affect prices more than do purchases.

Stock returns at the individual level are positively skewed, but the market's skewness is negative (Bakshi, Kapadia, and Madan 2003; Albuquerque 2012). Hence our analysis implies that, when liquidity demanders trade individual stocks, purchases move prices more than sales whereas the opposite is true when they trade the market portfolio. Both of these implications are consistent with empirical evidence.[11] We also show that similar results hold for *small* orders traded by liquidity demanders, where risk-neutral skewness substitutes for a physical one. In the limit of low risk aversion, the risk-neutral and physical moments are the same; therefore,

[11] Saar (2001) summarizes the evidence, in the case of individual stocks, for a greater price impact of buy orders versus sell orders. Chordia, Roll, and Subrahmanyam (2002) document the opposite effect at the market level.

in both cases our analyses link price response asymmetry to risk-neutral skewness.

Our second set of results examines equilibrium illiquidity (defined as the sensitivity of prices to supply shocks) in a single-asset version of the model. We show that expressions for illiquidity in the general case are remarkably similar to those in the Gaussian benchmark; one need make only the minor adjustment of substituting risk-neutral variance for the physical variance. Thus our model implies that illiquidity is proportional to risk-neutral variance. There is strong support for a positive relationship between illiquidity and risk-neutral variance in the data. For instance, Nagel (2012) documents that the VIX index, which captures the market's risk-neutral variance, exhibits a strong and positive relation with market illiquidity. Moreover, he finds that VIX has explanatory power beyond that of the market's physical variance.[12]

Despite the similarity of expressions, the comparative statics of illiquidity in the general case and in the Gaussian benchmark case can be profoundly different. In particular, we show that illiquidity can *increase* when liquidity providers are less risk averse or after disclosure of public information about asset's future payoffs—in contrast to CARA-normal models (see e.g. Vayanos and Wang 2012). This result is possible because risk-neutral variance decreases with an increase in the liquidity-providers' risk aversion γ. To understand why, suppose that liquidity providers must absorb a sell order and that the associated asset's minimal payoff is zero. For large γ, the mean of an asset's payoff under a risk-neutral measure (which is also the price of that asset) is close to zero. This result is intuitive because, when γ is large, liquidity providers cannot bear much risk and so are willing to buy the risky asset only at an extremely low price. In short: under the risk-neutral measure, all realizations of the asset payoff are above zero although the mean is close to zero. The implication is that the risk-neutral distribution is concentrated around zero, which in turn implies that its variance is small. As γ increases, the distribution becomes even more concentrated around zero and so its variance becomes smaller.

Our third set of results examines the determinants of the cross section of illiquidity and

[12] Nagel (2012) decomposes the VIX into conditional volatility (i.e., physical volatility) and a volatility risk premium (i.e., the residual) and establishes that *both* components contribute to the returns to liquidity provision strategies. Our analysis implies that illiquidity might be better explained by risk-neutral variance because that variance helps account for the effect of higher moments.

asset returns in a multi-asset version of our model. Starting with the cross section of returns, we show that—in an economy with strategic liquidity providers—a stochastic discount factor (SDF) exists and is given by a weighted average of SDFs in the economies with competitive liquidity providers who absorb orders of larger size. This finding, too, is intuitive. Note that liquidity providers exercise their market power by charging the price that competitive liquidity providers would charge to absorb a larger order, a manifestation of *demand reduction* that is common to auctions of strategic divisible goods (see Ausubel, Cramton, Pycia, Rostek, and Weretka 2014).

We then consider the cross section of illiquidity; we show that illiquidity of individual assets is proportional to the risk-neutral covariance between returns earned by liquidity providers and by asset returns. This result implies that illiquidity is measurable with options data and also speaks to the empirical content of option-implied correlations (Buss and Vilkov 2012). In the limit when the size of order traded by liquidity demanders is small, we show that "asset carry" (as defined in Koijen, Moskowitz, Pedersen, and Vrugt 2018) is proportional to the asset's illiquidity. In line with it, 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.

Throughout the paper we contrast our findings with the case of strategic liquidity providers—that is, to findings in the benchmark case of liquidity providers behaving competitively. We discover that our qualitative results are the same in both scenarios. However, market power reinforces some of these results while attenuating others. For example, if liquidity providers are strategic then the price response asymmetry is relatively more pronounced and thus the cross section of stock returns is more affected by illiquidity; yet the price response to order imbalances is less nonlinear (i.e., less concave).

The rest of our paper proceeds as follows. Section 2 presents the model. In Section 3, we solve for the equilibrium in the single-asset case and derive implications concerning illiquidity and the shape of the price response function. Section 4 considers equilibrium in the multi-asset case and derives implications for the cross section of illiquidity and asset returns. We summarize these implications in Section 5, and Section 6 reviews the related literature. We conclude in Section 7 with a brief summary and some suggestions for future research. Technical details are relegated to the 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 return of R_f. A stock k is a claim to a terminal dividend δ_k.

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

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

this function contains information on the distribution δ's moments as follows:

$$\frac{\partial g}{\partial y_i}(0) = E[\delta_i],$$
$$\frac{\partial^2 g}{\partial y_i \partial y_j}(0) = E[(\delta_i - E[\delta_i])(\delta_j - E[\delta_j])] \equiv \operatorname{cov}(\delta_i, \delta_j), \quad \text{and}$$
$$\frac{\partial^3 g}{\partial y_i \partial y_j \partial y_k}(0) = E[(\delta_i - E[\delta_i])(\delta_j - E[\delta_j])(\delta_k - E[\delta_k])] \equiv \operatorname{coskew}(\delta_i, \delta_j, \delta_k).$$

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

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

The following technical restrictions are imposed on the joint probability distribution of divi-

[13]Under more strict technical conditions on the distribution of δ than we shall impose, an equilibrium with $L = 2$ exists in our model. However, it is well known that if $L = 2$ then the equilibrium does not exist in an important benchmark—namely, when the distribution of δ is Gaussian (see 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).

dends.

Assumption 1. *The random variables $(\delta_1, \delta_2, \ldots, \delta_N)$ are linearly independent modulo constant. In other words, there exists no nontrivial linear combination of $(\delta_1, \delta_2, \ldots, \delta_N)$ that is almost surely constant.*

Assumption 1 simply requires that there be no redundant securities.

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

Assumption 2 is a natural one. Real-world investors are protected by limited liability, which implies that dividends (δ_i) are nonnegative; hence, there must be a lower bound. An upper bound is also natural when one considers that the resources of any firm are limited, which means that no asset can have an infinite payoff. However, our benchmark case with a 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 the truncation bounds approach infinity (see Appendix C). Our model can handle any distribution with unbounded support (for which the limit just described exists) in a similar way.

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. The following restrictions will be imposed on the net supply.

Assumption 3. *The net supply s is independent of the random vector δ of dividends.*

This assumption implies that one cannot learn about δ from s. Therefore, there is no information asymmetry between suppliers and demanders of liquidity.

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

Supply uncertainty is needed to rule out the extreme multiplicity of equilibria (cf. Klemperer and Meyer 1989; Vayanos 1999. As in Klemperer and Meyer, our Assumptions 3 and 4 imply that equilibrium quantities will depend on the realization of s but not on its distribution. Hence 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 number of noise traders whose random demands have full support and then taking a limit of the model as the number of noise traders approaches zero.[14]

The liquidity providers are identical and maximize the expected CARA utility from their terminal wealth W while accounting for their price impact. They are initially endowed with portfolio x_0. Suppose that all traders except trader i submit identical demand functions $D(p)$. Then the optimal demand $D^i(p)$ for trader i solves the following problem:

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

Before detailing our equilibrium concept, we shall define the set of arbitrage-free prices to be used in our definition of equilibrium.

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

Our concept of equilibrium, which is one of an arbitrage-free and symmetric Nash equilibrium with strictly decreasing demand (hereafter, simply "an equilibrium"), is defined formally as follows.

Definition 2. *A function $D(p)\colon \mathbb{R}^N \to \mathbb{R}^N$ is an equilibrium demand if the following statements hold. (i) For any $i = 1, 2, \ldots, L$, if traders $j \neq i$ submit demands $D^j(p) = D(p)$ then it is optimal for trader i to submit demand $D^i(p) = D(p)$; in other words, $D^i(p)$ solves problem (1). (ii) The function $D(p)$ is strictly decreasing—that is, $(D(p) - D(\hat{p}))^\top(p - \hat{p}) < 0$ for all $p \neq \hat{p}$.*

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

(iii) The function $D(p)$ is continuously differentiable, and the Jacobian ∇D is nondegenerate everywhere. Let $I(\cdot)$ denote the inverse of $D(\cdot)$.[15] *We also require that (iv) $I(q) \in \mathcal{A}$ for any q.*

Definition 2 (i) is simply a Nash equilibrium requirement. Parts (ii) and (iii) are technical; they ensure that the inverse demand, for which we solve when deriving the equilibrium, is well-defined. Part (iv) is required to ensure that the equilibrium is unique; then solving for the equilibrium amounts to solving an ODE and this requirement places a transversality condition on selecting a unique solution. The economic meaning of the condition (iv) is as follows. Suppose that, in addition to strategic liquidity providers, there is an arbitrarily small mass of competitive (price-taking) liquidity providers. Then, for prices that are *not* arbitrage-free, the price-taking liquidity providers would submit infinite demands and so the market would not clear. Hence there can be no equilibria when prices are not within $\mathcal{A}$. Thus requirement (iv) selects, among many potential equilibria, the one that is most robust to the presence of a vanishingly small number of competitive liquidity providers.

We also define the competitive benchmark—namely, an equilibrium in which liquidity providers take prices as given. We use a superscript c to denote equilibrium objects in this benchmark case.

Definition 3. *The competitive equilibrium demand $D^c(p)\colon \mathbb{R}^N \to \mathbb{R}^N$ is a solution to the problem $\max_D E[-\exp(-\gamma(\delta - R_f p)^\top (D + x_0))]$ for a given price vector p.*

Throughout the paper we use the following notation. At time $t = 0$, the 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 $f(q; x_0)$. By definition, $f(q; x_0)$ solves

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

[15] Part (ii) of this definition implies that D is bijective. Hence part (iii), when 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 certainty equivalent $f(q)$ is related to the CGF as follows:

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

We will often suppress the second argument, and simply write $f(q)$, provided that no confusion could arise.

3 The case of a single risky asset

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

3.1 Characterization of equilibrium

We first derive the equilibrium heuristically to show the intuition, after which we justify the derivation in Theorem 1. Consider first a competitive (price-taking) trader and his equilibrium demand. This trader solves

$$\max_q f(q) - pq. \tag{2}$$

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

$$f'(q) = p. \tag{3}$$

A strategic trader accounts for his ability to affect prices, and his first-order condition accordingly has a new term as follows:

$$f'(q) - \frac{\partial p}{\partial q} q = p. \tag{4}$$

A large trader who knows the price sensitivity $\frac{\partial p}{\partial q}$ can use the first-order condition above to solve for optimal demand.

Suppose that each trader has a *conjecture* $\Lambda(q) = \frac{\partial p}{\partial q}$ about how she can affect prices in equilibrium. This function reflects how much the focal trader can move prices by trading $q + dq$ instead of q. That conjecture, together with the first-order condition (4), determines her optimal (inverse) demand:

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

In a Nash equilibrium, the conjectures $\Lambda(q)$ must be *consistent* with the inverse demands of other traders. That is to say, $\Lambda(q)$ must be given by the slope of the inverse residual supply (as in Kyle (1989)).

Consistency implies a relationship between $\Lambda(q)$ and $I'(q)$. In a symmetric equilibrium there are $L - 1$ identical demands that contribute to the residual supply's slope, which is thus $-(L-1)\frac{1}{I'(q)}$. The negative sign accounts for the upward-sloping nature of residual supply, whereas demand is downward sloping. We also exploit that demand's slope is the reciprocal of inverse demand's slope. Hence the slope of the inverse residual supply is

$$\Lambda(q) = \frac{-1}{L-1} I'(q). \tag{6}$$

It follows that the equilibrium inverse demand should satisfy two conditions: optimality and consistency, or (respectively) Equation (5) and Equation (6). Observe that these two conditions result in a linear ODE,

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

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 foregoing derivation we also made implicit use of the following condition: 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 that prices be arbitrage-free, which translates into the no-free-lunch condition in our first, summary theorem.

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

(i) Optimality: The demand $I(q)$ is optimal (i.e., $D(p) = I^{-1}(p)$ solves (1)) given a conjecture about the price impact function $\Lambda(q)$. That is,

$$I(q) = f'(q) - \Lambda(q)q. \tag{8}$$

(ii) Consistency: The conjecture about the price impact function $\Lambda(q)$ is consistent with equilibrium demand $I(q)$; thus

$$\Lambda(q) = -\frac{1}{L-1}I'(q). \tag{9}$$

(iii) No free lunch:

$$I(q) \in \mathcal{A} \ \forall q. \tag{10}$$

We follow Rostek and Weretka (2015a) in representating equilibrium via optimality and consistency conditions because doing so nicely captures the decision making of real traders. As in real life, traders adopt a market impact model and determine their optimal bids (undertake optimal trade execution) in accordance with that model. In a Nash equilibrium, the market impact (price sensitivity) model is specified by the consistency condition. Our primary contribution, extending Rostek and Weretka (2015a), is to derive that consistency condition when price impact is a function of order size. The fixed-point condition then results in an ODE, not an algebraic equation. This representation is useful for solving the model. One can find an equilibrium demand by jointly solving (8) and (9), which reduces to a linear ODE that can be solved using standard methods. Condition (10) pins down this ODE's unique solution. We proceed in that fashion in Section 3.2.

We remark that Theorem 1 (and its multi-asset generalization, Theorem 2) gives both necessary and sufficient conditions for a function $I(q)$ to be equilibrium inverse demand. Establishing the global optimality of a candidate equilibrium inverse demand function $I(q)$ is a far from trivial problem, especially in the multi-asset case, that we address (using novel mathematical techniques) in Appendix B.10.

3.2 Closed-form solution

Theorem 1 implies that, in order to find equilibrium demand $I(q)$, one needs a solution to (7) that is strictly decreasing and also satisfies the no-arbitrage restriction (10).[16]

Using standard methods, one can write a generic solution to (7) for $q > 0$ as

$$I(q) = (L-1)\int_1^{\infty} \xi^{-L} f'(\xi q)\, d\xi + c q^{L-1} \tag{11}$$

with an arbitrary constant $c \leq 0$.[17] The condition $c \leq 0$ ensures that the solution is downward sloping.

For (11) to be an equilibrium inverse demand, it must satisfy the no-free-lunch condition; yet it is easy to show that solutions with $c < 0$ violate this condition. Indeed, these solutions are unbounded. Hence for q high enough, $I(q)$ will be smaller than the minimal payoff to the asset: for such a q, we have $I(q) \notin \mathcal{A}$. Our proof of Proposition 1 establishes that the solution with $c = 0$ does, in fact, satisfy the no-free-lunch condition. Therefore, if an equilibrium exists then (a) it is unique and (b) the equilibrium demand is given by (11) with $c = 0$. The following proposition summarizes our discussion so far.

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

[16]We claim no contribution due to our analyzing (7); similar ODEs were analyzed in Klemperer and Meyer (1989), Bhattacharya and Spiegel (1991), Wang and Zender (2002), and Boulatov and Bernhardt (2015). We present our derivation here for the sake of completeness.

[17]One way of obtaining (11) is to multiply both sides of (7) by the integrating factor q^{-L} and then to integrate both parts from q to ∞.

inverse demand $I(q)$ *and price impact* $\Lambda(q)$ *are given by:*

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

$$= \frac{L-1}{R_f}\int_1^\infty \xi^{-L} g'(-\gamma(\xi q + x_0))\, d\xi; \tag{13}$$

$$\Lambda(q) = -\int_1^\infty \xi^{1-L} f''(\xi q)\, d\xi \tag{14}$$

$$= \frac{\gamma}{R_f}\int_1^\infty \xi^{1-L} g''(-\gamma(\xi q + x_0))\, d\xi. \tag{15}$$

One can derive an expression for the equivalent martingale measure (EMM) in our economy. Doing so allows us to rewrite the equilibrium objects just described in a more compact way and also to gain additional insights. In what follows we use an asterisk (*) to denote those moments of δ that are evaluated under the equivalent martingale measure.

Corollary 1. *Let* $\zeta(t;q) = \frac{\exp(-\gamma(x_0+tq)\delta)}{E[\exp(-\gamma(x_0+tq)\delta)]}$. *Then the equilibrium inverse demand* $I(q)$ *and the price impact* $\Lambda(q)$ *can be written as*

$$I(q) = \frac{E[Z^*(q)\delta]}{R_f} = \frac{E^*[\delta]}{R_f} \quad \textit{and} \tag{16}$$

$$\Lambda(q) = \frac{E[(Z^{*c}(q) - Z^*(q))\delta]}{qR_f}, \tag{17}$$

respectively, where

$$Z^*(q) = (L-1)\int_1^\infty t^{-L}\zeta(t;q)\, dt \quad \textit{and} \quad Z^{*c} = \zeta(1;q). \tag{18}$$

Note that $\zeta(t;q)/R_f$ is an SDF in a competitive economy, where liquity providers absorb an order of size $t \times q$. Equation (18) shows that Z^*/R_f (i.e., the SDF in the economy with market power) is a weighted average of the SDFs in the competitive economies where liquidity providers absorb orders of size $t \times q$ with $t > 1$. This outcome is intuitive: liquidity providers exercise their market power by charging the price that competitive liquidity providers would charge for absorbing a larger order. Thus (18) manifests the *demand reduction* common to auctions of divisible goods (see Ausubel et al. 2014). The price impact $\Lambda(q)$ is affected by the difference, $Z^{*c}/R_f - Z^*/R_f$, between the SDF in the competitive benchmark and in the economy with market power. Here we have simply restated the first-order condition (8).

3.3 Equilibrium illiquidity

In this section we analyze equilibrium illiquidity by focusing on its comparative statics with respect to the model's parameters. In line with prior research (for a review, see Vayanos and Wang 2012), our measure of illiquidity is the effect of a first marginal unit traded by liquidity demanders at the equilibrium price:[18]

$$\text{IL} \equiv -\frac{\partial P}{\partial s}\bigg|_{s=0}. \tag{19}$$

A straightforward calculation yields that equilibrium illiquidity is related to the price impact of liquidity providers as follows:

$$\text{IL}(s) = \frac{L-1}{L}\Lambda(0).$$

We also define the illiquidity in our competitive benchmark. By (3), we can write

$$\text{IL}^c(s) \equiv -\frac{\partial P^c}{\partial s}\bigg|_{s=0} = -\frac{1}{L}f''(0).$$

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

[18]One can alternatively define illiquidity as the effect of the *last* marginal unit traded by liquidity demanders at the equilibrium price. That is, if liquidity demanders sell s^* units then one can define $\text{IL}^{\text{alt}}(s^*) \equiv -\frac{\partial P}{\partial s}|_{s=s^*}$. If this alternative definition were used, then Proposition 3 would remain unchanged whereas Proposition 2 would hold only for small s^*. Note that, thanks to linearity, the two definitions are equivalent in CARA-normal models. Real-world liquidity demanders can split their orders over time and trade small quantities in each period. We therefore believe that, from an empirical standpoint, the difference between these two definitions should be small.

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

$$\mathrm{IL} = \frac{\gamma(L-1)\sigma^2}{R_f L(L-2)} \quad \textit{and} \quad \mathrm{IL}^c = \frac{\gamma\sigma^2}{R_f L}. \tag{20}$$

Thus IL *and* IL^c *are each increasing in both* γ *and* σ^2*.*

This corollary accords with the prevailing wisdom about illiquidity. In particular, illiquidity is higher when uncertainty about the asset's payoff is higher or when liquidity providers' risk aversion is higher, as they have less capacity to bear the risk. In Proposition 2 we show that, for the general case, one need make only a minor adjustment to (20)—namely, substitute risk-neutral variance for the physical variance σ^2. However, this minor change may reverse the comparative statics: we show in Proposition 3 that the conventional wisdom holds only when risk aversion γ is small. In contrast, if risk aversion is high then Corollary 2's comparative statics are necessarily reversed.

Proposition 2. *The illiquidity can be written as*

$$\mathrm{IL} = \frac{\gamma(L-1)\mathrm{var}^*(\delta)}{R_f L(L-2)}, \ \mathrm{IL}^c = \frac{\gamma \mathrm{var}^*(\delta)}{R_f L} \tag{21}$$

here $\mathrm{var}^*(\delta)$ *is the risk-neutral variance computed under the EMM in the economy with zero supply shock, with Radon-Nikodym derivative given by* $Z^*(0)$*.*

Proposition 2 gives simple expressions for equilibrium illiquidity. Moreover, since the risk-neutral variance is measurable given option prices, these expressions can be taken to the data, as we discuss in the Section 5. The requirement that EMM used to compute the risk-neutral variance should correspond to the economy with zero supply shock, simply means that one measures the risk-neutral variance immediately before the supply shock. Despite the similarity between (21) and (20), the comparative statics in the Gaussian and general models can differ markedly, as we will demonstrate.

For the comparative statics with respect to the amount of uncertainty about the asset's payoff, we assume there is a public information release concerning δ. We model such release of information in a flexible way. Thus we assume that all traders start with an information set characterized by a sigma-algebra $\mathcal{F}_1$. Then, after new information is released, this information set is characterized by a larger sigma-algebra $\mathcal{F}_2$, $\mathcal{F}_1 \subset \mathcal{F}_2$. We impose the following restriction on the information structure.

Assumption 5. *The support of the distribution of δ conditional on any event in $\mathcal{F}_2$ is the same as the support of the unconditional distribution of δ.*

According to Assumption 5, the information released allows for updating the probabilities but not for updating the set of possible realizations of δ. This assumption is valid when, for example, $\mathcal{F}_1$ changes to $\mathcal{F}_2$ because of the announced signal $\delta + u$; here the noise u is independent of δ and has full support.

Proposition 3. *Denote by $\mathrm{IL}(\gamma, \mathcal{F}_i)$ the illiquidity when liquidity providers' risk aversion is equal to γ and when information is given by $\mathcal{F}_i$ in an equilibrium with market power, and let $\mathrm{IL}^c(\gamma, \mathcal{F}_i)$ be the competitive benchmark. Then, for γ small enough, we have*

$$\frac{\partial \mathrm{IL}(\gamma, \mathcal{F}_1)}{\partial \gamma} > 0 \quad \textit{and} \quad \frac{\partial \mathrm{IL}^c(\gamma, \mathcal{F}_1)}{\partial \gamma} > 0 \tag{22}$$

as well as the corresponding inequalities

$$\mathrm{IL}(\gamma, \mathcal{F}_1) > E[\mathrm{IL}(\gamma, \mathcal{F}_2)|\mathcal{F}_1] \quad \textit{and} \quad \mathrm{IL}^c(\gamma, \mathcal{F}_1) > E[\mathrm{IL}^c(\gamma, \mathcal{F}_2)|\mathcal{F}_1]. \tag{23}$$

Furthermore, if $x_0 \neq 0$ then the sign of both these inequalities is reversed for γ large enough.

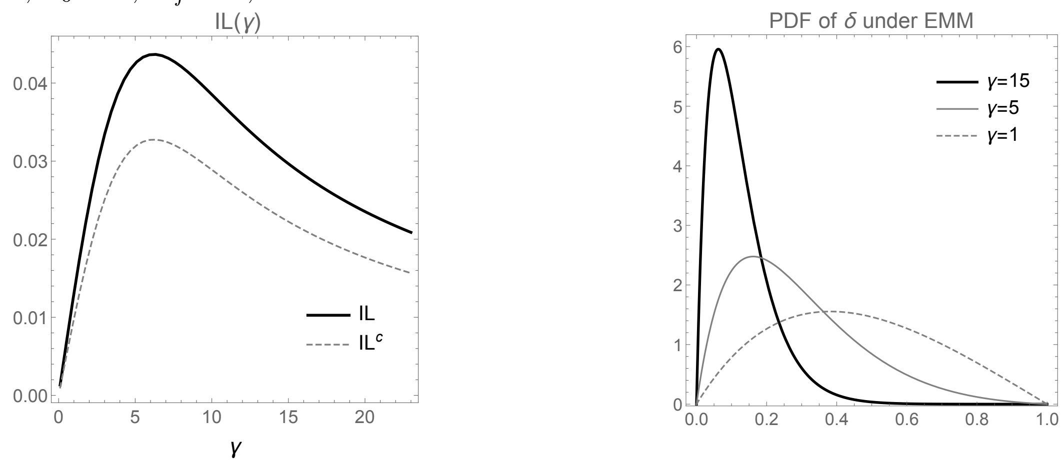
This proposition implies that, for sufficiently high levels of risk, the comparative statics in the Gaussian model change to their exact opposites when the distribution of δ has bounded support;[19] see Figure 1(a), which plots illiquidity as a function of γ. As the figure clearly shows,

[19]Proposition 3 states that the comparative statics change *at least* once. According to numerical experiments we performed with a variety of finite-support distributions (uniform, beta, Bernoulli, truncated normal), they change to their opposites *exactly* once.

this function is increasing for small γ but becomes decreasing for large enough γ.

Figure 1: Comparative statics of illiquidity and the risk-neutral distribution of δ.

Panel (a) plots illiquidity, as a function of γ, in an equilibrium with market power (IL) and in the competitive benchmark (IL^c); Panel (b) plots the probability density function (PDF) of δ under EMM with Radon–Nikodym derivative given by $Z^*(0)$. We assume that $\delta \sim \text{Beta}[a, b]$, $a = b = 2$, $x_0 = 1$, $R_f = 1$, and $L = 5$.



(a) Illiquidity in an equilibrium with market power (IL) and in the competitive benchmark (IL^c)

(b) PDF of δ under EMM for $\gamma = 1, 5, 15$

We look at Proposition 3's unusual comparative statics through the lens of expression (21). Take the comparative statics with respect to risk aversion γ; the terms that depend on it are γ and $\text{var}^*(\delta)$. Therefore, illiquidity depends directly on γ (the first term) yet also indirectly on changes in the risk-neutral distribution (second term). If γ is small then the risk-neutral and physical measures are close to each other and so the direct effect dominates, resulting in the conventional comparative statics. However, if γ is large then the risk-neutral variance decreases to zero (as we shall explain) and thereby generates the unconventional result.

Now we explain why the risk-neutral variance decreases to zero for large γ. For concreteness, assume that $s > 0$. Then liquidity providers must buy the asset, $R_f = 1$, and so we need not concern ourselves with discounting. We denote the lowest possible payoff (lower bound of the support of δ) as $\underline{\delta}$. Next, we make two observations. First, the distribution of δ under EMM and under the physical measure have the same supports. This makes sense when one consider that, under both measures, the maximal and minimal payoffs are the same. Second, if γ is large then the mean of δ under EMM (which is also the stock's price) is close to $\underline{\delta}$. This is likewise intuitive: if γ is large then liquidity providers cannot bear much risk, so they are willing to buy the asset only for a price at which they do not take a loss. Combining these two observations, we see that—under EMM—all realizations of δ are above $\underline{\delta}$ and the mean is close to $\underline{\delta}$, which implies that the distribution is concentrated around $\underline{\delta}$ and so its variance must be small (see Panel (b) of Figure 1).

3.4 Shape of the price response function

Our purpose in this section is to study how prices in an imperfectly competitive market may be affected by a block sell or buy order s. That effect is measured by the *price response function*

$$\pi(s) = |P(s) - P(0)|, \tag{24}$$

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. This function measures the total reaction of the equilibrium price to a block transaction of size s; positive(resp. negative) values of s correspond to block sales (resp. purchases). Empirical studies have shown that $\pi(s)$ is a nonlinear, asymmetric (i.e., $\pi(s) \neq \pi(-s)$), and typically concave function of s.[20] Hence our analysis will focus on the convexity and asymmetry of the price response function $\pi(s)$. We define that function for the competitive benchmark as follows:

$$\pi(s)^c = |P^c(s) - P^c(0)|. \tag{25}$$

Our analysis commences with the case of Gaussian distribution.

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

Corollary 3. *Suppose that* $\delta \sim N(\mu, \sigma^2)$. *Then*

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

As a result, both $\pi^c(s)$ *and* $\pi(s)$ *are linear functions of* $|s|$.

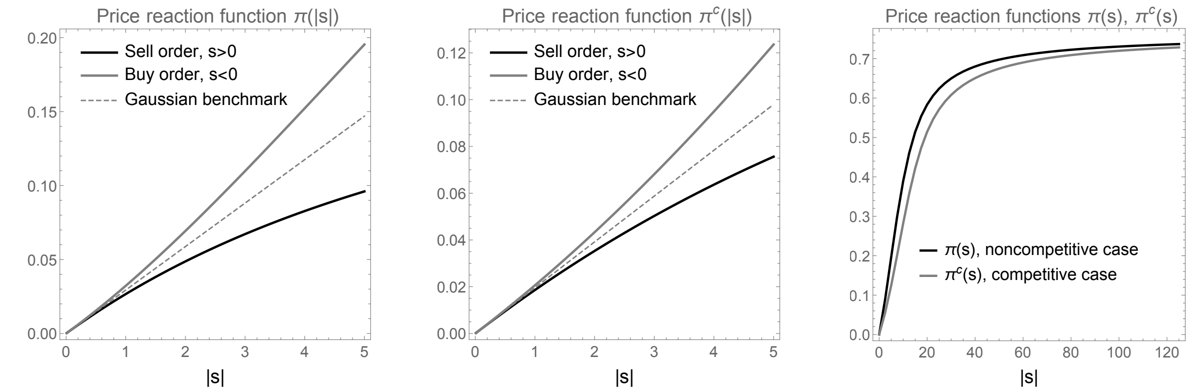
As it is clear from above, with Gaussian distribution, the price response function is linear in the order size, $|s|$. Yet this implies that the price response is linear and symmetric, which contradicts the empirical evidence summarized in Section 5. We examine the general case in Proposition 4 to follow. The analytical results that we derive depend on two variables: the size s of the supply shock; and risk aversion γ. We derive analytical results in the limiting cases of small and large supply shock size as well as for high and low levels of risk aversion γ. The limit of small (resp. large) s simply corresponds to small (resp. large) liquidity demand. In our model economy, the limit of small (large) γ corresponds to low (high) risk premia and so can be interpreted as good (crisis) times; see also Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016).

Proposition 4. *For small enough* γ *for a given* s, *or for small enough* $|s|$ *for a given* γ, *we have the following properties.*

(i) Asymmetry of $\pi(s)$*:* *Assume, without loss of generality, that* $s > 0$. *Then*

$$\text{sign}(\pi(s) - \pi(-s)) = \text{sign}(\pi^c(s) - \pi^c(-s)) = -\text{sign}(\text{skew}^*(\delta)),$$
$$|\pi(s) - \pi(-s)| > |\pi^c(s) - \pi^c(-s)|.$$

So both in equilibrium with market power and in the competitive benchmark, buys move prices more than sells when payoff skewness is positive; the converse holds when payoff skewness is negative. Also, asymmetry is more pronounced in the equilibrium with market power.



(a) Noncompetitive price response function, small order sizes

(b) Competitive price response function, small order sizes

(c) Competitive and noncompetitive price response functions, large order sizes

Figure 2: These graphs plot price response functions for the case of a beta distribution Beta$(1,2)$; the other parameter values are $x_0 = 1$, $\gamma = 2$, and $L = 4$.

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

$$\text{sign}(\pi''(s)) = \text{sign}((\pi^c)''(s)) = -\text{sign}(\text{skew}^*(\delta)s),$$
$$|\pi''(s)| > |(\pi^c)''(s))|.$$

That is: both in equilibrium with market power and in the competitive benchmark, the price response is convex for buys and concave for sells when payoff skewness is positive; when payoff skewness is negative, the opposite is true. In addition, convexity is more pronounced in the equilibrium with market power.

For large enough γ *for a given* s, $|s| > L|x_0|$, *or for large enough* $|s|$ *for a given* γ, *we have:*

(iii) Convexity of $\pi(s)$*: The inequalities* $(\pi^c)''(s) < \pi''(s) < 0$ *hold provided that* $g'''(x)$ *and* $g^{(4)}(x)$ *do not change sign for* $|x|$ *large enough. In other words, the price response is a concave function of the order size when risk aversion is large enough—and even more so in the competitive benchmark.*

This proposition demonstrates that higher moments (as in the case here of risk-neutral skewness) are linked to the price response function's convexity and asymmetry. Consider an instance of positive skewness. Proposition 4 implies that the price response to a sell order is a concave function of order size—not only in the competitive economy but also in the economy with market power. To appreciate the intuition, consider a benchmark Gaussian economy in which higher moments play no role and then examine how, in such economies, the price response depends on order size s. For concreteness, consider the case of small γ (so that we can ignore the difference between the physical and risk-neutral skewness). The Gaussian economy is identical to the original economy *except* that the asset's payoff is normally distributed with mean and variance adjusted so that, in the benchmark economy, the certainty equivalent is approximately equal to that in the original economy when s is small.[21] In the Gaussian economy, the price impact is linear in s (per Corollary 3). And for small order sizes s, the role of higher moments is negligible; hence the price impact function in the benchmark economy should be arbitrarily close to that in the original economy. This explains why, in Figure 2(a), the straight dashed line that represents the benchmark economy's price response is tangent to the price response function in the original economy.

A concave function has the characteristic property of lying below its tangent line. So in order to understand the concave shape of the original economy's price response function, it suffices to understand why the price impact there is smaller than that in the benchmark economy. The intuition is straightforward: when payoffs are positively skewed, the trading profit of investors who accommodate the sale order is also positively skewed—which they like because then they occasionally receive a strong boost to their profits. As a consequence, these investors require less price compensation for increasing their inventories relative to the case of zero skewness. By a symmetric intuition, the price response to a buy order is convex when skewness is positive. Hence the price response to a sell order lies above the tangent line in Figure 2 and is therefore greater than that for a buy order. These outcomes explain the intuition underlying our asymmetry result. The intuition showcased here (and in the preceding paragraph) applies in both the competitive and noncompetitive cases, as illustrated in (respectively) Panels (a)

[21]The marginal certainty equivalent (CE) in the original economy is $f'(q) = (1/R_f)g'(-\gamma(q + x_0)) \approx (1/R_f)g'(-\gamma x_0) - (\gamma q/R_f)g''(-\gamma x_0)$. For that CE to approximately equal the marginal CE in the benchmark Gaussian economy, or $(1/R_f)(\mu - \gamma\sigma^2 q)$, one must set the mean $\mu = g'(\gamma x_0)$ and the variance $\sigma^2 = g''(-\gamma x_0)$.

and (b) of Figure 2.

In our proof of this proposition we show that, for positive s, the price response is concave if the risk-neutral skewness is positive. In turn, the risk-neutral skewness is positive when either s or γ is large because then the risk-neutral distribution is concentrated at the lower bound (cf. Figure 1(b) and the discussion following Proposition 3), which implies that risk-neutral skewness is indeed positive.

Comparing our results for the competitive and noncompetitive cases reveals that, although the qualitative findings are much the same, market power reinforces parts (i) and (ii) of the proposition but attenuates part (iii). It is intuitive that parts (i) and (ii) are driven by liquidity providers demanding compensation for skewness. With market power, liquidity demanders can ask for greater compensation and thus reinforce these results. In contrast, part (iii) is driven by payoff bounds. In a noncompetitive economy, prices are closer to the payoff bounds and so liquidity providers have less room in which to exercise their market power; this dynamic leads to attenuation of the results.

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 she can move prices. When there are multiple assets, the price impact is a matrix whose (ij)th element measures the effect 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 she can move prices in equilibrium. Then the matrix $\Lambda(q)$ shows how much that trader can affect the prices of different assets if she trades a portfolio $q + dq$ instead of q. This conjecture determines her optimal (inverse) demand:

$$\nabla f(q) - \Lambda(q)q = P. \tag{26}$$

As in Section 3, the price impact $\Lambda(q)$ is determined by the consistency condition—which means that, when there are multiple assets, the price impact is a Jacobian of the price vector. Therefore, $\Lambda(q)$ is related to the Jacobian of $I(q)$ as follows:

$$\Lambda(q) = \frac{-1}{L-1}\nabla I(q). \tag{27}$$

Note the analogy between (27) and (6).

The *optimality* condition (26) and the *consistency* condition (27) result in the following system of equations:

$$\nabla f(q) + \frac{1}{L-1}\nabla I(q)q = I(q). \tag{28}$$

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

The next theorem summarizes our equilibrium characterization.

Theorem 2. *(Equilibrium characterization) A strictly decreasing function $I(q)$ is an equilibrium inverse demand iff it satisfies the following conditions.*

(i) Optimality: The demand $I(q)$ is optimal (i.e., $D(p) = I^{-1}(p)$ solves (1)) given a conjecture

about the price impact matrix $\Lambda(q)$,

$$I(q) = \nabla f(q) - \Lambda(q)q. \tag{29}$$

(ii) Consistency: The conjecture about the price impact matrix $\Lambda(q)$ *is consistent with the equilibrium demand* $I(q)$; *that is,*

$$\Lambda(q) = -\frac{1}{L-1}\nabla I(q). \tag{30}$$

(iii) No free lunch:

$$I(q) \in \mathcal{A} \ \forall q. \tag{31}$$

4.2 Closed-form solution

Theorem 2 implies that finding equilibrium demand $I(q)$ requires us to find a solution to a system of PDEs (27) that is strictly decreasing and that satisfies the no-arbitrage restriction (31). In Proposition 5, to follow, we show that this system can be solved in two steps: (i) solve a single-asset ODE similar to that from Section 3, where the asset is an index characterized by a vector q of asset holdings; and (ii) differentiate this solution with respect to q. We next provide a heuristic derivation based on this two-step approach.

Consider an economy in which all investors (including liquidity demanders) can trade only a single index $q \in \mathbb{R}^N$. We refer to this as a *restricted* economy and to our baseline case as an *unrestricted* economy. Let $\iota(t)$ denote the inverse demand that liquidity suppliers submit for t units of this portfolio. The certainty-equivalent utility they derive from holding those t units is

$$\phi(t) \equiv f(tq).$$

Our results for the case of a single risky asset (Section 3) apply as well in this restricted economy.

Hence $\iota(t)$ must satisfy the ODE (7); that is,

$$\frac{d}{dt}\phi(t) + \frac{t}{L-1}\frac{d\iota(t)}{dt} = \iota(t). \tag{32}$$

In the unrestricted economy and for supply realizations $s = tq$ ($t \in \mathbb{R}$), the restriction to trade only portfolio q is nonbinding in equilibrium.[22] So the price liquidity providers bid for t units of portfolio q in the unrestricted economy, or $q^\top I(tq)$, should be an optimal bid in the restricted economy. Hence $q^\top I(tq) = \iota(t)$ should satisfy ODE (32), which completes our first step.

The second step to solving for equilibrium demand $I(q)$ requires solving (32) for $q^\top I(tq) = \iota(t)$. Note that $\iota(1) = q^\top I(q)$ is the expenditure $e(q)$ for portfolio q (i.e., the dollar amount spent on buying the portfolio q). Once $e(q)$ is known, we can derive the inverse demand by differentiating the above definition of expenditure $e(q)$ with respect to q:

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

Combining this equality with (27) yields

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

completing the second of our two steps. This approach is summarized in our next proposition.

Proposition 5. *(From PDE to ODE) The inverse demand $I(q)$ satisfies the system (28) if*

$$I(q) = 1/L \cdot \nabla e(q) + (1 - 1/L)\, \nabla f(q). \tag{33}$$

Here $e(q) \equiv q^\top I(q)$ is the trader's expenditure on risky assets. This expenditure can be found from $e(q) = \iota(1; q)$, where $\iota(t; q) \equiv q^\top I(tq)$ is the inverse demand for a t units of portfolio q

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

that satisfies the ODE

$$\iota(t;q) = \frac{d}{dt} f(tq) + \frac{t}{L-1} \frac{d\iota(t;q)}{dt} \tag{34}$$

for every $t > 0$.

Thus the first step in solving for equilibrium demand $I(q)$ is to solve the ODE (34). As in Section 3.2, there is only one solution to this ODE such that $I(q) \in \mathcal{A}$. This solution is given by

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

Hence the expenditure can be written as

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

In the second step, we differentiate this equation with respect to q and then apply (33) to obtain (35) for $I(q)$. It remains then to establish the global optimality (1) of $D(p) = I^{-1}(q)$. In Appendix B.10 we use novel mathematical techniques to solve this extremely difficult problem. Our next proposition summarizes the closed-form solution.

Proposition 6. *<u>(Closed-form solution)</u> There exists a unique equilibrium. The equilibrium inverse demand $I(q)$ and the price impact matrix $\Lambda(q)$ are given by, respectively:*

$$\begin{aligned} I(q) &= (L-1) \int_1^\infty \xi^{-L} \nabla f(\xi q)\, d\xi && (35)\\ &= \frac{L-1}{R_f} \int_1^\infty \xi^{-L} \nabla g(-\gamma(x_0 + \xi q))\, d\xi; && (36)\\ \Lambda(q) &= -\int_1^\infty \xi^{1-L} \nabla^2 f(\xi q)\, d\xi && (37)\\ &= \frac{\gamma}{R_f} \int_1^\infty \xi^{1-L} \nabla^2 g(-\gamma(x_0 + \xi q))\, d\xi. && (38)\end{aligned}$$

4.3 Cross section of returns and illiquidity

Here we derive some implications of our model for the cross section of stock returns and illiquidty.

As in the single-asset case, we define the illiquidity of asset i as the effect of the first marginal unit of portfolio s traded by liquidity demanders at the equilibrium price i:

$$\mathrm{IL}_i \equiv -\frac{\partial P_i(ts)}{\partial t}\bigg|_{t=0}; \mathrm{IL}_i^c \equiv -\frac{\partial P_i^c(ts)}{\partial t}\bigg|_{t=0}.$$

Our next corollary characterizes the cross section of returns and illiquidity in the CARA-normal benchmark.

Corollary 4. *Put $q = s/L$, and suppose that $\delta \sim N(\mu, \Sigma)$. Then the cross section of returns is characterized by*

$$E[R_i] - R_f = c_1 \mathrm{cov}(R_i, R_{q+x_0}) + l_1 \mathrm{cov}(R_i, R_q), \tag{39}$$

where $c_1 = \gamma p_{q+x_0}$, $l_1 = \frac{\gamma p_q}{L-2}$, and $p_y = P^\top y$ denotes the price of portfolio y. In the competitive benchmark we have

$$E[R_i] - R_f = c_1 \mathrm{cov}(R_i, R_{q+x_0}). \tag{40}$$

The cross section of illiquidity is characterized by

$$\mathrm{IL}_i = \gamma \frac{L-1}{R_f L(L-2)} \mathrm{cov}(\delta_i, \delta^\top s); \tag{41}$$

in the competitive benchmark,

$$\mathrm{IL}_i^c = \frac{\gamma}{R_f L} \mathrm{cov}(\delta_i, \delta^\top s). \tag{42}$$

The first term in (39) (and the only term in (40)), $c_1 \mathrm{cov}(R_i, R_{q+x_0})$, is the standard capital asset pricing model (CAPM) term. Assets that covary positively with the $t = 1$ portfolio of liquidity providers will increase that portfolio's variance and must therefore compensate by offering higher returns. The second term, $l_1 \mathrm{cov}(R_i, R_q)$, is the illiquidity correction. A trade of portfolio q moves the price of asset i. In the competitive market, asset i's price is driven by the covariance $\text{cov}(\delta_i, \delta^\top(q + x_0))$. Strategic traders account for the impact of a marginal trade in the portfolio q on that covariance, a marginal effect that is given by

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

It follows that the illiquidity correction term is proportional to the covariance between R_i and R_q.[23] Because illiquidity is proportional to $\text{cov}(R_i, R_q)$, 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. Note also that—unlike the competitive benchmark, where the cross section of returns is priced with a single factor (R_{q+x_0})—assets in the noncompetitive case are priced with a combination of R_{q+x_0} and R_q; the result is a different cross section of asset returns.

We now turn to the general case. The following consumption-CAPM holds in our economy.

Proposition 7. *(Cross section of returns) Put $q = s/L$, and let $\zeta(t; q) = \frac{\exp(-\gamma(x_0+tq)^\top \delta)}{E[\exp(-\gamma(x_0+tq)^\top \delta)]}$. Then*

$$E[R_i] - R_f = -\text{cov}(Z^*(q), R_i), \quad \textit{where } Z^*(q) = (L-1)\int_1^\infty \xi^{-L}\zeta(\xi; q)\, d\xi. \tag{43}$$

In the competitive benchmark, we have

$$E[R_i] - R_f = -\text{cov}(Z^{*c}(q), R_i), \quad \textit{where } Z^{*c} = \zeta(1; q). \tag{44}$$

According to (43), the cross section of asset returns is governed by its covariance with the stochastic discount factor Z^*/R_f. That relation confirms the standard consumption-CAPM intuition that only the systematic risk (i.e., the part of return that covaries with the SDF) affects stock returns. We remark that, as in the single-asset case, $\zeta(t; q)/R_f$ is an SDF in a

[23] Rostek and Weretka (2015a) were the first to derive this correction term.

competitive economy where liquidity providers absorb an order of size $t \times q$. Equation (43) also shows that Z^*/R_f, the SDF in the economy with market power, is a weighted average of the SDFs in the competitive economies where liquidity providers absorb orders of a larger size. And just as in the single-asset case (see the discussion in Section 3.2), Z^*/R_f represents the demand reduction due to liquidity providers' market power. The reduction will differ for different assets, which means that the competitive benchmark and the noncompetitive case will feature different cross sections of returns.

Next we address the cross section of illiquidity.

Proposition 8. *(Cross section of illiquidity) The cross section of illiquidity is characterized by*

$$\text{IL}_i = \gamma \frac{L-1}{L(L-2)R_f} \text{cov}^*(\delta_i, \delta^\top s), \qquad \text{IL}_i^c = \frac{\gamma}{LR_f} \text{cov}^*(\delta_i, \delta^\top s); \tag{45}$$

here $\text{cov}^*(\delta_i, \delta^\top s)$ *is the risk-neutral covariance computed under the EMM in the economy with zero supply shock, with Radon–Nikodym derivative given by* $Z^*(0)$*.*

As in the single-asset case, the expressions for equilibrium illiquidity in general case are remarkably similar to those for a Gaussian economy. The intuition for why the covariance between δ_i and $q^\top \delta$ matters for illiquidity is similar to the Gaussian case. We can see intuitively that using risk-neutral (rather than physical) covariance helps account for the higher (physical) moments of δ. Note that illiquidity is not affected by higher moments under a risk-neutral measure. Requiring that the EMM used to compute the risk-neutral covariance correspond to the economy with zero supply shock simply means (cf. the single-asset case) that this covariance should be measured immediately prior to the supply shock.

Next we examine a tractabl limits of our general model when the supply shock s is small.

Proposition 9. *Suppose that liquidity demanders trade* t *units of the portfolio* s*. Then*

$$E^*[R_i] - R_f = t \cdot IL_i/P_i + o(t). \tag{46}$$

In the competitive benchmark we similarly have

$$E^*[R_i] - R_f = t \cdot IL_i^c/P_i + o(t),$$

where $E^*[R_i] = E^*[\delta_i]/P_i(s)$*. The risk-neutral expectation is computed under the EMM in an economy with zero supply shock, with Radon–Nikodym derivative given by* $Z^*(0)$*.*

Proposition 9 links the difference between spot-to-forward price ratio $E^*[\delta_i]/P_i(s)$ and the risk-free return R_f to asset illiquidity. In Section 5 we discuss how this proposition is related to empirical findings about asset carry (Koijen et al. 2018). Once again, the requirement that $E^*[\delta_i]$ be computed using the measure $Z^*(0)$ means that one should measure forward price $E^*[\delta_i]$ based on the options prices *before* realization of the supply shock s.

5 Implications

Sections 3.3 and 3.4 outlined our study's implications for equilibrium illiquidity and the price response function, respectively; in Section 4.3, we did likewise for the cross section of illiquidity and stock returns. In this section, these topics are discussed in more detail.

Equilibrium illiquidity

Proposition 2 demonstrates that an asset's equilibrium illiquidity is positively related to its risk-neutral variance. This positive link is strongly supported by the data: Nagel (2012) documents the strong positive relation between the VIX index (i.e., market risk-neutral variance) and the returns on liquidity provision reversal strategies. Moreover, he finds that VIX has explanatory power beyond that afforded by the market's physical variance. The proposition implies that risk-neutral variance might be a better explanatory variable for illiquidity because it helps account for the effect of higher moments.

In Proposition 2 we also show that the expression for equilibrium illiquidity is remarkably similar in the general case to its counterpart in the Gaussian case. One only has to substitute risk-neutral variance for the physical variance—a minor change that may nonetheless have profound effects on comparative statics (see Proposition 3). In particular, market illiquidity is widely presumed to be higher when liquidity providers are more risk averse or when the asset is riskier, given the information available (see e.g. the review paper by Vayanos and Wang 2012). This common wisdom is based on the CARA-normal models typically employed to analyze price impact. Our model with Gaussian distribution produces the same result (Corollary 2), but the common wisdom is not robust to the case of a general distribution. Proposition 3 shows that, whereas the conventional comparative statics hold when risk aversion γ is low, if γ is high enough then they must necessarily reverse. The implication is that, when risk aversion is high (i.e., in bad times), policy measures aimed at improving liquidity (say, by requiring more disclosure) will have exactly the opposite effect. Accounting for this dynamic is critical because policy interventions are more likely to occur during bad times.

Price response function

We start by summarizing known stylized facts about the shape of the price response function.

1. The price response function is an increasing and concave function of the order size.[24]

2. The price response function is asymmetric: the price effects of sell and buy orders are not the same.[25]

Proposition 4 implies that, for a large order size or for a sufficiently high level of liquidity providers' risk aversion, the price response function is concave; that explains the first stylized fact. The monotonicity of the two measures follows from Proposition 1. Proposition 4 also links the asymmetry of the price response function to the skewness of returns. Our model predicts

[24]For equities, see Holthausen, Leftwich, and Mayers (1990), Keim and Madhavan (1996), and Almgren et al. (2005); for options, see Muravyev (2016). In the case of equities, a specification popular among practitioners is a linear permanent price impact and power-law temporary price impact, where the exponent is estimated to be 1/2 (Plerou, Gopikrishnan, Gabaix, and Stanley 2002) or 3/5 (Almgren et al. 2005). Such a specification is employed in Citigroup's BECS software Almgren et al. (2005), for instance. A price response function is the sum of temporary and permanent price impacts.

[25]Saar (2001) summarizes the evidence for buy orders having a greater price impact than sell orders.

that, with positive skewness (a natural property for stocks at the individual level; Chen, Hong, and Stein 2001), the price impact of purchases is greater than that of sales—which agrees with the evidence summarized by Saar (2001). Stock returns are positively skewed at the individual level, but their skewness is negative at the aggregate level (Bakshi et al. 2003; Albuquerque 2012). The model implies that, for indices, purchases move prices less than sales do (consistently with the evidence presented by Chordia et al. 2002).[26]

Chiyachantana et al. (2004) find that the asymmetry of the permanent price impact is linked to the underlying market condition. Our model links that asymmetry to skewness, and Perez-Quiros and Timmermann (2001) present evidence that skewness varies with the underlying market condition. Thus the model's predictions are in line with previous findings.

Cross section of illiquidity and stock returns

Propositions 8 and 9 suggest that option data can be informative about equities' illiquidity. Thus Proposition 8 stresses the role of the risk-neutral covariance between asset returns and the returns on a portfolio traded by liquidity demanders. Buss and Vilkov (2012) document that option-implied covariances have power in addition to that of physical covariances in explaining the cross section of equity returns. Our theory implies that option-implied covariances help account for asset illiquidity.

Proposition 9 could shed some light on empirical findings about asset carry (Koijen et al. (2018)). To see how, consider Equation (46); its left-hand side is related to Koijen et al.'s *carry return* (i.e., the difference between the spot and forward prices normalized by the forward

[26]Our model links the asymmetry to *risk-neutral* skewness rather than to physical skewness. The two measures should be close to each other when the risk aversion of liquidity providers is low, or in good times, but in bad times the two measures might differ significantly. It is interesting that, according to Chiyachantana, Jain, Jiang, and Wood (2004), the results for price response asymmetry in bad times are opposite to those cited here.

price). Indeed, rewriting their equation (8) in our paper's notation yields

$$\text{carry} = \left(\underbrace{\frac{E^*[\delta_i]}{P_i}}_{E^*[R_i]} - R_f + 1 \right) \frac{P_i}{F_i}. \tag{47}$$

Koijen et al. note also that the last term—the ratio $\frac{P_i}{F_i}$ of spot to forward prices—is close to 1 in the data. Combining (46) and (47), we obtain

$$\text{carry} \approx t \cdot IL_i / P_i + 1.$$

That is, there is a positive association between carry and asset illiquidity. In line with this result, Koijen et al. (2018) find that carry is a strong positive predictor of returns and also 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 effects arise from traders' limited risk-bearing capacity. We model trade using the classic uniform-price double-auction protocol in which traders submit price-contingent demand schedules. For the single-asset case, see Klemperer and Meyer (1989), Kyle (1989), Vayanos (1999), Wang and Zender (2002), Vives (2011), Rostek and Weretka (2012), Ausubel et al. (2014), Bergemann, Heumann, and Morris (2015), Rostek and Weretka (2015b), Du and Zhu (2017), Kyle, Obizhaeva, and Wang (2017), and Lee and Kyle (2018); for the multi-asset case, see Rostek and Weretka (2015a) and Malamud and Rostek (2017).[27] Antill and Duffie (2017) and Duffie and Zhu (2017) consider models in which the uniform-price auction market is augmented by price discovery sessions.

[27] Sannikov and Skrzypacz (2016) develop an alternative trading protocol, a "conditional double auction", in which traders can condition their demand schedules on the trading rates of other players.

All of these papers feature traders with marginal utilities that are linear in trade size (which is either assumed directly or follows from the combination of CARA utility and normality of asset payoffs).[28] With the exception of Du and Zhu (2017), they derive linear equilibria in which (a) slopes of the demand schedules are independent of the price level and (b) the equilibrium price impact (which is given by the inverse of the residual supply's slope) is constant irrespective of the trade size. As noted before, such linear equilibria are poorly aligned with empirical evidence for the nonlinearity and asymmetry of the price response to order imbalances. Du and Zhu (2017) derive nonlinear equilibria when there are two agents, in which case no linear equilibria exist. Du and Zhu also show that there often exist nonlinear equilibria. This nonlinearity is not linked to higher moments, which is a fundamental aspect of our paper; instead, in Du and Zhu it is linked to strategic behavior by traders.[29] As far as we know, 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 with higher moments of asset payoffs.[30]

There is a large body of literature on *competitive* trading with nonstrategic liquidity providers in set-ups that deviate from CARA normal. 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 (1990a,b), Bhattacharya and Spiegel (1991), DeMarzo and Skiadas (1998, 1999), Yuan (2005), Albagli, Hellwig, and Tsyvinski

[28]Bagnoli, Viswanathan, and Holden (2001) derive necessary and sufficient conditions for linear equilibria in Kyle-type models. They use a characteristic function approach to show that linear equilibria are possible even when the distributions are not Gaussian. In contrast, we focus on nonlinear equilibria and— in our model—linearity is possible only in the Gaussian case; also, we adopt a cumulant generating function approach.

[29]Other papers that analyze nonlinear equilibria in settings with linear marginal utility include Bhattacharya and Spiegel (1991), Wang and Zender (2002), and Boulatov and Bernhardt (2015). In these works, some of the equilibria (among many) are nonlinear. As in Du and Zhu (2017), the nonlinearity is not linked to higher moments but rather to traders' strategic behavior. Moreover, in all these papers only linear equilibria remain after the selection criterion is applied.

[30]Another class of strategic trading models assumes that strategic traders use market orders to trade; see Kyle (1985), Subrahmanyam (1991), Rochet and Vila (1994), Foster and Viswanathan (1996), and Vayanos (2001), among others. Rochet and Vila go beyond the CARA-normal framework; they analyze a model à la Kyle (1985) without normality and prove the uniqueness of an equilibrium. However, Rochet and Vila (1994) derive no implications regarding the cross section of illiquidity and asset returns, price response asymmetry, or the comparative statics of illiquidity.

(2015), Breon-Drish (2015), Pálvölgyi and Venter (2015), and Chabakauri, Yuan, and Zachariadis (2017). Peress (2003) and Malamud (2015) examine noisy rational expectations equilibria 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 this assumption has notable implications for the cross section of stock returns.

Several studies seek to explain the shape of the price impact. Roşu (2009) presents a model of the limit order book in which the main friction is the costs associated with waiting for the limit orders to be executed. Keim and Madhavan (1996) explain concave price effects in terms of a search friction in the "upstairs" market for block transactions. Saar (2001) gives an institutional account of the price impact asymmetry across buys and sells. We add to this literature by providing a unified treatment of the properties of the price response function and then linking them to the shape of the probability distribution that describes asset payoffs.

Our paper is related also 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). Our study differs from these in that we assume transaction costs to be endogenous. In addition, we demonstrate that commonality in transaction costs (illiquidity) emerges endogenously in our model. This paper speaks to the literature on optimal dynamic execution algorithms for price effects that are exogenous and nonconstant (see Bertsimas and Lo 1998; Almgren and Chriss 2001; Almgren et al. 2005; Huberman and Stanzl 2005; Obizhaeva and Wang 2013). Our paper complements this literature by providing equilibrium foundations for nonlinear price functions.

Finally, there is a related strand of the literature that considers strategic liquidity provision and uses discriminatory price mechanisms to model trade. Notable examples are the studies of 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 both Biais et al. and Back and Baruch assume that liquidity providers are risk neutral. Hence there is no inventory risk, on which our model focuses.

7 Conclusion

We present a tractable model of strategic trading in an economy populated by a finite number of large and strategic CARA investors who trade a finite number of assets with arbitrary distribution of asset payoffs. We show that departing from the common (but unrealistic) assumption of normal payoffs has far-reaching economic implications for illiquidity and asset returns. More specifically: (i) the equilibrium price is nonlinear, and the price response is an asymmetric function of order size; (ii) liquidity is a non-monotonic function of risk aversion, and of uncertainty about future asset payoffs; and (iii) there is (endogenous) commonality in illiquidity that is summarized by the second moment of assets' joint risk-neutral distribution. These results are consistent with extant empirical evidence on liquidity.

We develop a novel constructive approach to solve for the equilibrium in closed form. We establish that solving for the equilibrium is equivalent to solving a linear ODE, which can be done using standard methods. It would be instructive to extend, along several relevant dimensions, our departure from the common CARA-normal assumption in strategic trading models. We are currently examining the equilibrium implications of wealth effects (i.e., removing the CARA assumption) and heterogeneity for investors' wealth. Other extensions worth exploring include the cases of heterogeneity in investors' risk aversion (as a means 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
$\nabla f(q),\ \text{where } f:\ \mathbb{R}^N \to \mathbb{R}$	Gradient of f, $(\nabla f)_l = \frac{\partial f}{\partial q_l}$
$\nabla^2 f(q),\ \text{where } f:\ \mathbb{R}^N \to \mathbb{R}$	Hessian of f, $(\nabla^2 f)_{kl} = \frac{\partial^2 f}{\partial q_k \partial q_l}$
$\nabla I(q),\ \text{where } I:\ \mathbb{R}^N \to \mathbb{R}^N$	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., $I_k^i(q)$ 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 6.

B.3 Proof of Corrolary 1

Proof of Corrolary 1. The function g satisfies

$$\frac{\partial g}{\partial y} = \frac{E[\delta \exp(y\delta)]}{E[\exp(y\delta)]}\,.$$

It follows from Equations (13) and (15) and integration by parts that

$$
\begin{aligned}
\Lambda(q) &= \frac{\gamma}{R_f}\int_1^\infty \xi^{1-L} g''(-\gamma(\xi q + x_0))d\xi \\
&= \frac{1}{qR_f}\left[g'(-\gamma(q+x_0)) \;-\; (L-1)\int_1^\infty \xi^{-L} g'(-\gamma(\xi q + x_0))d\xi\right] \\
&= \frac{1}{qR_f}\left[E[\delta Z^{*c}(q)] \;-\; R_f I(q)\right] .
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
I(q) &= \frac{L-1}{R_f}\int_1^\infty \xi^{-L} g'(-\gamma(\xi q + x_0))d\xi \\
&= \frac{L-1}{R_f}\int_1^\infty \xi^{-L}\frac{E[\delta\exp(-\gamma(\xi q + x_0)\delta)]}{E[\exp(-\gamma(\xi q + x_0)\delta)]}d\xi \\
&= \frac{L-1}{R_f}E\left[\int_1^\infty \xi^{-L}\frac{\delta\exp(-\gamma(\xi q + x_0)\delta)}{E[\exp(-\gamma(\xi q + x_0)\delta)]}d\xi\right] \\
&= \frac{1}{R_f}E\left[\delta\,(L-1)\int_1^\infty \xi^{-L}\frac{\exp(-\gamma(\xi q + x_0)\delta)}{E[\exp(-\gamma(\xi q + x_0)\delta)]}d\xi\right] \\
&= \frac{E[Z^*(q)\delta]}{R_f} ,
\end{aligned}
$$

where the change of expectation and integration follows from Fubuni's theorem and Lemma 6. The result then follows. ■

B.4 Proof of Corollary 2

The proof is a special case of the proof of Corollary 4

B.5 Proof of Proposition 2

Proof of Proposition 2. We know that

$$
\mathrm{IL} = \frac{L-1}{L}\Lambda(0) \;=\; \frac{L-1}{L}\frac{\gamma}{R_f}\frac{1}{L-2}g''(-\gamma x_0) .
$$

The functions g and f satisfy

$$\frac{\partial^2 f}{\partial y^2} = -\frac{\gamma}{R_f}\frac{\partial^2 g}{\partial y^2}$$
$$\frac{\partial^2 g}{\partial y^2} = E\left[\delta^2 \frac{e^{y\delta}}{E[e^{y\delta}]}\right] - \left(E\left[\delta\frac{e^{y\delta}}{E[e^{y\delta}]}\right]\right)^2 \implies \left.\frac{\partial^2 g}{\partial y^2}\right|_{y=-\gamma x_0} = E\left[\delta^2\zeta(0;q)\right] - \left(E\left[\delta\zeta(0;q)\right]\right)^2 .$$

The result then follows since $Z^*(0) = \zeta(0;q)$. ■

B.6 Proof of Proposition 3

Proof of Proposition 3.

Proposition 2 implies that

$$\frac{L(L-2)}{L-1}\mathrm{IL}(\gamma,\mathcal{F}_i) = L\,\mathrm{IL}^c(\gamma,\mathcal{F}_i) = \gamma\mathrm{var}^*(\delta|\mathcal{F}_i) = h(\gamma,\mathcal{F}_i)\,,$$

where we define (within this proof)

$$h(\gamma) \equiv \gamma g''(-\gamma x_0,\mathcal{F}_i) = \gamma\left(E\left[\delta^2\frac{e^{-\gamma x_0\delta}}{E[e^{-\gamma x_0\delta}|\mathcal{F}_i]}\middle|\mathcal{F}_i\right] - \left(E\left[\delta\frac{e^{-\gamma x_0\delta}}{E[e^{-\gamma x_0\delta}]}\middle|\mathcal{F}_i\right]\right)^2\right) .$$

and we denoted conditional CGF $g(x,\mathcal{F}_i) = \log E[\exp(x^\top\delta)|\mathcal{F}_i]$.

We have $h'(0) = g''(0) = \mathrm{var}[\delta|\mathcal{F}_i]$. It thus follows that for γ sufficiently small,

$$\frac{\partial \mathrm{IL}(\gamma,\mathcal{F}_1)}{\partial\gamma} > 0, \qquad \frac{\partial \mathrm{IL}^c(\gamma,\mathcal{F}_1)}{\partial\gamma} > 0\,,$$

and by the law of total variance that (23) also holds.

To see that these inequalities have to change note that the illiquidity can be written as

$$\mathrm{IL}(\gamma,\mathcal{F}_1) = \frac{\gamma g''\left(-\gamma x_0,\mathcal{F}_1\right)}{R_f(L-2)}.$$

It then follows that

$$\int_0^\infty g''(-\gamma x_0, \mathcal{F}_i) d\gamma = R_f(L-2) \int_0^\infty \frac{\text{IL}(\gamma, \mathcal{F}_1)}{\gamma} d\gamma. \tag{48}$$

LHS of (48):

$$\int_0^\infty g''(-\gamma x_0) d\gamma = \frac{1}{-x_0} g'(t)|_{t=0}^{-\infty} = \frac{a - E[\delta|\mathcal{F}_1]}{x_0}.$$

where we used lemma 3 to conclude that $g'(-\infty) = \text{ess inf}\delta$, and we denoted ess $\text{inf}\delta = a$. Note that

RHS of (48):

$$R_f(L-2) \int_0^\infty \frac{\text{IL}(\gamma, \mathcal{F}_1)}{\gamma} d\gamma.$$

If $\text{IL}(\gamma)$ increases in γ the RHS of (48)diverges, whereas the LHS is bounded. A contradiction. A proof for the competitive case is similar and is omitted for brevity.

■

B.7 Proof of Corollary 3

Proof of Corollary 3. Recall that $P(s) = I(s/L)$. It follows from the proof of Corollary 2 (see the proof of Corollary 2) that

$$P(s) \;=\; \frac{1}{R_f}\left[\mu - \gamma(x_0)\sigma^2\right] - \frac{L-1}{L(L-2)}\frac{\gamma\sigma^2}{R_f}s \implies P(s) - P(0) \;=\; -\frac{L-1}{L(L-2)}\frac{\gamma\sigma^2}{R_f}s\,.$$

Therefore,

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

For the competitive case, we know that $P^c(s) = f'(s/L)$. Thus,

$$\pi^c(s) \;=\; \frac{1}{LR_f}\gamma\sigma^2|s|\,.$$

■

B.8 Proof of Proposition 4

Proof of Proposition 4.

The proof below is for the case γ small enough. The proof for small enough order size is analogous and is omitted for brevity.

First, as $\gamma \to 0$, we have

$$\lim_{\gamma \to 0} \zeta(t;q) \;=\; 1,\; a.s.$$

Thus, it follows from the DCT that

$$\lim_{\gamma \to 0} \text{skew}^*(\delta) \;=\; \text{skew}(\delta)\,,$$

where

$$\text{skew}^*(\delta) \;=\; E[\delta^3\zeta(1;q)] - 3E[\delta^3\zeta(1;q)]Var[\delta^3\zeta(1;q)] - E^3[\delta\zeta(1;q)]\,.$$

Therefore, $\text{sign}(\text{skew}^*(\delta)) \;=\; \text{sign}(\text{skew}(\delta))$ for γ sufficiently small.

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. $\text{sign}(\pi(s)-\pi(-s)) = \text{sign}(\pi^c(s)-\pi^c(-s)) = -\text{sign}(\text{skew}(\delta))$, *and* $|\pi(s)-\pi(-s)| > |\pi^c(s)-\pi^c(-s)|$.

Recall that $p(s;\gamma) = I(s/L;\gamma)$. Differenting (13) with respect to γ we get

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

and

$$\frac{\partial^2}{\partial\gamma^2} P(s;\gamma)|_{\gamma=0} = \frac{g'''(0)}{R_f}\left(x_0^2 + \left(\frac{s}{L}\right)^2 \frac{(L-1)}{(L-3)} + 2x_0 \frac{s}{L}\frac{(L-1)}{(L-2)}\right)$$
$$= \frac{\text{skew}(\delta)}{R_f}\left(x_0^2 + \left(\frac{s}{L}\right)^2 \frac{(L-1)}{(L-3)} + 2x_0 \frac{s}{L}\frac{(L-1)}{(L-2)}\right).$$

Thus, it follows from Taylor's Theorem that, to second order,

$$P(s;\gamma) - P(s;0) \;=\; -\gamma\frac{\text{var}(\delta)}{R_f}\left(x_0 + \frac{s}{L}\frac{(L-1)}{(L-2)}\right) + \frac{1}{2}\gamma^2\frac{\text{skew}(\delta)}{R_f}\left(x_0^2 + \left(\frac{s}{L}\right)^2\frac{(L-1)}{(L-3)} + 2x_0\frac{s}{L}\frac{(L-1)}{(L-2)}\right).$$

Similarly,

$$P(0;\gamma) - P(0;0) \;=\; -\gamma\frac{\text{var}(\delta)}{R_f}x_0 \;+\; \frac{1}{2}\gamma^2\frac{\text{skew}(\delta)}{R_f}x_0^2.$$

Therefore, we have

$$\pi(s;\gamma) = \; P(0;\gamma) - P(s;\gamma) \;=\; P(0;\gamma) - P(0;0) - (P(s;\gamma) - P(s;0)) + (P(0;0) - P(s;0))$$
$$= \; P(0;\gamma) - P(0;0) - (P(s;\gamma) - P(s;0))$$
$$= \gamma\left(\frac{L-1}{L-2}s/L\right)\frac{\text{var}(\delta)}{R_f} - \frac{1}{2}\gamma^2\left(\left(\frac{s}{L}\right)^2\frac{(L-1)}{(L-3)} + 2x_0\frac{s}{L}\frac{(L-1)}{(L-2)}\right)\frac{\text{skew}(\delta)}{R_f} + o\left(\gamma^2\right).$$

Proceeding similarly for $P(0;\gamma) = g'(-\gamma x_0)/R_f$, we get

$$\pi(s;\gamma) = P(0;\gamma) - P(s;\gamma)$$
$$= \gamma\left(\frac{L-1}{L-2}s/L\right)\frac{\text{var}(\delta)}{R_f} - \frac{1}{2}\gamma^2\left(\left(\frac{s}{L}\right)^2\frac{(L-1)}{(L-3)} + 2x_0\frac{s}{L}\frac{(L-1)}{(L-2)}\right)\frac{\text{skew}(\delta)}{R_f} + o\left(\gamma^2\right).$$

Similarly,

$$\pi(-s;\gamma) = P(-s;\gamma) - P(0;\gamma)$$
$$= \gamma\left(\frac{L-1}{L-2}s/L\right)\frac{\text{var}(\delta)}{R_f} - \frac{1}{2}\gamma^2\left(-\left(\frac{s}{L}\right)^2\frac{(L-1)}{(L-3)} + 2x_0\frac{s}{L}\frac{(L-1)}{(L-2)}\right)\frac{\text{skew}(\delta)}{R_f} + o\left(\gamma^2\right).$$

Correspondingly,

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

Proceeding similarly for $P^c(s;\gamma) = g'(-\gamma s/L)/R_f$ we get

$$\pi^c(s;\gamma) - \pi^c(-s;\gamma) = -\left(\frac{s}{L}\right)^2 \frac{\gamma^2}{R_f}\text{skew}(\delta) + o\left(\gamma^2\right).$$

The statement follows directly from the last two displayed equations.

Part 2. $\text{sign}(\pi''(s)) = \text{sign}((\pi^c)''(s)) = -\text{sign}(\text{skew}(\delta)s)$, *and* $|\pi''(s)| > |(\pi^c)''(s))|$.

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

$$\pi''(s;\gamma) = -\gamma^2 \frac{L-1}{R_f L^2} \int_1^\infty \xi^{2-L} g'''(-\gamma(\xi s/L + x_0))d\xi. \tag{49}$$

One can see that

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

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

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

Similarly, for $(\pi^c)''(s;\gamma)$:

$$\lim_{\gamma \to 0} \frac{\pi''(s;\gamma)}{\gamma^2} = -\frac{1}{R_f L^2}\text{skew}(\delta), \text{ for } s > 0, \text{ and}$$

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

The claim follows from the last four displayed equations.

Parts 3 and 4. $(\pi^c)''(s) < \pi''(s) < 0$ *for large enough* γ *(for given* s*:* $|s| > L|x_0|$*) or for large enough* $|s|$ *(for given* γ*), provided that* $g'''(x)$ *and* $g^{(4)}(x)$ *does not change sign for* $|x|$ *large* *enough.*

Consider the case $s > 0$. The case $s < 0$ can be considered analogously. 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 (49) we see that $\pi''(s,\gamma)$ is negative both when s is large enough or γ is large enough.

Differentiating $\pi^c(s,\gamma) = \frac{g'(-\gamma s/L)}{R_f}$ twice and using (49) one can derive

$$\pi''(s;\gamma) - (\pi^c)''(s;\gamma) = -\gamma^2 \frac{L-3}{R_f L^2} \int_1^\infty \xi^{2-L} \left(\frac{L-1}{L-3} g'''(-\gamma(\xi s/L + x_0)) - g'''(s/L + x_0) \right) d\xi.$$

Provided that $g^{(4)}(-x)$ is positive the bracketed term inside the integral is positive and so is $\pi''(s;\gamma) - (\pi^c)''(s)$. This derivative cannot be negative because $g''(x)$ is positive (hence bounded from below) and decreasing. Since it does not change sign, it must be positive, from which the statement follows.

■

B.9 Proof of Corollary 3

Proof of Corollary 3. Recall that $P(s) = I(s/L)$. It follows from the proof of Corollary 2 (see the proof of Corollary 2) that

$$P(s) = \frac{1}{R_f}\left[\mu - \gamma(x_0)\sigma^2\right] - \frac{L-1}{L(L-2)}\frac{\gamma\sigma^2}{R_f}s \implies P(s) - P(0) = -\frac{L-1}{L(L-2)}\frac{\gamma\sigma^2}{R_f}s\,.$$

Therefore,

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

For the competitive case, we know that $P^c(s) = f'(s/L)$. Thus,

$$\pi^c(s) = \frac{1}{LR_f}\gamma\sigma^2|s|\,.$$

■

B.10 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

$$\nabla f\left(q^i\right) + \frac{1}{L-1}\nabla I\left(\frac{s-q^i}{L-1}\right)q^i = I\left(\frac{s-q^i}{L-1}\right). \tag{50}$$

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:

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

The equilibrium inverse demand $I(q)$ must be a strictly decreasing solution to (51) 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 (51), then $q = s/L$ is the unique solution to FOCs (50). Moreover, $q = s/L$ is a local maximum.*

Proof. Denote

$$\xi = \frac{s-q^i}{L-1} \tag{52}$$

and rewrite (50) as follows:

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

Instead of solving for $q^i(s)$ from (50), we will solve an equivalent system of equations (53) and (52).

Step 1. There is at most one solution to (53).

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

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

$$\nabla f\left(q_2\right)+\frac{1}{L-1}\nabla I\left(\xi\right)q_2=I\left(\xi\right). \tag{55}$$

Multiply (54) and (55) by $(q_2-q_1)^\top$ and subtract one equation from the other, as follows:

$$(q_2-q_1)^\top(\nabla f\left(q_2\right)-\nabla f(q_1))+\frac{1}{L-1}(q_2-q_1)^\top\nabla I\left(\xi\right)\left(q_2-q_1\right)=0. \tag{56}$$

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 (56) is negative; the right-hand side is zero.

Step 2. The only solution to (53) is $q^i=\xi$.

Indeed, $q^i=\xi$ is a solution, since for such q^i, equation (53) becomes equation (51). By the previous step, there is at most one solution. Hence, $q^i=\xi$ is the only solution to (53).

Step 3. The only solution to (50) is $q^i=s/L$.

Indeed, (50) is equivalent to a system of equations (53) and (52). We know that the only solution to (53) is $q = \xi$. Therefore, the system of equations (53) and (52) becomes

$$q^i = \xi, \tag{57}$$

$$\xi = \frac{s - q^i}{L - 1}, \tag{58}$$

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 (50) and substituting $q^i = q^* \equiv s/L$, we get

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

where the partial derivatives in ∇ are taken with respect to the components of q^*. Differentiating (51), we get

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

Combining the two preceding equations we get

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

■

Lemma 2. *Given that $I(q)$ solves (51) 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

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

The result then follows.

■

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

Lemma 5. *The unique solution to (51) such that $I(q) \in \mathcal{A}$ is*

$$I(q) = (L-1)\int_1^\infty t^{-L}\nabla f(tq)dt. \tag{59}$$

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

$$\text{ess inf}(q^\top\delta) < q^\top I(q). \tag{60}$$

Writing (60) for a portfolio tq as well as $-tq$, we also get

$$\text{ess inf}(q^\top \delta) < \iota(t;q) < \text{ess sup}(q^\top \delta), \tag{61}$$

which must hold for any t. According to Proposition 5, finding a solution to (51) amounts to solving linear ODE (34). This solution implies that $I(q) \in \mathcal{A}$ iff $\iota(t;q)$ is such that for any t, and any q (61) holds.

Step 1. Solving ODE (34).

We multiply both sides of (34) 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 (60) imply that $\iota(t;q)$ is bounded, we get a particular solution to (34)

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

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 (34) is given by

$$\iota(x;q) = (L-1)\int_1^\infty \xi^{-L} q^\top \nabla f(x\xi q)d\xi + ct^{L-1}, \tag{62}$$

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

Step 2. The solution (62) 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 (62) *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, (61) cannot hold.

Step 4. The solution (62) *with* $c = 0$ *implies* $I(q) = (L-1)\int_1^\infty \xi^{-L}\nabla f(\xi q)d\xi$.

Indeed, the solution (62) 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 (33) to get (59) ■

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.

$$I\left(\frac{s-t_k\theta_k}{L-1}\right)^\top\theta_k \tag{63}$$

$$= \; -L\int_1^\infty z^{-L-1}\nabla f\left(-zc\frac{s-t_k\theta_k}{L-1}\right)^\top\theta_k dz. \tag{64}$$

$$\tag{65}$$

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

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

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,

$$E[e^{t_k\theta_k^\top\delta}] \;=\; E[e^{t_k\theta_k^\top\delta}\mathbf{1}_{A_k}] \;+\; E[e^{t_k\theta_k^\top\delta}(1-\mathbf{1}_{A_k})] \;\leq\; E[e^{t_k\theta_k^\top\delta}\mathbf{1}_{A_k}] + e^{t_k(b_k-\epsilon)}\,. \tag{66}$$

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.11 Proof of Proposition 5

Proof of Proposition 5.

Step 1. PDE (28) implies ODE (34).

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 (34) 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 (28) 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 (33).

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 (28) and noting that $\frac{1}{L-1}(\nabla I(q)\, q + I(q)) = \nabla e(q)$ we get (33). ■

B.12 Proof of Proposition 6

Proof of Proposition 6. Equilibrium inverse demand is a solution to PDE (28), which is strictly decreasing and such that $I(q) \in \mathcal{A}$. Lemma 5 implies that there is unique such solution, given by (35) or, equivalently, (36). Expressions (37) and (38) are obtained by differentiating (35) and (36). ■

B.13 Proof of Corollary 4

Proof of Corollary 4. Suppose that $\delta \sim N(\mu, \Sigma)$. Then

$$g(y) \;=\; y^\top \mu + \frac{1}{2} y^\top \Sigma y; \quad f(q) \;=\; -\frac{1}{\gamma R_f}\left[-\gamma(x_0+q)^\top \mu + \frac{1}{2}\gamma^2 (x_0+q)^\top \Sigma (x_0+q)\right].$$

It follows that

$$\nabla f(q) \;=\; -\frac{1}{R_f}\left[-\mu + \gamma \Sigma (x_0+q)\right] \quad \text{and} \quad \nabla^2 f(q) \;=\; -\frac{1}{R_f}\gamma\Sigma\,.$$

The first-order condition is

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

The solution to this first-order ODE with variable coefficient is

$$I(q) \;=\; \frac{1}{R_f}\left[\mu - \gamma\Sigma x_0\right] - \frac{L-1}{L-2}\frac{\gamma}{R_f}\Sigma q \;\Longrightarrow\; \nabla I(q) \;=\; -\frac{L-1}{L-2}\frac{\gamma}{R_f}\Sigma\,.$$

Write the above displayed equation for security i (recall that $P(s) = I(q = s/L)$):

$$R_f P_i = \mu_i - \gamma \mathrm{cov}\left(\delta_i, \delta^\top \left(x_0 + q\right)\right) - \frac{\gamma}{(L-2)}\mathrm{cov}\left(\delta_i, \delta^\top q\right)$$

Divide by P_i and rearrange to obtain

$$E[R_i] - R_f = \gamma p_{q+x_0}\mathrm{cov}\left(R_i, R_{x_0+q}\right) + \frac{\gamma p_q}{L-2}\mathrm{cov}\left(R_i, R_q\right).$$

The statement for the CAPM in the non-competitive case then follows. The proof for competitive case is analogous and is omitted for brevity

For illiquidity, observe that

$$\mathrm{IL}_i \;=\; \frac{L-1}{L}e_i^\top \Lambda(0)s \qquad \text{and} \qquad \mathrm{IL}_i^c \;=\; -\frac{1}{L}e_i^\top \nabla^2 f(0)s$$

The result then follows.

■

B.14 Proof of Proposition 7

Proof of Proposition 7.

We follow a common approach to proving CAPM. For any random variable M with

$E[M] = 1$, we have

$$
\begin{aligned}
\operatorname{cov}(M, \delta_l) + E[\delta_l] &= E[M\delta_l] \\
\Longrightarrow E[R_l] - \frac{E[MR_l]}{P_l} &= -\operatorname{cov}(M, R_l)
\end{aligned}
$$

We can adapt the proof of Corollary 1 to show that the price of l-th asset can be written as

$$
P_l(s) \equiv I_l(s) = \frac{E[Z^*(s)\delta_l]}{R_f}.
$$

Similarly, the corresponding competitive price is (see Equation (3))

$$
P_l^c(q) \equiv \frac{\partial f(s)}{\partial q_l} = \frac{E[Z^{*c}(s)\delta_l]}{R_f}.
$$

The result then follows.

■

B.15 Proof of Proposition 8

Proof of Proposition 8. In the multi-assets case, the functions g and f satisfy (see the proof of Corollary 1 for technical details)

$$
\begin{aligned}
\nabla^2 f &= -\frac{\gamma}{R_f}\nabla^2 g \\
\nabla^2 g(y) &= E\left[\delta\delta^\top \frac{e^{y^\top\delta}}{E[e^{y^\top\delta}]}\right] - E\left[\delta \frac{e^{y^\top\delta}}{E[e^{y^\top\delta}]}\right] E\left[\delta^\top \frac{e^{y^\top\delta}}{E[e^{y^\top\delta}]}\right] \\
\Longrightarrow \quad \nabla^2 g\big|_{y=-\gamma x_0} &= \operatorname{cov}^*(\delta, \delta) \\
\Longrightarrow \Lambda(0) &= \frac{\gamma}{R_f}\frac{1}{L-2}\operatorname{cov}^*(\delta, \delta).
\end{aligned}
$$

We know that

$$
\mathrm{IL}_i = \frac{L-1}{L} e_i^\top \Lambda(0) s \qquad \text{and} \qquad \mathrm{IL}_i^c = -\frac{1}{L} e_i^\top \nabla^2 f(0) s
$$

The result then follows. ■

B.16 Proof of Proposition 9

Proof of Proposition 9. Denote

$$\zeta(\xi;\epsilon) = \frac{\exp\left(-\gamma\left(x_0+\xi\epsilon q\right)^\top\delta\right)}{E\left[\exp\left(-\gamma\left(x_0+\xi\epsilon q\right)^\top\delta\right)\right]}$$

Note that

$$\frac{\partial\zeta(\xi;\epsilon)}{\partial\epsilon} = -\gamma\xi\left(\zeta(\xi;\epsilon)q^\top\delta - \zeta(\xi;\epsilon)E[\zeta(\xi;\epsilon)q^\top\delta]\right).$$

Note also that

$$\zeta(\xi,0) = \frac{\exp\left(-\gamma x_0^\top\delta\right)}{E\left[\exp\left(-\gamma x_0^\top\delta\right)\right]} \equiv \zeta_0$$

does not depend on ξ.

Prie of l-th asset is given by

$$\begin{aligned}
R_f P_l &= E\left[\left((L-1)\int_1^\infty \xi^{-L}\zeta(\xi;\epsilon)d\xi\right)\delta_l\right] \\
&= E\left[\left((L-1)\int_1^\infty \xi^{-L}\left(\zeta(\xi;0)+\epsilon\frac{\partial\zeta(\xi;0)}{\partial\epsilon}\right)d\xi\right)\delta_l\right] + o(\epsilon) \\
&= E\left[\zeta_0\delta_l\right] - \gamma\epsilon\left((L-1)\int_1^\infty \xi^{1-L}\left(E\left[\zeta_0\delta_l q^\top\delta\right] - E\left[\zeta_0\delta_l\right]E[\zeta_0 q^\top\delta]\right)d\xi\right) + o(\epsilon) \\
&= E^*\left[\delta_l\right] \underbrace{- \frac{\gamma\epsilon(L-1)}{L-2}\text{cov}^*(\delta_l, q^\top\delta)}_{\text{illiquidity term}} + o(\epsilon)
\end{aligned}$$

Similarly, in the competitive case

$$\begin{aligned} R_f P_l &= E\left[\zeta(1;\epsilon)\delta_l\right] \\ &= E\left[\left(\zeta(1;0) + \epsilon\frac{\partial\zeta(1;0)}{\partial\epsilon}\right)\delta_l\right] + o(\epsilon) \\ &= E^*\left[\delta_l\right] - \underbrace{\gamma\epsilon\mathrm{cov}^*(\delta_l, q^\top\delta)}_{\text{illiquidity term}} + o(\epsilon). \end{aligned}$$

■

C CARA-Normal Benchmark as a Limit

We analyze the 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 (12) and (14) converge to their corresponding counterparts in the Gaussian benchmark as the truncation bounds go to infinity. We start by deriving the Gaussian benchmark.

C.1 CARA-Normal Benchmark

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

$$g(y) \;=\; y^\top\mu + \frac{1}{2}y^\top\Sigma y; \quad f(q) \;=\; -\frac{1}{\gamma R_f}\left[-\gamma(x_0+q)^\top\mu + \frac{1}{2}\gamma^2(x_0+q)^\top\Sigma(x_0+q)\right].$$

It follows that

$$\nabla f(q) \;=\; -\frac{1}{R_f}\left[-\mu + \gamma\Sigma(x_0+q)\right] \quad \text{and} \quad \nabla^2 f(q) \;=\; -\frac{1}{R_f}\gamma\Sigma\,.$$

The first-order condition is

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

The *unique* solution to this first-order ODE with variable coefficient is

$$I(q) = \frac{1}{R_f}\left[\mu - \gamma\Sigma x_0\right] - \frac{L-1}{L-2}\frac{\gamma}{R_f}\Sigma q \tag{67}$$

$$\Longrightarrow \Lambda(q) \equiv -\frac{1}{L-1}I'(q) = \frac{1}{L-2}\frac{\gamma}{R_f}\Sigma\,. \tag{68}$$

C.2 CARA-Normal Benchmark as a Limit I: Single Asset Case

Supposed that the random variable δ is a truncated normal random variable with bounds $a < b$. That is, there exists a normal random variable X with mean μ and variance σ such that the random variable δ satisfies

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

Moreover, for simplicity, set

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

Then,

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

It follows that

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

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

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

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.

C.3 CARA-Normal Benchmark as a Limit I: Multi-Asset Case

Suppose that the random variable δ is a truncated multivariate normal random variable. That is, there exists a multivariate normal random variable X with mean μ and covariance Σ such that the random variable δ satisfies

$$\delta \sim X \text{ conditional on } -b < X_i < b\,,$$

for a positive real number b. Define

$$I_b = \{x \in \mathbb{R}^N | -b < x_i < b, \forall i\} \quad \text{and} \quad \mu_b = E[\mathbb{1}_{I_b}(\delta)]$$

. For simplicity, set

$$R_f = 1\,.$$

These assumptions imply that

$$\begin{aligned} e^{-\gamma f(q)} &= E[e^{-\gamma(x_0+q)^\top \delta}] \\ &= \int_{\mathbb{R}^N} e^{-\gamma(x_0+q)^\top y} \mathbb{1}_{I_b}(y) f_\delta(y) dy \\ &= \int_{\mathbb{R}^N} e^{-\gamma(x_0+q)^\top y} \mathbb{1}_{I_b}(y) f_X(y) \frac{1}{\mu_b} dy\,. \end{aligned}$$

Suppose that $b > b_0 > 0$. We have

$$\left| e^{-\gamma(x_0+q)^\top y} \mathbb{1}_{I_b}(y) f_X(y) \frac{1}{\mu_b} \right| < e^{-\gamma(x_0+q)^\top y} f_X(y) \frac{1}{\mu_{b_0}}\,. \tag{69}$$

Moreover, the left-hand side is integrable since X is a multivariate normal random variable. This shows that

$$e^{-\gamma(x_0+q)^\top \delta}$$

is uniformly integrable. The Bounded Convergence Theorem implies that

$$\lim_{b\to\infty} g(q) \;=\; g_X(q)\,,$$

where g_X is the function g under the assumption that the payoffs are multivariate normal distributions X. A similar approach establishes that

$$\lim_{b\to\infty} g^{(n)}(q) \;=\; g_X^{(n)}(q)\,,$$

Inequality 69 also implies that the BCT applies to $I(q)$ and $\Lambda(q)$:

$$\begin{aligned}
\lim_{b\to} I(q) &= \frac{L-1}{R_f}\lim_{b\to}\int_1^\infty \xi^{-L}g'(-\gamma(\xi q+x_0))d\xi. \\
&= \frac{L-1}{R_f}\int_1^\infty \lim_{b\to}\xi^{-L}g'(-\gamma(\xi q+x_0))d\xi. \\
&= \frac{L-1}{R_f}\int_1^\infty \xi^{-L}g'_X(-\gamma(\xi q+x_0))d\xi. \\
\lim_{b\to}\Lambda(q) &= \frac{\gamma}{R_f}\int_1^\infty \xi^{1-L}g''_X(-\gamma(\xi q+x_0))d\xi.
\end{aligned}$$

This completes the proof since the equilibrium is unique in the Gaussian case is unique.

D 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

$$\pi(t; s) = |s^\top (P(ts) - P(0))|. \tag{70}$$

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 10. *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 10.

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

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

It follows that

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

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

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

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

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

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

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

Correspondingly,

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

from which the statement follows.

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

Note that

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

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

$$\pi''(t;\gamma) = -\frac{\partial^2}{\partial t^2}\iota(t/L;\gamma) \tag{71}$$

$$= -L^{-2}\gamma^2\frac{L-1}{R_f}\int_1^\infty \xi^{2-L}\,\mathrm{rg}_{111}(-\gamma\xi t, -\gamma\xi; q, x_0)d\xi. \tag{72}$$

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 $s < 0$,

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

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 (71) 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 (71) 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.

Alex Boulatov and Dan Bernhardt. Robustness of equilibrium in the kyle model of informed speculation. Annals of Finance, 11(3-4):297–318, 2015.

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.

Adrian Buss and Grigory Vilkov. Measuring equity risk with option-implied correlations. The Review of Financial Studies, 25(10):3113–3140, 2012.

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.

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.

Marcin Kacperczyk, Stijn Van Nieuwerburgh, and Laura Veldkamp. A rational theory of mutual funds' attention allocation. Econometrica, 84(2):571–626, 2016.

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.

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.