Illiquidity and Higher Cumulants*

Sergei Glebkin, Semyon Malamud, and Alberto Teguia

June 20, 2022

Abstract

We characterize the unique equilibrium in an economy populated by strategic CARA investors who trade multiple risky assets with arbitrarily distributed payoffs. We use our explicit solution to study the joint behavior of illiquidity of option contracts. Option bid-ask ask spreads are proportional to risk aversion and risk-neutral variances of option payoffs. Contrary to the conventional wisdom, spreads may decrease in risk aversion, physical variance, open interest, and increase after earnings announcements. All these predictions are confirmed empirically using a large panel dataset of US stock options.

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

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

---

*Sergei Glebkin (glebkin@insead.edu) is at INSEAD, Semyon Malamud (semyon.malamud@epfl.ch) is at EPFL and the Swiss Finance Institute, and Alberto Teguia (alberto.mokakteguia@sauder.ubc.ca) is at UBC Sauder. This paper was previously circulated under the title "Asset Prices and Liquidity with Market Power and Non-Gaussian Payoffs." For their valuable feedback, we thank Bruno Biais, Svetlana Bryzgalova, Georgy Chabakauri, Jean-Edouard Colliard, Bernard Dumas, Daniel Ferreira, Thierry Foucault, Mike Gallmeyer (discussant), Craig Holden, Yunzhi Hu (discussant), 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, S."Vish" Viswanathan (discussant), Kathy Yuan, and Kostas Zachariadis. We are especially grateful to Marzena Rostek for her comments and suggestions. Semyon Malamud gratefully acknowledges financial support from the Swiss Finance Institute and the Swiss National Science Foundation (project Nr. 100018_192692).

1 Introduction

Illiquidity, or the market's inability to accommodate large trades without a price change, impacts the trading and pricing of financial assets even for a market as developed as US equities.[1] The effects of illiquidity are particularly severe for derivative contracts, where even short-term at-the-money (ATM) options written on the largest stocks can have bid-ask spreads on the order of 2%.[2] Large traders and institutional investors, such as mutual and pension funds, respond to illiquidity by trading *strategically*—that is, accounting for their price impact. Some investors (e.g., J.P. Morgan and Citigroup) have in-house "optimal execution" desks that devise trading strategies to minimize trading costs. Other investors use 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] Importantly, modern markets are largely dominated by institutions managing billions of dollars and, hence, strategic trading might be a key driver of a typical investors' behavior.

How are illiquidity and asset prices determined in equilibrium when investors internalize their price impact? The literature on strategic trading addresses this question adopting a *CARA-normal* framework for tractability. Traders are either risk-neutral or have constant absolute risk aversion (CARA) utility functions, and asset payoffs have Gaussian distribution. These assumptions make such models inapplicable to derivative markets, where payoffs are nonlinear functions of the underlying asset prices and, hence, cannot be Gaussian. Notably, multiple derivatives written on the same asset must be studied jointly. Thus, to study illiquidity

[1]For example, Koijen and Yogo (2019) estimate that—for the median US risky asset—the price impact of a 10% demand shock was consistently greater than 20% between 1980 and 2017. Koijen and Yogo (2019) also document that most of the variation in the cross-section of risky asset 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 risky asset returns. That such shocks would not affect returns in a perfectly liquid market underscores the importance of illiquidity for the cross-section of risky asset returns.

[2]For example, on May 17, 2021, the last date in our sample, average bid-ask spreads for ATM options with one month to expiry (the most liquid and actively traded contracts) on Apple Inc. were about 1.5%. This is about 30 times higher than a spread of about 0.05% for the Apple Inc. stock. Muravyev and Pearson (2020) argue that the true option bid-ask spreads are about one-quarter smaller than direct estimates. Yet, even after this correction, spreads remain very large.

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

in derivative markets, we need a model of strategic trading for multiple assets with non-Gaussian payoffs. Our paper develops such a model.

We allow for multiple assets and a general distribution of asset payoffs. Despite significant technical challenges, we characterize equilibrium explicitly and are able to derive its properties analytically. In an application of our theory, we derive several surprising implications regarding option bid-ask spreads and provide supporting empirical evidence. In particular, option bid-ask spreads may decrease in risk aversion, physical variance, and open interest, but they may increase after earnings announcements.

We assume that CARA traders, whom we refer to as *liquidity providers* (LPs), exchange multiple risky assets for a riskless asset over one period while internalizing their price impact. LPs all have the same risk aversion and are symmetrically informed. The absence of information asymmetry implies that, in our setting, the unique source of price impact is inventory risk.[4] Trading is organized as a uniform-price double auction: Traders simultaneously submit demand functions specifying the number of units of the 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.[5] In addition to LPs, there are uninformed *liquidity demanders* (LDs) who submit market orders. We express all equilibrium quantities as functions of the aggregate liquidity demand.

In equilibrium, traders need to determine their optimal demand functions (the map from the vector of prices to the vector of positions), knowing the demand functions of all other traders. However, we show that an LP's problem is, in fact, equivalent to that of a trader just knowing, for each order size, his price impact matrix (i.e., how his trades move prices of all assets at the

[4]For the sake of tractability, we abstract from asymmetric information. However, our focus on inventory risk is justified by recent empirical results documenting that inventory risk is a dominant source of price impact in options markets (Muravyev (2016)).

[5]We can also handle distributions with unbounded support. For example, in the benchmark case of a Gaussian distribution, 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 similarly handle any distribution with unbounded support (for which the limit just described exists).

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.[6] The equilibrium price impact matrix is pinned down by the requirement that it is consistent with the demand functions of all other traders and must be equal to the inverse of the "slope" of the total residual supply of all other traders for any level of liquidity demand. We show that optimality and consistency conditions imply a partial differential equations (PDEs) system for the equilibrium demand. Remarkably, solving this complex system of PDEs can be reduced to solving a single-asset ordinary differential equation (ODE). Such an ODE is linear and, thus, can be solved in closed-form.[7] This ODE characterizes the price function in an economy whose single asset is an index defined by a vector of asset holdings. We establish equilibrium uniqueness in the class of symmetric equilibria with strictly decreasing, continuously differentiable demands and arbitrage-free equilibrium prices.

We then look at the implications for bid-ask spreads and derive several surprising results.[8] For example, we show that bid-ask spreads may decrease in LPs' risk aversion and the size of their inventory, when LPs' risk aversion is high. The key to our surprising result is the interaction between LPs' inventories and bid-ask spreads. Consider what happens when LPs' initial inventory increases. The LP will decrease the price (both bid and ask) to reflect the higher inventory cost. Such an effect—a decrease in price when inventory increases— is present in both CARA-normal models and our general model. The effect that is unique to our model is that the ask price will decrease by more than the bid. A more profound decrease in the ask price makes buying from LPs relatively more attractive, which LPs like: They are eager to decrease inventory when risk aversion is high. Such interaction is absent in CARA-normal models, where the size of inventory affects the price level but not the bid-ask spread. The same interaction is behind our other surprising results: Bid-ask spreads may decrease in LPs'

[6] 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.

[7] Similar ODEs arose and were analyzed in Klemperer and Meyer (1989), Bhattacharya and Spiegel (1991), Wang and Zender (2002), and Boulatov and Bernhardt (2015).

[8] Our model has the same surprising implications for price impacts. We focus here on bid-ask spreads because this is the illiquidity measure we work with in our empirical exercise.

risk aversion and physical variance and may increase after earnings announcements—again, in contrast to CARA-normal models. See, e.g., Vayanos and Wang (2012) for a review.

We also derive closed-form expressions for bid-ask spreads and show that they are proportional to LPs' risk-aversion and risk-neutral variance of asset payoffs. These expressions in the general case are remarkably similar to those in the Gaussian benchmark; one needs to make only the minor adjustment of substituting the risk-neutral variance for the physical variance. However, this minor adjustment changes comparative statics dramatically, as the risk-neutral variance depends on model parameters such as risk aversion. It is this dependence that is responsible for our surprising results mentioned above. The empirical implication of our closed-form expressions is that the cross-section of bid-ask spreads of options is explained by the risk-neutral (not physical) variance of these options' payoffs.

In the paper's empirical part, we confront this and other predictions using US options data. In stark contrast to the conventional wisdom, and in line with the surprising implications of our theory, we find: (i) a *negative* relationship between bid-ask spreads and VIX, commonly interpreted as a proxy for market-wide risk aversion;[9] (ii) a non-monotonic relationship between bid-ask spreads and physical variance; (iii) an increase in the bid-ask spread following earnings announcements; and (iv) a negative relationship between option bid-ask spreads and the size of LPs' inventory, proxied by the options' open interest.[10] In line with our explicit solutions showing that bid-ask spreads are proportional to LPs' risk-aversion and risk-neutral variance of options payoffs we find that: (i) bid-ask spreads are positively related to an exact measure of options risk-neutral variance that we construct from option prices using the Carr and Madan (1998) payoff decomposition formula and (ii) the betas in the regression of spreads on risk-neutral variances are positively related to proxies of risk aversion.

Our findings stand in stark contrast with existing evidence for the stock market liquidity. First, equity market liquidity and VIX are negatively related (Nagel (2012)). Second, a release

[9]Our results are unchanged if we use Bekaert, Engstrom, and Xu (2021) measure of risk aversion instead of VIX.

[10]As options are in zero net supply, open interest is a natural proxy for the absolute size of LPs' inventories.

of public information (disclosure) is associated with improved liquidity for the underlying stock (Healy and Palepu (2001)). Third, there is a negative association between stock market liquidity and the size of market makers' inventories (Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010)). All three relationships hold in the opposite direction for stock options.

Although this contrasting evidence might be initially surprising, both sets of findings are consistent with our theory if we allow for some form of market segmentation. Indeed, we show that unconventional results only hold when LPs' risk aversion is sufficiently high. In contrast, we recover conventional results when this risk aversion is small. We argue that LPs in options markets might have a much smaller risk-bearing capacity than LPs in equity markets. First, this capacity is exhausted much more quickly due to options' embedded leverage. Second, LPs have many options contracts to intermediate, even for a single underlying asset. Much higher average percentage bid-ask spreads in the options markets also point toward a significantly lower risk-bearing capacity for options' LPs.

The rest of our paper proceeds as follows. Section 2 presents the model. Section 3 considers equilibrium in the model and presents closed-form solutions. Section 4 derives implications for bid-ask spreads and presents supportive empirical evidence. Section 5 reviews the related literature. We conclude in Section 6 with a summary and some suggestions for future research. Technical details are relegated to the appendices.

2 The Model

There are two time periods $t \in \{0, 1\}$.[11] A number $L > 2$ of strategic *liquidity providers* (LPs) trade assets with *liquidity demanders* (LDs) at $t = 0$ and consume at $t = 1$.[12] There are N risky

[11] In the baseline model presented in this section, all securities pay off at $t = 1$, when the consumption also takes place. In our empirical application of the US options market, the options have different maturities. In Section IA.4, we show how the model can easily be extended to allow for multiple maturities and multiple consumption dates (while still having one trading date).

[12] Under some technical conditions, 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).

assets and a risk-free asset (a bond). The bond is a numeraire and, thus, earns a net return of zero. A risky asset 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:[13]

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

We define skewness as the third central moment of the distribution, as follows:

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

We put forward the following technical restrictions on the model parameters.

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 and, hence, there are no redundant securities.*

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

Assumption 1 simply requires that there be no redundant securities.[14] 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

[13]The link between derivatives of $g(\cdot)$ of order higher than 3 and the central moments of δ is more complex. See, for example, Shiryaev (1996), Ch II.12 for more general formulas.

[14]Redundant securities in our model are priced by arbitrage. We exclude them to streamline the exposition.

natural when one considers that the resources of any firm are limited, which means that no asset can have an infinite payoff.

As is common in the literature, we assume that LDs' aggregate trade is characterized by the aggregate supply shock $s \in \mathbb{R}^N$, which has full support[15] and is independent of δ. As in Klemperer and Meyer, our assumptions imply that equilibrium quantities will depend on the realization of s but not its distribution.

The LPs are identical, are symmetrically informed, and maximize the expected CARA utility from their terminal wealth W while accounting for their price impact. Each LP is initially endowed with the same portfolio x_0. In a symmetric equilibrium, all traders submit identical demand functions $D(p)$. Then, the optimal demand $D^i(p)$ for trader i given the residual demand $(L-1)D(p)$ of other $L-1$ traders solves the following problem:

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

Before detailing our equilibrium concept, we define the set of arbitrage-free prices.

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

As is common in the literature, we focus on arbitrage-free, symmetric Nash equilibria with strictly decreasing demands (hereafter, simply "an equilibrium").

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, $D^i(p) = D(p)$ solves problem (1). (ii) The function $D(p)$ is strictly decreasing—that is, $(D(p) - D(\hat{p}))^\top (p - \hat{p}) < 0$*

[15] Supply uncertainty is needed to rule out the extreme multiplicity of equilibria (cf. Klemperer and Meyer (1989); Vayanos (1999)).

for all $p \neq \hat{p}$. (iii) The function $D(p)$ is continuously differentiable, and the Jacobian ∇D is nondegenerate everywhere. Let $I(\cdot)$ denote the inverse of $D(\cdot)$.[16] *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. Solving for the equilibrium amounts to solving an ODE, and this requirement places a transversality condition that yields a unique solution. The economic meaning of condition (iv) is as follows. Suppose that, in addition to strategic LPs, there is an arbitrarily small mass of competitive (price-taking) LPs. Then, for prices that are *not* arbitrage-free, the price-taking LPs would submit infinite demands; thus, the market would not be 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 robust to the presence of a vanishingly small number of competitive LPs.

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 $\exp(-\gamma f(q; x_0)) \equiv E_\delta[\exp(-\gamma(x_0 + q)^\top \delta)]$. The certainty equivalent $f(q)$ is related to the CGF as follows:

$$f(q; x_0) = -\frac{1}{\gamma} g(-\gamma(x_0 + q)). \tag{2}$$

We will often suppress the second argument, and simply write $f(q)$, provided that no confusion could arise. The following remarks are in order.

Remark 1 (On bounded support of δ)*.* Assumption 2 ensures equilibrium uniqueness, as we discuss below, following Proposition 2. However, our benchmark case with Gaussian distribution does not satisfy Assumption 2. In Appendix IA.7, we analyze the case of unbounded support

[16] 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.

as a limit of our model. In the Gaussian case, the procedure selects the linear equilibrium commonly considered in the literature. We also note that some of our key results do not require boundedness (Proposition 3) or only require that δ be bounded on one side (Proposition 4).

Remark 2 (Symmetric information). Absence of information asymmetry implies that the only source of the price impact in the model is inventory risk. Our focus is on the effects of departures from normality on illiquidity, and we keep the model simple in other dimensions. Our empirical application is the US options market, where it has been shown that inventory risk is a dominant factor in options' illiquidity (Muravyev (2016)).[17] Thus, our model with inventory risk is a good starting point to understand the illiquidity of US options. Further research can enrich our model to also incorporate information asymmetry. We also show below (both theoretically and empirically) that option bid-ask spreads increase after the release of public information. We believe that it would be nontrivial to generate this surprising behavior in a model driven purely by asymmetric information. Indeed, in a typical asymmetric information model, the bid-ask spread increases when information asymmetry increases.

3 Equilibrium

This section provides a characterization of equilibrium and derives the closed-form solutions.

3.1 Characterization of equilibrium

We start with a heuristic derivation. Consider first a price-taking LP. His inverse demand $P = I(q)$ is determined by the first-order condition in terms of the certainty equivalent (2), as follows:[18]

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

[17] Earlier research (Cao and Wei (2010)) argues that asymmetric information is an important driver of option spreads. This claim is based on the observation that spreads co-move positively with volume. We do not find support for this result in our data. After controlling for other characteristics, spreads and volume are negatively related.

[18] See Appendix A for the summary of notation.

In contrast, 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. 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:

$$\nabla f(q) - \Lambda(q) q = P. \tag{3}$$

The price impact $\Lambda(q)$ is determined by the consistency condition. To derive it, suppose that a trader of interest modifies her demand, while other traders' demands are still given by $I(q)$. The equilibrium price is given by $I(q_o^*)$, where q_o^* denotes equilibrium allocation to other traders. If the trader of interest increases his equilibrium allocation by dq, then by market clearing, q_o^* changes by $-dq/(L-1)$. Therefore, $\Lambda(q)$ is related to the Jacobian of $I(q)$ as follows:

$$\Lambda(q) = \frac{-1}{L-1} \nabla I(q). \tag{4}$$

The *optimality* condition (3) and the *consistency* condition (4) result in the following system of partial differential equations:

$$\nabla f(q) + \frac{1}{L-1} \nabla I(q) q = I(q). \tag{5}$$

The next theorem summarizes our equilibrium characterization.

Theorem 1. *(Equilibrium characterization) A strictly decreasing function $I(q)$ is an equilibrium inverse demand if and only if 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 $\Lambda(q)$,

$$I(q) = \nabla f(q) - \Lambda(q)q. \tag{6}$$

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

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

(iii) No free lunch:

$$I(q) \in \mathcal{A} \qquad for\ all\ q. \tag{8}$$

3.2 Closed-form solution

Theorem 1 implies that finding equilibrium demand $I(q)$ is reduced to finding a solution to the system of PDEs (4) that is strictly decreasing and that satisfies the no-arbitrage restriction (8). It is instructive to write (5) in the special case of a single risky asset, where it becomes a linear ODE:

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

Such an ODE can be solved in a closed form using standard methods.[19] In contrast, solving (systems of) PDEs usually presents significant technical challenges. Surprisingly, we now show that solving the (seemingly complex) system of PDEs (5) boils down to solving a linear ODE that is similar to (9). Hence, our approach is tractable even in the case of multiple assets.

To gain some intuition behind our approach, consider an economy in which all investors (including LDs) can trade only a single index (portfolio) $q \in \mathbb{R}^N$. We refer to this as a *restricted* economy and to our baseline case as an *unrestricted* economy. Let $\iota(t)$ denote the inverse

[19]Similar ODEs arose and were analyzed in Klemperer and Meyer (1989), Bhattacharya and Spiegel (1991), Wang and Zender (2002), and Boulatov and Bernhardt (2015).

demand that LPs submit for $t > 0$ units of the index. The certainty-equivalent utility they derive from holding those t units is denoted $\phi(t)$

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

Since the restricted economy is a one-asset economy, $\iota(t)$ must satisfy the ODE (9) for every $t > 0$:

$$\frac{d}{dt}\phi(t) + \frac{t}{L-1}\frac{d\iota(t)}{dt} = \iota(t). \tag{10}$$

Now consider the unrestricted economy. In the symmetric equilibrium, for supply shock realizations $s = tq$ ($t \in \mathbb{R}^+$), it should be optimal for LPs to absorb $1/L$ fraction of supply shock s—that is, to trade t/L units of portfolio q. Hence, the price LPs bid for t/L units of portfolio q in the unrestricted economy, or $q^\top I(t/Lq)$, should be an optimal bid in the restricted economy. Therefore, $q^\top I(t/Lq) = \iota(t/L)$ should satisfy ODE (10), which completes the first step.

The second step in solving for equilibrium demand $I(q)$ requires solving the ODE (10) for inverse demand $q^\top I(tq) = \iota(t)$. 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 previous definition of expenditure $e(q)$ with respect to q:

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

Combining this equality with (4) and (6) yields

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

completing the derivation. This approach is summarized in our next proposition.

Proposition 1. *(From PDE to ODE) The inverse demand $I(q)$ satisfies the system (5) if*

$$I(q) = 1/L \cdot \nabla e(q) + (1 - 1/L)\,\nabla f(q). \tag{11}$$

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 t units of portfolio q that satisfies the ODE

$$\iota(t; q) = \frac{d}{dt} f(tq) + \frac{t}{L-1} \frac{d\iota(t; q)}{dt} \tag{12}$$

for every $t > 0$.

As we show in the Appendix, the assumption of bounded support implies that there is only one solution to the ODE (12) such that $I(q) \in \mathcal{A}$. This solution is given by

$$\iota(t; q) = (L-1) \int_1^\infty \xi^{-L} \phi'(t\xi)\, d\xi \; = q^\top \bigg((L-1) \int_1^\infty \xi^{-L} \nabla f(t\xi q)\, d\xi \bigg). \tag{13}$$

Hence the expenditure can be written as

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

In the second step, we differentiate this equation with respect to q and then apply (11) to obtain (14) for $I(q)$. It then remains to establish the global optimality (1) of $D(p) = I^{-1}(q)$. This is highly nontrivial due to the complex, non-convex nature of (1). In Appendix B.1, we develop novel mathematical techniques to tackle it. The following is true.

Proposition 2. *(Closed-form solution) There exists a unique equilibrium. The equilibrium inverse demand $I(q)$ and the price impact matrix $\Lambda(q)$ are given by, respectively:*

$$\begin{aligned} I(q; x_0) &= (L-1) \int_1^\infty \xi^{-L} \nabla f(\xi q)\, d\xi && (14)\\ &= (L-1) \int_1^\infty \xi^{-L} \nabla g(-\gamma(x_0 + \xi q))\, d\xi; && (15)\\ \Lambda(q; x_0) &= -\int_1^\infty \xi^{1-L} \nabla^2 f(\xi q)\, d\xi && (16)\\ &= \gamma \int_1^\infty \xi^{1-L} \nabla^2 g(-\gamma(x_0 + \xi q))\, d\xi. && (17) \end{aligned}$$

One of the key results of the above Proposition is the equilibrium uniqueness. Two features of the model help to achieve it: (a) the bounded support of δ and (b) the no-arbitrage restriction (iv) of Definition 2. Essentially, (a) and (b) help to select the unique solution of the ODE (12) among the continuum of solutions. Indeed, one can show that the general solution to (12) is given by $\hat{\iota}(t;q) = \iota^*(t;q) + C \cdot t^{L-1}$, where C is a constant and $\iota^*(t;q)$ is given by (13).[20] If $C < 0$, then $\hat{\iota}(t;q)$ becomes arbitrary small as t increases, so $\hat{\iota}(t;q)$ becomes smaller than the minimum payoff of the portfolio q (which is finite, due to boundedness of δ). A price less than a minimal payoff clearly violates no-arbitrage restriction. A similar argument applies to the case $C > 0$, where the price will be above maximal payoff. Thus, the only solution that could satisfy the no-arbitrage restriction is $\iota^*(t;q)$. Note that the boundedness of δ is important here, without it, the solutions of (12) with $C \neq 0$ will not violate the no-arbitrage restrictions, because the maximum or minimum payoffs on the portfolio q are not finite.

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 as well as gain additional insights. In what follows we use an asterisk (*) to denote those moments of δ evaluated under the EMM.

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

$$I(q) = E[Z^*(q)\delta] = E^*[\delta], \tag{18}$$

where

$$Z^*(q) = (L-1)\int_1^\infty t^{-L}\zeta(t;q)\,dt. \tag{19}$$

Note that $\zeta(t;q)$ is an SDF in a competitive economy, where LPs absorb an order of size $t \times q$. Equation (19) shows that Z^* (i.e., the SDF in the economy with market power)

[20] The general solution to (12) is a sum of a particular solution (13) and a solution to the homogeneous ODE, in our case, $\iota(t;q) = \frac{t}{L-1}\frac{d\iota(t;q)}{dt}$. The solution to the homogeneous ODE is $\iota = C \cdot t^{L-1}$.

is a weighted average of the SDFs in the competitive economies, where LPs absorb orders of size $t \times q$ with $t > 1$. This outcome is intuitive: LPs exercise their market power by charging the price that competitive LPs would charge for absorbing a larger order. Thus, (19) manifests the *demand reduction* common to auctions of divisible goods (see Ausubel, Cramton, Pycia, Rostek, and Weretka (2014)).

The function $Z^*(q)$ is a Radon-Nikodym derivative for the risk-neutral measure in the economy with the per-capita supply shock q. Therefore, $Z^*(0) = \exp(-\gamma x_0^\top \delta)/E[\exp(-\gamma x_0^\top \delta)]$ is associated with the risk-neutral measure in the economy where there is no supply shock. Let the density of δ be $\eta(\delta)$. In the sequel, we refer to the probability measure with the density $\eta(\delta) Z^*(0)/E[Z^*(0)]$ as the *risk-neutral measure.* We use superscript * to denote moments under this measure. Below, we show that bid-ask spreads are related to the second moments of the distribution under this measure.

4 Bid-Ask Spreads

In this section we derive implications of our theory for the joint behavior of bid-ask spreads when payoffs are non-Gaussian. We then test our key predictions using a large panel of exchange-traded options on US stocks.

4.1 Bid-ask Spreads: Theory

We define the bid-ask spread for an asset k as the difference in equilibrium prices when LPs absorb (buy or sell) a small amount of n_k units of asset i, normalized by n_k. We interpret n_k as a minimal lot size.

$$\mathrm{BA}_k \equiv \lim_{n_k \to 0} \frac{I_k(-n_k 1_k) - I_k(n_k 1_k)}{n_k} = 2(L-1)\Lambda_{kk}(0). \tag{20}$$

The last equality is a direct corollary of Proposition 2. The advantage of such a measure, compared to the one without normalization by n_k, is that it is independent of the minimal lot size.[21]

We start by characterizing bid-ask spreads for the Gaussian case.[22] We are interested in how bid-ask spreads change when (i) risk aversion increases; (ii) public information is released; and (iii) there is a systematic increase in riskiness of the assets,[23] modeled as a rescaling of the distribution of δ, preserving its mean. To have a shift parameter that affects the riskiness of the assets, in this section we maintain the following assumption.

Assumption 3. *The dividend vector is given by $\hat{\delta} = E[\delta] + \sigma(\delta - E[\delta])$. The scalar parameter σ is called asset riskiness.*

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

$$\mathrm{BA}_k = 2\frac{L-1}{L-2}\gamma\sigma^2\Sigma_{kk}. \tag{21}$$

Correspondingly, BA_k increases in γ, σ, and Σ_{kk}. Suppose, in addition, that traders observe a public signal $s_p = \delta + u$, where $u \sim N(0, \Sigma_u)$. Denote by $\mathrm{BA}_k(s_p)$ (resp., $BA_k(\emptyset)$) the bid-ask spread after (resp., before) observing the signal. Then, $BA_k(s_p) < BA_k(\emptyset)$; i.e., a release of public information decreases the bid-ask spread for every asset. Furthermore, BA_k is independent of x_0.

The comparative statics in Corollary 2 are standard for CARA-normal models. As the risk faced by LPs decreases due to the release of information or a decrease in the asset riskiness σ, they demand a lower compensation for providing liquidity and the bid-ask spread decreases. Similarly, bid-ask spread widens with an increase in risk aversion. In stark contrast to this conventional wisdom, empirical evidence from stock options data (see Section 4.2) documents

[21]Most of our results would also hold for the percentage bid-ask spread measure $\mathrm{BA}_k^{\%}$, where the BA is normalized by the mid-price $I_k(0)$—in other words, $\mathrm{BA}_k^{\%} = 2(L-1)\Lambda_{kk}(0)/I_k(0)$. However, the comparative statics of such a measure would be driven by both illiquidity Λ and prices $I_k(0)$. Thus, BA_k is our preferred measure, as it solely reflects illiquidity.

[22]Formally, Gaussian distributions do not satisfy Assumption 2. However, as we show in the Appendix, the unique linear equilibrium of the Gaussian model can be constructed as the limit of bounded-support equilibria obtained by truncating the payoff distribution.

[23]For example, when volatility increases for the underlying stock, this constitutes a systematic increase in riskiness for all derivatives written on the stock.

(i) a *negative* relationship between bid-ask spreads proxies for market-wide risk aversion; (ii) a non-monotonic relationship between bid-ask spreads and physical variance; and (iii) an increase in bid-ask spread following the earnings announcements. We also find a negative relationship between option bid-ask spreads and the size of LPs' inventory $|x_{0,k}|$, proxied by the option open interest. This result clearly contradicts equation (21): In a CARA-normal setting, inventory has no impact on bid-ask spreads.

Below, we theoretically demonstrate that the comparative statics of Corollary 2 might change signs when we abandon the Gaussian assumption, in line with our empirical results. We first consider the comparative statics for risk aversion γ and the asset riskiness σ and then derive necessary and sufficient conditions for $\partial \mathrm{BA}_k/\partial\gamma < 0$ and $\partial \mathrm{BA}_k/\partial\sigma < 0$ in Proposition 3. We next consider the case when LPs' risk aversion is high and show in Proposition 4 that all results highlighted in Corollary 2 are overturned in that case.

Proposition 3. *Denote* $\Sigma^*_{kk} = \mathrm{var}^*(\delta_k)$. *The bid-ask spread can be written as*

$$BA_k = 2\frac{L-1}{L-2}\gamma\sigma^2\Sigma^*_{kk}. \tag{22}$$

Moreover,

$$\mathrm{sign}\left(\frac{\partial}{\partial\gamma}BA_k\right) = \mathrm{sign}\left(\Sigma^*_{kk} - \gamma\sigma\mathrm{coskew}^*(\delta_k, \delta_k, \delta^\top x_0)\right) \ \ \textit{and} \tag{23}$$

$$\mathrm{sign}\left(\frac{\partial}{\partial\sigma}BA_k\right) = \mathrm{sign}\left(2\Sigma^*_{kk} - \gamma\sigma\mathrm{coskew}^*(\delta_k, \delta_k, \delta^\top x_0)\right). \tag{24}$$

For small enough γ*, we have* $\mathrm{sign}\left(\frac{\partial}{\partial\gamma}BA_k\right) = \mathrm{sign}\left(\frac{\partial}{\partial\sigma}BA_k\right) > 0$.

There are two main takeaways from Proposition 3. First, the expression for bid-ask spread remains remarkably similar to that in the Gaussian case, with physical variance substituted for the risk-neutral one (compare (21) and (22)). Importantly, the risk-neutral variance depends on model parameters such as the risk aversion γ. The co-skewness terms in (23) arise precisely because of that: We show in the proof that they are proportional to the sensitivity of

$\text{var}^*(\delta_k)$ to γ (see Lemma 8).

Second, the proposition gives necessary and sufficient conditions for the Gaussian comparative statics with respect to the γ and σ to be overturned. This happens when the term $\gamma \cdot \text{coskew}^*(\delta_k, \delta_k, \delta^\top x_0)$ is large compared to $\text{var}^*(\delta_k)$. Thus, if the distribution of δ has zero higher cumulants or if risk aversion γ is small, we have $\text{sign}\left(\frac{\partial}{\partial\gamma} BA_k\right) = \text{sign}\left(\frac{\partial}{\partial\gamma} BA_k\right) = \text{sign}\left(\text{var}^*(\delta_k)\right) > 0$, explaining why in the Gaussian case the bid-ask spreads are always increasing in γ and also why the same is true when γ is small (even if the distribution is non-Gaussian).[24]

We next show that, when γ increases, the Gaussian comparative statics are necessarily reversed.

Proposition 4. *Suppose that the density of δ, $\eta(\delta)$, is strictly positive everywhere on its support. Suppose also that $x_{0,k} \neq 0$. Then, for large enough γ, we have $BA_k = 2\frac{L-1}{L-2}\gamma^{-1}x_{0,k}^{-2} + O(\gamma^{-2})$, where the term $O(\gamma^{-2})$ depends on η. Furthermore, for large enough γ, the bid-ask spread*

- *decreases in the risk-aversion γ,*
- *decreases in $|x_{0,k}|$, and*
- *can be increasing or decreasing in the physical variance of δ, $\Sigma_{k,k}$, depending on the shape of η.*[25]

Suppose further that there is a release of public information ι of the form $s_p = \delta + u$, where $u \sim N(0, \Sigma)$. Denote by $\text{BA}(s_p)$ the bid-ask spreads, given this information. Then, $E[\text{BA}(s_p)] - \text{BA}(\emptyset)$ is positive for small enough γ and changes sign (at least once) as γ increases.

The implications of Proposition 4 are intriguing. Contrary to conventional wisdom based on Corollary 2, (i) physical distribution has only second-order effects on bid-ask spreads and (ii) these spreads decrease in γ and depend on LPs' inventories, x_0. We now discuss these counter-intuitive comparative statics through the lens of formula (23). When γ is small, so is the second

[24]We note that the proposition does not require δ to be bounded. In Appendix B.6, we formulate and prove a more general version of Proposition 3 that does not require δ to be bounded.

[25]We provide the details of this dependence in the Appendix.

term in (23), and the bid-ask spread increases in γ. As γ increases, the risk-neutral measure puts progressively higher weight on the "bad" states of the world—that is, the states where the payoff to LPs' inventory of asset k, $x_{0,k}\delta_k$ is small. Such greater weight put on bad states of the world implies that the risk-neutral distribution of $x_{0,k}\delta_k$ will be concentrated around small realisations of $x_{0,k}\delta_k$ and thus skewed to the right, with a relatively small variance. See Figure 1 for an illustration. This makes the right-hand side of (23) negative, implying that the bid-ask spread decreases in γ.

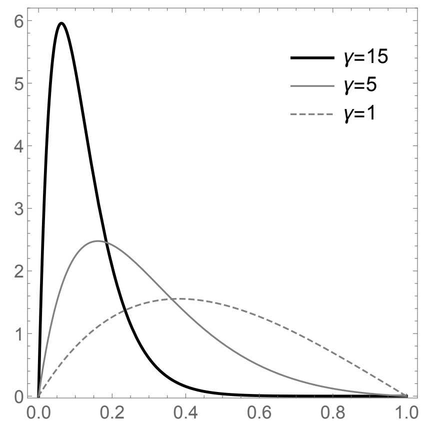

Figure 1: Risk-neutral distribution of $x_{0,k}\delta_k$ for $\gamma = 1, 5, 15$. We assume that $\delta \sim \text{Beta}[a, b]$, $a = b = 2$, $N = 1$, $x_0 = 1$, and $L = 5$.

4.2 Bid-ask Spreads: Empirics and Discussion

This section documents several surprising empirical patterns in the prices of options on US stocks that are consistent with our model's predictions.

4.2.1 Data description

We obtain *daily* option prices and Black-Scholes implied volatilities from Option Research and Technology Services (ORATS), a data provider for historical options quotes and implied volatilities.[26] The dataset covers the period of 2007-01-03 to 2021-05-17 and contains data for

[26] https://www.orats.com/.

options bids, asks, volume, open interest, implied volatilities, and the price of the underlying asset for all US stocks with traded options.[27]

Many prices of options on small stocks are stale, and the underlying option contracts have extremely low liquidity. To highlight that small stocks do not drive our findings, we focus on only 145 stocks comprising the NASDAQ index as of February 22, 2022. In addition, we pre-filter data to exclude options with a volume of fewer than 20 contracts[28], so we consider only those options with more than one and less than 60 days to expiry. We define the bid-ask spreads as $(Ask - Bid)/(0.5(Ask + Bid))$. Each option contract is uniquely determined by its underlying asset, expiration date, option type (call or put), and strike. To avoid dealing with extremely illiquid contracts, we keep only those option contracts in our panel for which (a) bid prices are positive for both call and put for that strike, (b) bid-ask spreads are positive for both put and call at that strike, and (c) either call or put at that strike have $Ask < 1.5 \cdot Bid$.

We use the rolling 20-day standard deviation of stock returns times the square root of the number of days to expiry as a measure of anticipated physical variance of the payoff at expiry.[29] We compute option moneyness using the standard formula

$$moneyness = log(strike/underlyingPrice)/\Sigma_{k,k}^{1/2}, \tag{25}$$

and we exclude options with moneyness above 2 and below -2, as these are extremely illiquid.

To test for potential non-monotonicity of BA spreads in a given variable X, we include both X and $X \cdot 1_{X>q}$ in the list of independent variables, where q is the 50% quantile (the median) of X across the whole panel. For example, var is variance and $\text{var} \cdot 1_{\text{var}>q}$ is variance times the indicator that this variance is above its panel upper 50% quantile value.[30]

[27]Most of the existing academic literature uses the IvyDB database provided by *OptionMetrics*. We chose to use ORATS instead for two reasons: First, unlike OptionMetrics, which is updated yearly, ORATS provides real-time data, which allows us to fully include recent data up to and including the recent COVID-crisis. Second, because the ORATS data are provided to us exactly as they are available to market participants in real time, we are certain that our results do not suffer from look-ahead bias.

[28]Options are written on lots of 100 shares of a stock. Thus, a volume of 20 contracts corresponds to a volume of 2000 stock shares.

[29]This is the natural counterpart of $\Sigma_{k,k}$.

[30]All our results remain virtually unchanged when we replace 50% with higher levels (such as 60%–80%).

4.2.2 Results and discussion

We split our results into three parts. Before we proceed we make a note on the notation. Throughout the section we use subscript s to index stocks, subscript k to index option contracts on this stock and subscript t to index dates. We use two proxies for risk aversion: (i) VIX, the CBOE S&P 500 implied volatility index and (ii) RA, the proxy for risk aversion from Bekaert et al. (2021) kindly provided on the authors website. For each option contract, $\Sigma^*_{s,k,t}$ denotes the proxy for the risk-neutral variance of it's payoff (see below for the description of the construction of the proxy) and $\mathrm{IV2}_{s,k,t}$ denotes the squared implied volatility for a particular option contract. We also use option volume, open interest (oi), implied volatility (IV) (provided by ORATS) and its square ($IV2 = IV^2$), as well as physical variance of the underlying stocks $\Sigma_{k,k}$ in our regressions.

Cross-section of bid-ask spreads and risk-neutral variances

Our key prediction (Proposition 3) is that bid-ask spreads are positively related to the product of the *risk-neutral variance of the payoff of the asset* times risk aversion of market participants, as follows:

$$\mathrm{BA}_{s,k,t} \;\propto\; \gamma_t \, \Sigma^*_{s,k,t}. \tag{26}$$

Constructing risk-neutral variance of the payoff is nontrivial. For example, for call and put options with strike K, these variances are given by

$$\begin{aligned} \mathrm{var}^*[(S-K)^+] &= E^*[((S-K)^+)^2] - (E^*[(S-K)^+])^2 \text{ and} \\ \mathrm{var}^*[(K-S)^+] &= E^*[((K-S)^+)^2] - (E^*[(K-S)^+])^2 \,. \end{aligned} \tag{27}$$

Thus, to compute the risk-neutral variance, one would need to observe prices of highly nonlinear derivatives with payoffs $((S-K)^+)^2$ and $((K-S)^+)^2$, respectively. While such derivatives are not directly available for trading, their prices can be constructed synthetically using the Carr and Madan (1998) payoff decomposition formula. We describe our empirical procedure in

Appendix IA.2. The outcome of this procedure is a proxy $\Sigma^*_{k,s,t}$ for the risk-neutral variance of each option contract. Using these risk-neutral variances, we run for every stock and date combination (s,t) the following panel regression:[31]

$$\mathrm{BA}_{s,k,t} \;=\; \alpha_{s,t} \;+\; \beta_{s,t}\Sigma^*_{s,k,t} + \epsilon_{s,k,t}. \tag{28}$$

That is, consistent with our theory, we assume that each date the markets for all options on a given stock are integrated, implying that the theoretical relationship of Proposition 3 should hold simultaneously for all options. For each stock and date combination (s,t), regression (28) is a linear panel regression across the whole panel of option contracts available on that date for that stock. For robustness, we also run this regression separately only for put and only for call options. Thus, we end up with the following three panels of β estimates: $\beta^{Call}_{s,t}$, $\beta^{Put}_{s,t}$, and $\beta_{s,t}$.

Proposition 3 implies the following empirical predictions for the behavior of the β estimates in (28):

(1) $\beta^{Call}_{s,t}$, $\beta^{Put}_{s,t}$, and $\beta_{s,t}$ are all positive.

(2) $\beta^{Call}_{s,t}$, $\beta^{Put}_{s,t}$, and $\beta_{s,t}$ are positively related to risk aversion, as proxied either by VIX or by RA.

We first look at prediction (1). We find that all three β estimates are positive for more than 99.95% of observations in our panel. They have a mean of about 1 and a median of about 0.6. These findings are in perfect agreement with our prediction (1). Moving on to prediction (2), we can see from Table 1 that both VIX and RA correlate positively with all three beta measures, with the RA correlation being significantly higher. We then take a deeper look at our prediction. We compute the correlation between β and proxies of risk aversion for every single stock and find that the correlation between β and RA is positive for 117 out of 145 stocks, with

[31] We use the standard OLS formula to compute the β coefficient: If $X = \alpha + \beta Y + noise$, then $\beta = \mathrm{cov}(X,Y)/\mathrm{var}(Y)$. However, due to extreme nonlinearities in the data, $\mathrm{var}(Y)$ is often unstable. We winsorize it to make it less sensitive to extreme outliers.

	RA	VIX	β^{Call}	β^{Put}	β
RA	1.00	0.81	0.39	0.26	0.28
VIX	0.81	1.00	0.35	0.11	0.16
β^{Call}	0.39	0.35	1.00	0.86	0.95
β^{Put}	0.26	0.11	0.86	1.00	0.97
β	0.28	0.16	0.95	0.97	1.00

Table 1: Correlation matrix of betas and proxies of risk aversion.

a mean of 0.1 and a standard deviation of 0.2, which, with 145 observations, gives a t-statistic of 6. All these findings stand in perfect agreement with our predictions.

Another way to test prediciton (2) is to perform a panel regression:[32]

$$\mathrm{BA}_{s,k,t} \;=\; a \;+\; b_1\,\Sigma^*_{s,k,t} \;+\; b_2\,RA_t \;+\; b_3\,RA_t \cdot \Sigma^*_{s,k,t} + \epsilon_{s,k,t}\,. \tag{29}$$

As Panels (3) of Tables 2 and 3 show, in both regressions the coefficient β_3 is positive and highly significant, consistent with the theory predictions. One interesting additional test we run is whether $\Sigma^*_{s,k,t}$ can be replaced by implied volatility. Although theoretically there is no direct relationship between implied volatility and $\Sigma^*_{s,k,t}$, intuitively one might expect that these quantities are related as both are related to volatility. We therefore run a version of (29), with $\Sigma^*_{s,k,t}$ replaced by squared implied volatility (IV2). As Panels (2) of Tables 2 and 3 show, β_3 is indeed positive for puts, but not for calls, suggesting that for calls, $\Sigma^*_{s,k,t}$ is capturing a dimension of risk-neutral variance that is not spanned by implied volatility.

Bid-ask spreads and risk-aversion, open interest, and physical variance

We now proceed to our other predictions. If one does not control for the risk-neutral variance, then conventional wisdom, largely based on the CARA-normal setting (see Corollary 2) suggests that the bid-ask spread is increasing in physical variance and risk aversion and that a release of public information should lead to a reduction in spreads. In contrast, Proposition 4 implies

[32]The results for the same regression but with RA replaced by VIX are similar and are reported in Tables 8 and 9.

that the opposite can be true when risk aversion γ is sufficiently high. We take RA as our main proxy for risk aversion here. The results when RA is replaced with VIX are very similar and are reported in Tables 8 and 9 in the Appendix.

For markets in which liquidity providers have a sufficiently high risk aversion, we predict that bid-ask spreads:

(1) may be decreasing in risk aversion,

(2) may be decreasing in open interest,

(3) may be decreasing in physical variance, and

(4) may increase after the release of public information.

We believe that options markets and the markets of the respective underlying assets are highly fragmented; as a result, risk aversion (the risk-bearing capacity of liquidity providers) differs across them. For example, a trading desk can be intermediating a particular subset of options; capital requirements are extremely high for illiquid option contracts, implying that the effective risk aversion (the ability of dealers to take on balance sheet risk) will be much higher for options than for stocks. Yet, aggregate risk aversion is known to move in cycles, driven by VIX (see Nagel (2012)). Hence, we think about the "true" γ as being proportional to RA, $\gamma = aRA$, but with the proportionality constant a being much higher for options than for stocks. For markets with a small a, we are in the "normal" regime, and Proposition 3 conforms with the findings for equities: Bid-ask spreads are increasing in risk-aversion (as proxied by VIX, see Nagel (2012)).[33] For markets with a large a, we are in the regime of Proposition 4, and the sign of the relationship may reverse.

Our findings for options agree with the predictions of Proposition 4. Namely, we find that a negative relationship between bid-ask spreads and RA (Panel (1) of Tables 2 and 3).[34] To the

[33]In addition, (i) a large body of accounting literature has found that the release of public information (disclosure) is associated with improved liquidity of equities (Healy and Palepu (2001)) and (ii) there is a negative association between stock liquidity and size of market makers' inventories Comerton-Forde et al. (2010).

[34]Note that a linear model $BA \sim a\,RA \; + \; b\,RA1_{RA>q}$ means that the slope of the relationship is a for data points with $RA < q$, while the slope of $a + b$ for data points with $RA > q$.

best of our knowledge, our paper is the first to document (and provide a theoretical foundation for) this surprising and counterintuitive relationship between risk aversion and option bid-ask spreads.[35]

We now proceed to our second novel prediction: a negative relationship between option bid-ask spreads and open interest.[36] Although this link has not received much attention in academic literature, it is an important part of option traders' folklore. For example, according to Investopedia:[37] *"Open interest also gives you key information regarding the liquidity of an option... All other things being equal, the bigger the open interest, the easier it will be to trade that option at a reasonable spread between the bid and ask."* Panel (1) of Tables 2 and 3 confirms this intuition, which is also fully consistent with the result of Proposition 4.

We now proceed to discussing the dependence on physical variance. Panel (1) of Table 3 shows that for put options, this dependence is quite subtle: The relationship is strongly negative for variances below the panel median and then becomes positive for the part of the sample above the median.[38] We interpret this result as strong evidence for the importance of nonlinear models in explaining the illiquidity of options. Finally, for completeness, we also include option volume in the list of explanatory variables. Cao and Wei (2010) argue that the relationship between bid-ask spreads and volume is positive, potentially driven by asymmetric information. In contrast, in Panel (1) of both tables, we find a strong, negative relationship between spreads and volume.[39]

---

[35]In additional experiments available upon request, we find that the relationship is even more negative if we exclude from the dataset periods where VIX is extremely high (above its 90% quantile). In particular, the coefficients on RA become more negative and become statistically significant for calls.

[36]See also Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010) and Chen, Joslin, and Ni (2019).

[37]See, for example, `https://www.investopedia.com/trading/options-trading-volume-and-open-interest`

[38]We obtain similar results for calls if we exclude from the dataset periods where VIX is extremely high (above its 90% quantile). These results are available upon request.

[39]We also report additional results of individual regressions of bid-ask spreads on RA, var, oi, etc., in Appendix IA.1.

Table 2: Results for a panel regression of levels of call bid-ask spreads on explanatory variables. RA is the proxy for risk aversion from Bekaert et al. (2021) and kindly provided on the authors website, and $1_{RA>q}$ is the indicator of RA being above its median; $var = 10000 \cdot \sigma^2 \cdot \text{daysToExpiry}/365$, where σ is the rolling 20-day standard deviation of underlying returns, and $1_{var>q}$ is the indicator of var being above its median; oi=open interest/1000; $volume$=volume/1000; and Σ^*, the risk-neutral variance of the call payoff, defined in (27). Spreads are proportional, defined as spread $= 10000 \cdot (\text{Ask(Option)} - \text{Bid(Option)})/\text{Mid(Option)}$.

	call (1)	call (2)	call (3)
RA	-0.08 (0.14)	51.91*** (0.58)	64.81*** (0.51)
$RA \cdot \mathbf{1}_{RA>q}$	35.68*** (0.56)		
var	7.65 (20.58)		
$var \cdot \mathbf{1}_{var>q}$	20.77 (20.48)		
$RA \cdot \Sigma^*$			437.05*** (26.19)
Σ^*			-25074.62*** (173.71)
$IV2$		17.38*** (3.57)	
$RA \cdot IV2$		-3.12*** (0.46)	
oi	-0.65*** (0.06)		
volume	-13.31*** (0.17)		
const	891.71*** (2.05)	786.79*** (1.92)	794.53*** (1.71)
R-squared	0.01	0.00	0.01
R-squared Adj.	0.01	0.00	0.01

Table 3: Results for a panel regression of levels of put bid-ask spreads on explanatory variables. RA is the proxy for risk aversion from Bekaert et al. (2021) and kindly provided on the authors website, and $1_{RA>q}$ is the indicator of RA being above its median; $var = 10000 \cdot \sigma^2 \cdot \text{daysToExpiry}/365$, where σ is the rolling 20-day standard deviation of underlying returns, and $1_{var>q}$ is the indicator of var being above its median; oi=open interest/1000; $volume$=volume/1000; and Σ^*, the risk neutral variance of the option payoff, defined in (27). Spreads are proportional, defined as spread $= 10000 \cdot (\text{Ask(Option)} - \text{Bid(Option)})/\text{Mid(Option)}$.

	put (1)	put (2)	put (3)
RA	-11.84***	-34.75***	-8.54***
	(0.12)	(0.51)	(0.47)
$RA \cdot \mathbf{1}_{RA>q}$	29.68***		
	(0.49)		
var	-111.19***		
	(17.99)		
$var \cdot \mathbf{1}_{var>q}$	134.24***		
	(17.91)		
$RA \cdot \Sigma^*$			1138.80***
			(14.53)
Σ^*			-21563.09***
			(119.89)
$IV2$		134.10***	
		(3.13)	
$RA \cdot IV2$		9.43***	
		(0.40)	
oi	-3.14***		
	(0.05)		
volume	-9.03***		
	(0.17)		
const	1046.65***	942.73***	934.62***
	(1.79)	(1.68)	(1.57)
R-squared	0.01	0.00	0.01
R-squared Adj.	0.01	0.00	0.01

Bid-ask around earnings announcements

Finally, we discuss our last empirical prediction—namely, the fact that, contrary to conventional wisdom, bid-ask spreads may increase after public information release. To test this prediction, we compare option bid-ask spreads on the day before an earnings announcement with the spread on the same option on the day after the earnings announcement. We proceed as follows:

- In our panel dataset (pre-filtered for volume and liquidity, as previously explained), we select all options that exist both before and after the earnings announcement date. For these options, we compute the change in the bid-ask spread,

$$\Delta BA \; = \; BA_{after\ earnings} - BA_{before\ earnings}\,. \tag{30}$$

- We then run a contemporaneous panel regression of spread changes on various controls. The controls are included to make sure the change in the bid-ask spread is not mechanical.

Table 4 reports the results of the regressions, separately for calls and for puts. We introduce three controls in our regression: moneyness change computed over the same time period as ΔBA, number of days to expiry,[40] and change in the implied volatility. Consistent with our key prediction (and completely inconsistent with the CARA-normal model), we see that the constant term is positive and highly significant. Although all three controls in our regressions are significant, their explanatory power is quite low. The change in bid-ask spreads is dominated by the positive constant term—that is, bid-ask spreads increase significantly after earnings announcements.

[40] This control is particularly important. It is known that options with fewer days to expiry are less liquid. Conventional wisdom suggests that this happens because of a fast theta decay of the option. Proposition 4 suggests an alternative explanation: Shorter-term options have a lower physical payoff variance. See also Wei and Zheng (2010).

Table 4: Results for a panel regression of changes in bid-ask spreads around earnings announcements on controls. Moneyness change is the change in moneyness between the day after and the day before earnings, where moneyness of an option is defined as $\log(\text{strike}/\text{underlyingStockPrice})/(\sigma\sqrt{\text{daysToExpiry}/365})$ where σ is the rolling 20-day standard deviation of underlying returns; yte=daysToExpiry/365. IV is Black-Scholes implied volatility of option, provided by ORATS. Spreads are in basis points, defined as spread $= 10000 \cdot (\text{Ask(Option)} - \text{Bid(Option)})/\text{underlyingStockPrice}$.

	calls	puts
const	44.63***	51.74***
	(0.81)	(0.91)
moneyness change	-0.09***	0.22***
	(0.01)	(0.01)
yte change	32.86**	-48.77***
	(13.77)	(15.59)
IV change	8.18***	39.75***
	(1.90)	(2.15)
R-squared	0.00	0.01
R-squared Adj.	0.00	0.01

5 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).[41] Antill and Duffie (2017) and Duffie and Zhu (2017) consider models in which the uniform-price auction market is augmented by price discovery sessions.

All of these papers feature traders with marginal utilities that are linear in trade size

[41] 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.

(which is either assumed directly or follows from the combination of CARA utility and normality of asset payoffs).[42] With the exception of Du and Zhu (2017), they derive linear equilibria with price impact that is constant.[43] As we previously noted, such models cannot speak to empirical evidence for US options that we present. 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 nonlinear equilibria often exist. 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.[44] 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.[45]

A large body of literature exists on *competitive* trading with nonstrategic LPs in setups 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 (2015), Breon-Drish (2015), Pálvölgyi and Venter (2015), and Chabakauri, Yuan, and Zachariadis (2017). Peress

[42]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; we also adopt a cumulant-generating function approach.

[43]Several studies derive models that 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 accounting 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.

[44]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.

[45]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 the 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.

(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 LPs are strategic and demonstrate that this assumption has notable implications for the (cross-)reversals of option returns.

Our paper is also related to the literature on transaction costs and asset prices; see Heaton and Lucas (1996), Vayanos (1998), Vayanos and Vila (1999), Lo, Mamaysky, and Wang (2004), Acharya and Pedersen (2005), and Buss and Dumas (2019). Our study differs from these studies 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, Thum, Hauptmann, and Li 2005; Huberman and Stanzl 2005; Obizhaeva and Wang 2013). Our paper complements this literature by providing equilibrium foundations for nonlinear price functions.

The paper of Liu and Wang (2016) derives implications of earnings announcements for bid-ask spreads, similar to our paper. Assuming that there is less information asymmetry after earnings announcements, the results in Liu and Wang (2016) imply that expected bid–ask spreads may decrease with information asymmetry. This happens because strategic market makers may optimally shift some trades with some investors to other investors by adjusting bid or ask. We believe their findings are complementary to ours. Our effect operates through changes in uncertainty, while their effect operates through changes in information asymmetry.

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 LPs are risk neutral and there is no inventory risk, on which our model focuses.

6 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 an arbitrary distribution of asset payoffs. We show that departing from the common (but unrealistic) assumption of normal payoffs has far-reaching economic implications for asset illiquidity. More specifically, (i) illiquidity may decrease risk aversion, physical variance, and LPs' inventory size; and (ii) it may increase after earnings announcements. These results are consistent with the empirical evidence for US stock options that we present.

We develop a novel constructive approach to solve for the equilibrium in a multi-asset strategic trading model in a 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 of 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
$\nabla f(q),$ where $f:\ \mathbb{R}^N \to \mathbb{R}$	Gradient of f, $(\nabla f)_l = \frac{\partial f}{\partial q_l}$
$\nabla^2 f(q),$ where $f:\ \mathbb{R}^N \to \mathbb{R}$	Hessian of f, $(\nabla^2 f)_{kl} = \frac{\partial^2 f}{\partial q_k \partial q_l}$
$\nabla I(q),$ where $I:\ \mathbb{R}^N \to \mathbb{R}^N$	Jacobian of I, $(\nabla f)_{ik} = \frac{\partial I^i}{\partial q_k}$
$a = \text{ess inf}(h(\delta))$	a is essential infimum of $h(\delta)$. Consider $h_l = \{\hat{a} \in \mathbb{R} : \hat{a} \leq\ h(\delta),\ \text{a.s.}\}$. Then $a = \sup h_l$ if $h_l \neq \emptyset$, and $a = -\infty$ otherwise.
$b = \text{ess sup}(h(\delta))$	b is essential supremum of $h(\delta)$. Consider $h_u = \{\hat{b} \in \mathbb{R} : \hat{b} \geq\ h(\delta),\ \text{a.s.}\}$. Then $b = \inf h_u$ if $h_u \neq \emptyset$, and $b = +\infty$ otherwise.
A_{ij}	ij-th element of a matrix A.
a_i	i-th element of a vector a.
	Model variables

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

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

B Proofs

B.1 Proof of Theorem 1

Proof of Theorem 1. 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

$$\nabla f\left(q^i\right) + \frac{1}{L-1}\nabla I\left(\frac{s-q^i}{L-1}\right)q^i = I\left(\frac{s-q^i}{L-1}\right). \tag{31}$$

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

$$\nabla f(q) + \frac{1}{L-1} \nabla I(q) \, q = I(q) \,. \tag{32}$$

The equilibrium inverse demand $I(q)$ must be a strictly decreasing solution to (32) 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 (32), then $q = s/L$ is the unique solution to FOCs (31). Moreover, $q = s/L$ is a local maximum.*

Proof. Denote

$$\xi = \frac{s - q^i}{L-1} \tag{33}$$

and rewrite (31) as follows:

$$\nabla f\left(q^i\right) + \frac{1}{L-1} \nabla I(\xi) \, q^i = I(\xi) \,. \tag{34}$$

Instead of solving for $q^i(s)$ from (31), we will solve an equivalent system of equations (34) and (33).

Step 1. There is at most one solution to (34).

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

$$\nabla f(q_1) + \frac{1}{L-1} \nabla I(\xi) \, q_1 = I(\xi) \tag{35}$$

$$\nabla f(q_2) + \frac{1}{L-1} \nabla I(\xi) \, q_2 = I(\xi) \,. \tag{36}$$

Multiply (35) and (36) by $(q_2 - q_1)^\top$ and subtract one equation from the other, as follows:

$$(q_2 - q_1)^\top (\nabla f(q_2) - \nabla f(q_1)) + \frac{1}{L-1} (q_2 - q_1)^\top \nabla I(\xi) (q_2 - q_1) = 0. \tag{37}$$

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, $(q_2 - q_1)^\top \nabla I(\xi) (q_2 - q_1)$, 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 (37) is negative and the right-hand side is zero.

Step 2. The only solution to (34) is $q^i = \xi$.

Indeed, $q^i = \xi$ is a solution, since for such q^i, equation (34) becomes equation (32). By the previous step, there is at most one solution. Hence, $q^i = \xi$ is the only solution to (34).

Step 3. The only solution to (31) is $q^i = s/L$.

Indeed, (31) is equivalent to a system of equations (34) and (33). We know that the only solution to (34) is $q = \xi$. Therefore, the system of equations (34) and (33) becomes

$$q^i = \xi, \tag{38}$$

$$\xi = \frac{s - q^i}{L - 1}, \tag{39}$$

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 (31) 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 (32), 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 (32) 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

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

Moreover, for any realization w, we have

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

The result then follows.

∎

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

Proof. Since

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

is equivalent to

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

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

Lemma 5. *The unique solution to (32) such that $I(q)\in\mathcal{A}$ is*

$$I(q)=(L-1)\int_1^\infty t^{-L}\nabla f(tq)dt. \tag{40}$$

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

$$\text{ess inf}(q^\top\delta)<q^\top I(q). \tag{41}$$

Writing (41) for a portfolio tq as well as $-tq$, we also get

$$\text{ess inf}(q^\top\delta) < \iota(t;q) < \text{ess sup}(q^\top\delta), \tag{42}$$

which must hold for any t. According to Proposition 1, finding a solution to (32) amounts to solving linear ODE (12). This solution implies that $I(q) \in \mathcal{A}$ iff $\iota(t;q)$ is such that for any t, and any q (42) holds.

Step 1. Solving ODE (12).

We multiply both sides of (12) 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 (41) implies that $\iota(t;q)$ is bounded, we get a particular solution to (12)

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

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 (12) is given by

$$\iota(x;q) = (L-1)\int_1^\infty \xi^{-L}q^\top\nabla f(x\xi q)d\xi + ct^{L-1}, \tag{43}$$

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

Step 2. The solution (43) 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 $\text{ess inf}(q^\top\delta) < \iota(0;q) = E\left[q^\top\delta\right] < \text{ess sup}(q^\top\delta)$. Therefore, it suffices to prove that

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

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

Step 3. The solution (43) 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, (42) cannot hold.

Step 4. The solution (43) with $c = 0$ implies $I(q) = (L-1)\int_1^\infty \xi^{-L}\nabla f(\xi q)d\xi$.

Indeed, the solution (43) 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 (11) to get (40) ■

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.

$$
\begin{aligned}
& I\left(\frac{s - t_k\theta_k}{L-1}\right)^\top \theta_k && (44)\\
&= \; -L \int_1^\infty z^{-L-1} \nabla f\left(-zc\frac{s - t_k\theta_k}{L-1}\right)^\top \theta_k dz. && (45)\\
& && (46)
\end{aligned}
$$

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,

$$E[e^{t_k\theta_k^\top\delta}] \;=\; E[e^{t_k\theta_k^\top\delta}\mathbf{1}_{A_k}] \;+\; E[e^{t_k\theta_k^\top\delta}(1-\mathbf{1}_{A_k})] \;\le\; E[e^{t_k\theta_k^\top\delta}\mathbf{1}_{A_k}]+e^{t_k(b_k-\epsilon)}\,. \tag{47}$$

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,

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

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

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

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

B.2 Proof of Proposition 1

Proof of Proposition 1.

Step 1. PDE (5) implies ODE (12).

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 (12) 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 (5) 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 (11).

It follows from the 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 (5) and noting that $\frac{1}{L-1}\left(\nabla I\left(q\right)q + I(q)\right) = \nabla e(q)$, we get (11). ■

B.3 Proof of Proposition 2

Proof of Proposition 2. Equilibrium inverse demand is a solution to PDE (5), which is strictly decreasing and such that $I(q) \in \mathcal{A}$. Lemma 5 implies that there is unique such solution, given by (14) or, equivalently, (15). Expressions (16) and (17) are obtained by differentiating (14) and (15). ■

B.4 Proof of Corollary 1

Proof of Corollary 1. The function g satisfies

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

It follows from equation (15) that

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

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

B.5 Proof of Corollary 2

Proof of Corollary 2. It follows from Proposition 2 that for the case of Gaussian distribution,

$I(0) = \mu - \gamma\Sigma x_0$ and $\Lambda(0) = \frac{\gamma}{L-2}\Sigma$, from which the corollary follows. $\blacksquare$

B.6 Proof of Proposition 3

We first formulate a more general version of the proposition.

Proposition. *Suppose that the dividend vector is given by $\hat{\delta} = E[\delta] + \sigma(\delta - E[\delta])$, where σ is a scalar. Suppose that the equilibrium characterized in Proposition 2 exists. The claims that follow correspond to that equilibrium. Then,*

$$BA_k = 2\frac{L-1}{L-2}\gamma\sigma^2\Sigma^*_{kk}\,, \tag{48}$$

and

$$\operatorname{sign}\left(\frac{\partial}{\partial\gamma}BA_k\right) = \operatorname{sign}\left(\operatorname{var}^*(\delta_k) - \gamma\sigma\operatorname{coskew}^*(\delta_k,\delta_k,\delta^\top x_0)\right) \ \textit{and} \tag{49}$$

$$\operatorname{sign}\left(\frac{\partial}{\partial\sigma}BA_k\right) = \operatorname{sign}\left(2\operatorname{var}^*(\delta_k) - \gamma\sigma\operatorname{coskew}^*(\delta_k,\delta_k,\delta^\top x_0)\right). \tag{50}$$

Hence, for small enough γ, we have $\operatorname{sign}\left(\frac{\partial}{\partial\gamma}BA_k\right) = \operatorname{sign}\left(\frac{\partial}{\partial\sigma}BA_k\right) > 0$.

Proof. It follows from Proposition 2 that

$$\mathrm{BA}_k = 2\gamma\sigma^2\frac{L-1}{L-2}g_{kk}(-\gamma\sigma x_0). \tag{51}$$

Here we denote g_{kk} the second derivative of g with respect to its k-th argument. The functions g and f satisfy

$$\begin{aligned}
\frac{\partial^2 f}{\partial y_k^2} &= -\frac{\gamma}{1}\frac{\partial^2 g}{\partial y_k^2}\\
\frac{\partial^2 g}{\partial y_k^2} &= E\left[\delta_k^2\frac{e^{y^\top\delta}}{E[e^{y^\top\delta}]}\right] - \left(E\left[\delta_k\frac{e^{y^\top\delta}}{E[e^{y^\top\delta}]}\right]\right)^2 \implies\\
\left.\frac{\partial^2 g}{\partial y_k^2}\right|_{y=-\gamma\sigma x_0} &= E\left[\delta_k^2\zeta(0;q)\right] - \left(E\left[\delta_k\zeta(0;q)\right]\right)^2.
\end{aligned}$$

Then,

$$\mathrm{BA}_k = 2\frac{\gamma}{1}\frac{L-1}{L-2}\sigma^2\operatorname{var}^*(\delta_k). \tag{52}$$

We turn to deriving expression for $\frac{\partial \mathrm{BA}_k}{\partial\gamma}$. We have

$$\frac{\partial \mathrm{BA}_k}{\partial\gamma} = 2\frac{L-1}{L-2}\sigma^2\left(\operatorname{var}^*(\delta_k) + \gamma\frac{\partial\operatorname{var}^*(\delta_k)}{\partial\gamma}\right).$$

We compute $\frac{\partial \text{var}^*(\delta_k)}{\partial \gamma}$ in Lemma 8 below. Similarly,

$$\frac{\partial \text{BA}_k}{\partial \sigma} = \gamma 2 \frac{L-1}{L-2} \left(2\sigma \text{var}^*(\delta_k) + \sigma^2 \frac{\partial \text{var}^*(\delta_k)}{\partial \sigma} \right).$$

Again, we compute $\frac{\partial \text{var}^*(\delta_k)}{\partial \sigma}$ in Lemma 8 below.

■

Lemma 8. $\frac{1}{\sigma}\frac{\partial \text{var}^*(\delta_k)}{\partial \gamma} = \frac{1}{\gamma}\frac{\partial \Sigma^*_{kk}}{\partial \sigma} = -\text{coskew}^*\left(\delta_k, \delta_k, x_0^\top \delta\right).$

Proof. Denote

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

Note that

$$\frac{\partial}{\partial \gamma}(Z(-\gamma\sigma x_0)) = -\sigma\left(x_0^\top \delta - E[x_0^\top \delta\, Z(-\gamma\sigma x_0))]\right) Z(-\gamma\sigma x_0)).$$

Denote also $q \equiv -\gamma\sigma x_0$. Then we have

$$\begin{aligned}
\frac{1}{\sigma}\frac{\partial \text{var}^*(\delta_k)}{\partial \gamma} &= \frac{1}{\sigma} E[\delta_k^2 \frac{\partial}{\partial \gamma} Z(q)] - 2E\left[\delta_k Z(q)\right] E\left[\delta_k \frac{\partial}{\partial \gamma} Z(q)\right] \\
&= -\left(E[\delta_k^2\left(x_0^\top \delta - E[x_0^\top \delta\, Z(q)]\right) Z(q)] - 2E\left[\delta_k Z(q)\right] E\left[\delta_k\left(x_0^\top \delta - E[x_0^\top \delta\, Z(q)]\right) Z(q)\right]\right) \\
&= -\left(E^*\left[\delta_k^2 x_0^\top \delta\right] - E^*[x_0^\top \delta] E^*\left[\delta_k^2\right] - 2E^*\left[\delta_k\right]\left(E^*\left[\delta_k x_0^\top \delta\right] - E^*\left[x_0^\top \delta\right] E^*\left[\delta_k\right]\right)\right) \\
&= -\text{coskew}^*\left(\delta_k, \delta_k, x_0^\top \delta\right).
\end{aligned}$$

Proceeding similarly, we obtain

$$\frac{1}{\gamma}\frac{\partial \Sigma^*_{kk}}{\partial \sigma} = -\text{coskew}^*\left(\delta_k, \delta_k, x_0^\top \delta\right).$$

■

B.7 Proof of Proposition 4

We first state the more general statement of the proposition, which does not require δ to be bounded on both sides. We prove the proposition in the more general form.

Proposition. *Suppose that the density of δ, $\eta(\delta)$, is strictly positive everywhere on its support. Suppose also that for all i, the distribution of $\delta_i x_{0,i}$ is bounded from below. Suppose that the dividend vector is given by $\hat{\delta} = E[\delta] + \sigma(\delta - E[\delta])$, where σ is a scalar. Suppose that $x_{0,k} \neq 0$. Suppose that the equilibrium characterized in Proposition 2 exists. The claims that follow correspond to that equilibrium. Then, the derivatives $\frac{\partial}{\partial\gamma} BA_k$ and $\frac{\partial}{\partial\sigma} BA_k$ are positive for small enough γ and change sign (at least once) as γ increases. Moreover, we have $BA_k = 4\frac{L-1}{L-2}\gamma^{-1} x_{0,k}^{-2} + O(\gamma^{-1})$. Correspondingly, for large enough γ: the bid-ask spread decreases in risk aversion γ and initial inventory $|x_{0,k}|$. Suppose further that there is a release of public information s_p of the form $s_p = \delta + u$, where $u \sim N(0, \Sigma)$. Denote by* $\text{BA}(\mathcal{F})$ *the bid ask-spreads, given the information*

$\mathcal{F}$. *Then,* $E[\mathrm{BA}(s_p)] - \mathrm{BA}(\emptyset)$ *is positive for small enough* γ*, and changes sign (at least once) as* γ *increases.*

We start with the following useful lemmas.

Lemma 9. *Suppose that* δ *is such that* $\inf_\delta (x_{0,j}\delta_j)$ *is attained for* $\delta_j = 0$ *for all* $j \in \{1, 2, ..N\}$*. Then, as* $\gamma \to \infty$*, we have*

$$\frac{\int \delta_i^m \delta_k^n \exp\left(-\gamma\sigma x_0^\top \delta\right) \eta(\delta) d\delta}{\int \exp\left(-\gamma\sigma x_0^\top \delta\right) \eta(\delta) d\delta} \sim$$
$$(x_{0,k})^{-n} (x_{0,i})^{-m} n!\, m!\, (\gamma\sigma)^{-(m+n)} \left(1 + (\gamma\sigma c_0)^{-1} \left(m c_{1,i} x_{0,i}^{-1} + n c_{1,k} x_{0,k}^{-1}\right) + O\left(\gamma^{-2}\right)\right).$$

Proof. Denote $c_0 = \eta(0)$ and $c_1 = \nabla\eta(0)$. Make a change of variables $y = \gamma\sigma\delta$.
Then we have

$$\frac{\int \delta_i^m \delta_k^n \exp\left(-\gamma\sigma x_0^\top \delta\right) \eta(\delta) d\delta}{\int \exp\left(-\gamma\sigma x_0^\top \delta\right) \eta(\delta) d\delta} =$$

$$\begin{aligned}
&= (\gamma\sigma)^{-(m+n)} \frac{\int y_i^m y_k^n \exp\left(-x_0^\top y\right) \eta(y/(\gamma\sigma)) dy}{\int \exp\left(-x_0^\top y\right) \eta(y/(\gamma\sigma)) dy} \\
&\sim (\gamma\sigma)^{-(m+n)} \left(\frac{\int y_i^m y_k^n \exp\left(-x_0^\top y\right) \left(1 + (\gamma\sigma c_0)^{-1} c_1^\top y\right) dy}{\int \exp\left(-x_0^\top y\right) \left(1 + (\gamma\sigma c_0)^{-1} c_1^\top y\right) dy} + O\left(\gamma^{-2}\right)\right) \\
&\sim (\gamma\sigma)^{-(m+n)} \left(A + (\gamma\sigma c_0)^{-1} (B - C) + O\left(\gamma^{-2}\right)\right).
\end{aligned}$$

Here $A = \frac{\int y_i^m y_k^n \exp\left(-x_0^\top y\right) dy}{\int \exp\left(-x_0^\top y\right) dy}$, $B = \frac{\int y_i^m y_k^n \exp\left(-x_0^\top y\right) c_1^\top y dy}{\left(\int \exp\left(-x_0^\top y\right) dy\right)}$, and $C = \frac{\int \exp\left(-x_0^\top y\right) c_1^\top y dy \int y_i^m y_k^n \exp\left(-x_0^\top y\right) dy}{\left(\int \exp\left(-x_0^\top y\right) dy\right)^2}$.
The first term: Consider

$$\begin{aligned}
A &= \frac{\int y_i^m y_k^n \exp\left(-x_0^\top y\right) dy}{\int \exp\left(-x_0^\top y\right) dy} \\
&\sim \frac{\int y_k^n \exp\left(-x_{0,k} y_k\right) dy_k \int y_i^m \exp\left(-x_{0,i} y_i\right) dy_i \times I_{-(k,i)}}{\int \exp\left(-x_{0,k} y_k\right) dy_k \int \exp\left(-x_{0,i} y_i\right) dy_i \times I_{-(k,i)}} \\
&\sim \frac{\int y_k^n \exp\left(-x_{0,k} y_k\right) dy_k \int y_i^m \exp\left(-x_{0,i} y_i\right) dy_i}{\int \exp\left(-x_{0,k} y_k\right) dy_k \int \exp\left(-x_{0,i} y_i\right) dy_i} \\
&\sim (x_{0,k})^{-n} (x_{0,i})^{-m} n!\, m!\,,
\end{aligned}$$

where $I_{-(k,i)} \equiv \int \exp\left(-x_{-(k,i)}^\top y_{-(k,i)}\right) dy_{-(k,i)}$ and $x_{-(k,i)}$ denotes a vector x_0, excluding its k-th and i-th components.

To see the last transition, consider $\int y_k^n \exp\left(-x_k(t) y_k\right) dy_k$. Make a change of variable $z_k = x_k(t) y_k$. Since $\inf_\delta (x_k(t)\delta_k)$ is attained for $\delta_k = 0$, the support of z_k becomes $\mathbb{R}^+$ as

$\gamma \to \infty$. Then

$$\int y_k^n \exp\left(-x_{0,k} y_k\right) dy_k = x_{0,k}^{-n-1} \int_0^\infty z_k^n \exp(-z_k) dz_k = x_{0,k}^{-n-1} \Gamma(n+1) = x_{0,k}^{-n-1} n!$$

The second term: Analogous to A, we get

$$\begin{aligned}
B &= \frac{\int y_i^m y_k^n \exp\left(-x_0^\top y\right) c_1^\top y dy}{\left(\int \exp\left(-x_0^\top y\right) dy\right)} \\
&= c_{1,i} \frac{\int y_i^{m+1} y_k^n \exp\left(-x_0^\top y\right) c_1^\top y dy}{\left(\int \exp\left(-x_0^\top y\right) dy\right)} + c_{1,k} \frac{\int y_i^m y_k^{n+1} \exp\left(-x_0^\top y\right) c_1^\top y dy}{\left(\int \exp\left(-x_0^\top y\right) dy\right)} + \\
&+ \sum_{j \notin \{i,k\}} c_{1,j} \frac{\int y_i^m y_k^n y_j \exp\left(-x_0^\top y\right) dy}{\left(\int \exp\left(-x_0^\top y\right) dy\right)} \\
&\sim c_{1,i} \left(x_{0,k}\right)^{-n} \left(x_{0,i}\right)^{-m-1} n! \left(m+1\right)! + c_{1,k} \left(x_{0,k}\right)^{-n-1} \left(x_{0,i}\right)^{-m} \left(n+1\right)! \, m! + \\
&+ \sum_{j \notin \{i,k\}} c_{1,j} n! \, m! \left(x_{0,k}\right)^{-n} \left(x_{0,i}\right)^{-m} x_{0,j}^{-1} \\
&= \left(x_{0,k}\right)^{-n} \left(x_{0,i}\right)^{-m} n! \, m! \left(\sum_j c_{1,j} x_{0,j}^{-1} + m c_{1,i} x_{0,i}^{-1} + n c_{1,k} x_{0,k}^{-1}\right).
\end{aligned}$$

The third term: Analogous to A, we get

$$\begin{aligned}
C &= \frac{\int \exp\left(-x_0^\top y\right) c_1^\top y dy \int y_i^m y_k^n \exp\left(-x_0^\top y\right) dy}{\left(\int \exp\left(-x_0^\top y\right) dy\right)^2} \\
&= A \frac{\int \exp\left(-x_0^\top y\right) c_1^\top y dy}{\int \exp\left(-x_0^\top y\right) dy} \\
&\sim \left(x_{0,k}\right)^{-n} \left(x_{0,i}\right)^{-m} n! \, m! \left(\sum_j c_{1,j} x_{0,j}^{-1}\right).
\end{aligned}$$

Thus,

$$B - C \sim \left(x_{0,k}\right)^{-n} \left(x_{0,i}\right)^{-m} n! \, m! \left(m c_{1,i} x_{0,i}^{-1} + n c_{1,k} x_{0,k}^{-1}\right).$$

■

Lemma 10. *Suppose that strategic traders receive a signal*

$$s_p \;=\; \delta + u\,,$$

where u is multivariate normal random variable independent of other random variables. Let $CGF(q, s_p)$ be the CGF conditional on receiving the signal. It is given by

$$CGF(q; s_p) = \; g(q; s_p) \;=\; h_I(q + \Sigma^{-1} s_p) \;+\; h_I(\Sigma^{-1} s_p)\,, \tag{53}$$

where

$$h_I(y) \;=\; \ln E\left[\exp\left(y^T\delta - \frac{1}{2}\delta^T\Sigma^{-1}\delta\right)\right]$$

and Σ is the variance-covariance matrix of u, assumed to be invertible.

Proof of Lemma 10. Since u is a multivariate normal random variable, we have

$$f(s_p|\delta) \;=\; \frac{1}{A}\exp\left(-\frac{1}{2}(\delta - s_p)^T\Sigma^{-1}(\delta - s_p)\right) \quad \text{where} \quad A \;=\; \sqrt{(2\pi)^N\det(\Sigma)}\,.$$

Thus,

$$\begin{aligned} f(\delta|s_p) = \;& \frac{f(s_p|\delta)f(\delta)}{\int f(s_p|\delta)f(\delta)d\delta} \\ =\;& \exp\left(-\frac{1}{2}\delta^T\Sigma^{-1}\delta\right)\frac{\exp\left(-\frac{1}{2}s_p^T\Sigma^{-1}s_p\right)}{\int\exp\left(\frac{1}{2}(\delta - s_p)^T\Sigma^{-1}(\delta - s_p)\right)d\delta}\exp\left(s_p^T\Sigma^{-1}\delta\right) \\ =\;& \xi(\delta)\frac{1}{\exp\left(h_I(\Sigma^{-1}s_p)\right)}\exp\left(s_p^T\Sigma^{-1}\delta\right)\,. \end{aligned}$$

Thus, it follows that

$$\frac{1}{\exp\left(h_I(\Sigma^{-1}s_p)\right)}\int\xi(\delta)\exp\left(s_p^T\Sigma^{-1}\delta\right)d\delta \;=\; \int f(\delta|s_p)d\delta \;=\; 1, \quad \forall s_p\,.$$

Now consider the conditional MGF:

$$\begin{aligned} M =\;& E\left[\exp\left(q^T\delta\right)|s_p\right] \\ =\;& \frac{1}{\exp\left(h_I(\Sigma^{-1}s_p)\right)}\int\xi(\delta)\exp\left(s_p^T\Sigma^{-1}\delta\right)\exp\left(q^T\delta\right)d\delta \\ =\;& \frac{1}{\exp\left(h_I(\Sigma^{-1}s_p)\right)}\int\xi(\delta)\exp\left((s_p + \Sigma q)^T\Sigma^{-1}\delta\right)d\delta \\ =\;& \frac{\exp\left(h_I(\Sigma^{-1}(s_p + \Sigma q))\right)}{\exp\left(h_I(\Sigma^{-1}s_p)\right)}\frac{1}{\exp\left(h_I(\Sigma^{-1}(s_p + \Sigma q))\right)}\int\xi(\delta)\exp\left((s_p + \Sigma q)^T\Sigma^{-1}\delta\right)d\delta \\ =\;& \frac{\exp\left(h_I(\Sigma^{-1}(s_p + \Sigma q))\right)}{\exp\left(h_I(\Sigma^{-1}s_p)\right)}\,. \end{aligned}$$

The result then follows.

■

Proof of Proposition 4.

Without loss of generality, assume that the infimum $\inf_\delta(x_{0,k}\delta_k)$ is attained for $\delta_k = 0$ (we can always shift δ by a constant without changing its risk-neutral variance). For the statements that do not involve the parameter σ, set $\sigma = 1$ in what follows.

As we have shown,

$$\text{BA}_k = 2\frac{\gamma}{1}\frac{L-1}{L-2}\sigma^2\text{var}^*(\delta_k). \tag{54}$$

We have

$$\text{var}^*(\delta_k) = \sigma^2 \frac{\int \delta_k^2 \exp\left(-\gamma\sigma x_0^\top \delta\right)\eta(\delta)d\delta}{\int \exp\left(-\gamma\sigma x_0^\top \delta\right)\eta(\delta)d\delta} - \sigma^2 \left(\frac{\int \delta_k \exp\left(-\gamma\sigma x_0^\top \delta\right)\eta(\delta)d\delta}{\int \exp\left(-\gamma\sigma x_0^\top \delta\right)\eta(\delta)d\delta}\right)^2 .$$

Applying Lemma 9, we get

$$\frac{\int \delta_k^2 \exp\left(-\gamma\sigma x_0^\top \delta\right)\eta(\delta)d\delta}{\int \exp\left(-\gamma\sigma x_0^\top \delta\right)\eta(\delta)d\delta} \sim (x_{0,i})^{-2}\, 2\,(\gamma\sigma)^{-2}\left(1 + (\gamma\sigma c_0)^{-1}\left(2c_{1,k}x_{0,k}^{-1}\right) + O\left(\gamma^2\right)\right)$$

and

$$\left(\frac{\int \delta_k \exp\left(-\gamma\sigma x_0^\top \delta\right)\eta(\delta)d\delta}{\int \exp\left(-\gamma\sigma x_0^\top \delta\right)\eta(\delta)d\delta}\right)^2 \sim (x_{0,i})^{-2}\,(\gamma\sigma)^{-2}\left(1 + 2\,(\gamma\sigma c_0)^{-1}\left(c_{1,k}x_{0,k}^{-1}\right) + O\left(\gamma^2\right)\right).$$

It follows that BA_k decreases in γ and $|x_0|$ and that $\partial^2\text{BA}_k/(\partial\gamma\partial|x_0|) > 0$ for large enough γ. Similarly, if $c_{1,k}x_{0,k}^{-1} > 0$, we have that BA_k decreases in σ for large enough γ.

If $c_{1,k}x_{0,k}^{-1} < 0$, we note that $\partial\text{BA}_k/(\partial\sigma) > 0$ for small γ and then approaches zero from below as $\gamma \to \infty$. Thus, by the Intermediate Value Theorem, $\partial\text{BA}_k/(\partial\sigma)$ has to become negative for some $\gamma \in (0, \infty)$.

Now we turn our attention to the second part of the proposition. Recall that

$$BA_k \;=\; 2(L-1)\Lambda_{kk}(0) \quad \text{and} \quad \Lambda(q) \;=\; \gamma\int_1^\infty \xi^{1-L}\nabla^2 g(-\gamma(x_0+\xi q))d\xi\,.$$

We have

$$\begin{aligned}
\frac{1}{\gamma}\Lambda(q;s_p) =& \int_1^\infty \xi^{1-L}\nabla^2 g(-\gamma(x_0+\xi q);s_p)d\xi \\
=& \int_1^\infty \xi^{1-L}\nabla^2 h_I(-\gamma(x_0+\xi q)+\Sigma^{-1}s_p)d\xi \\
\Longrightarrow \Lambda(0;s_p) =& \frac{\gamma}{(L-2)}\nabla^2 h_I(-\gamma x_0+\Sigma^{-1}s_p)\,.
\end{aligned}$$

Let $\nu(q;s_p) \;=\; \frac{\exp\left(q^T\delta+\left(s_p^T-\frac{1}{2}\delta^T\right)\Sigma^{-1}\delta\right)}{E\left[\exp\left(q^T\delta+\left(s_p^T-\frac{1}{2}\delta^T\right)\Sigma^{-1}\delta\right)\right]}$. We have

$$\begin{aligned}
\nabla^2 h_I(q) =& \; E\left[\delta\delta^T\nu(q;0)\right] - E[\delta^T\nu(q;0)]E\left[\delta\nu(q;0)\right] \\
\Longrightarrow \Lambda(0;s_p) =& \; \frac{\gamma}{(L-2)}\left(E\left[\delta\delta^T\nu(-\gamma x_0;s_p)\right] - E[\delta^T\nu(-\gamma x_0;s_p)]E\left[\delta\nu(-\gamma x_0;s_p)\right]\right) \\
\Lambda_{kk}(0;s_p) =& \; \frac{\gamma}{(L-2)}\left(E\left[\delta_k^2\nu(-\gamma x_0;s_p)\right] - E^2[\delta_k\nu(-\gamma x_0;s_p)]\right)\,.
\end{aligned}$$

Thus,

$$E[\text{BA}_k(s_p)] - \text{BA}_k(\emptyset) \;=\; 2(L-1)\left(\Lambda_{kk}(0;s_p) - \Lambda_{kk}(0;0)\right) \;=\; 2(L-1)\frac{\gamma}{(L-2)}\left(Var^{**}[\delta] - Var^*[\delta]\right)\,,$$

where * and ** indicates that expectations are taken under the change of measure $\nu(q; s_p)$ and $\nu(q; 0)$, respectively.

Consider

$$E\left[\delta_k^2 \nu(-\gamma x_0; s_p)\right] .$$

Make a change of variables $y = \gamma x_{0k}\delta$. We have

$$\begin{aligned}
E\left[\delta_k^2 \nu(-\gamma x_0; s_p)\right] &= \frac{\int \delta_k^2 \exp\left(-\gamma x_0^T \delta + \left(s_p^T - \frac{1}{2}\delta^T\right)\Sigma^{-1}\delta\right)\eta(\delta)d\delta}{\int \exp\left(-\gamma x_0^T \delta + \left(s_p^T - \frac{1}{2}\delta^T\right)\Sigma^{-1}\delta\right)\eta(\delta)d\delta} \\
&= \frac{1}{\gamma^2 x_{0k}^2}\frac{\int y_k^2 \exp\left(-x_0^T y/x_{0k} + \left(s_p^T - \frac{1}{2}y^T/\gamma x_{0k}\right)\Sigma^{-1}y/\gamma x_{0k}\right)\eta(y/\gamma x_{0k})dy}{\int \exp\left(-\gamma x_0^T y/\gamma x_{0k} + \left(s_p^T - \frac{1}{2}y^T/\gamma x_{0k}\right)\Sigma^{-1}\delta\right)\eta(y/\gamma x_{0k})dy} \\
&= \frac{1}{\gamma^2 x_{0k}^2}\frac{\int y_k^2 \exp\left(-x_0^T y/x_{0k}\right)\exp\left[\left(s_p^T - \frac{1}{2}y^T/\gamma x_{0k}\right)\Sigma^{-1}y/\gamma x_{0k}\right]\eta(y/\gamma x_{0k})dy}{\int \exp\left(-\gamma x_0^T y/\gamma x_{0k} + \left(s_p^T - \frac{1}{2}y^T/\gamma x_{0k}\right)\Sigma^{-1}\delta\right)\eta(y/\gamma x_{0k})dy} \\
&\approx \frac{1}{\gamma^2 x_{0k}^2}\frac{\int y_k^2 \exp\left(-x_0^T y/x_{0k}\right)\left[1 + s_p^T\Sigma^{-1}y/\gamma x_{0k}\right]\left(c_0 + c_1^T y/\gamma x_{0k}\right)dy}{\int \exp\left(-x_0^T y/x_{0k}\right)\left[1 + s_p^T\Sigma^{-1}y/\gamma x_{0k}\right]\left(c_0 + c_1^T y/\gamma x_{0k}\right)dy} \\
&\approx \frac{1}{\gamma^2 x_{0k}^2}\frac{\int y_k^2 \exp\left(-x_0^T y/x_{0k}\right)\left[c_0 + \left(c_0 s_p^T\Sigma^{-1} + c_1^T\right)y/\gamma x_{0k}\right]dy}{\int \exp\left(-x_0^T y/x_{0k}\right)\left[c_0 + \left(c_0 s_p^T\Sigma^{-1} + c_1^T\right)y/\gamma x_{0k}\right]dy} \\
&= \frac{1}{\gamma^2 x_{0k}^2}\frac{A_k(2) + B_k(2)\frac{1}{\gamma x_{0k}}}{A_k(0) + B_k(0)\frac{1}{\gamma x_{0k}}} \\
&\approx \frac{1}{\gamma^2 x_{0k}^2}\left[\frac{A_k(2)}{A_k(0)} + \frac{A_k(0)B_k(2) - A_k(2)B_k(0)}{A_k(0)^2}\frac{1}{\gamma x_{0k}} + O(1/\gamma^2)\right]
\end{aligned}$$

where

$$\begin{aligned}
A_k(n) &= c_0 \int y_k^n \exp\left(-x_0^T y/x_{0k}\right)dy \\
B_k(n) &= \left(c_0 s_p^T \Sigma^{-1} + c_1^T\right)\int y_k^n \exp\left(-x_0^T y/x_{0k}\right)y dy\,.
\end{aligned}$$

Similarly,

$$E\left[\delta_k \nu(-\gamma x_0; s_p)\right] \approx \frac{1}{\gamma x_{0k}}\left[\frac{A_k(1)}{A_k(0)} + \frac{A_k(0)B_k(1) - A_k(1)B_k(0)}{A_k(0)^2}\frac{1}{\gamma x_{0k}} + O(1/\gamma^2)\right]$$

It follows that

$$\Lambda_{kk}(0; s_p) = \frac{1}{\gamma^2 x_{0k}^2}\left[A_{\Lambda k} + B_{\Lambda k}\frac{1}{\gamma x_{0k}} + O(1/\gamma^2)\right],$$

where

$$
\begin{aligned}
A_{\Lambda k} = & \ \frac{A_k(2)}{A_k(0)} - \left(\frac{A_k(1)}{A_k(0)}\right)^2 \\
B_{\Lambda k} = & \ \frac{A_k(0)B_k(2) - A_k(2)B_k(0)}{A_k(0)^2} - 2\frac{A_k(1)}{A_k(0)}\frac{A_k(0)B_k(1) - A_k(1)B_k(0)}{A_k(0)^2}\,.
\end{aligned}
$$

■

Internet Appendix for
"Illiquidity and Higher Cumulants"

IA.1 Additional Regression Results

IA.1.1 Additional regressions of call spreads

Table 6: Results for a panel regression of levels of call bid-ask spreads on explanatory variables. RA is the proxy for risk aversion from Bekaert et al. (2021) and kindly provided on the authors website, and $1_{RA>q}$ is the indicator of RA being above its median; $var = 10000 \cdot \sigma^2 \cdot \text{daysToExpiry}/365$, where σ is the rolling 20-day standard deviation of underlying returns, and $1_{var>q}$ is the indicator of var being above its median; oi=open interest/1000; $volume$=volume/1000; and Σ^*, the risk neutral variance of the call payoff, defined in (27). Spreads are proportional, defined as spread $= 10000 \cdot (\text{Ask(Option)} - \text{Bid(Option)})/\text{Mid(Option)}$.

	call (0)	call (1)	call (2)	call (3)	call (4)
RA	49.68***	8.43***			
	(0.45)	(0.83)			
$RA \cdot \mathbf{1}_{RA>q}$		32.38***			
		(0.54)			
var				43.55***	-191.20***
				(0.64)	(20.42)
$var \cdot \mathbf{1}_{var>q}$					233.54***
					(20.31)
oi			-2.58***		
			(0.05)		
const	794.36***	864.50***	956.45***	935.27***	941.58***
	(1.56)	(1.95)	(0.70)	(0.71)	(0.90)
R-squared	0.00	0.00	0.00	0.00	0.00
R-squared Adj.	0.00	0.00	0.00	0.00	0.00

IA.1.2 Additional regressions of put spreads

Table 7: Results for a panel regression of levels of put bid-ask spreads on explanatory variables. RA is the proxy for risk aversion from Bekaert et al. (2021) and kindly provided on the authors website, and $1_{RA>q}$ is the indicator of RA being above its median; $var = 10000 \cdot \sigma^2 \cdot \text{daysToExpiry}/365$, where σ is the rolling 20-day standard deviation of underlying returns, and $1_{var>q}$ is the indicator of var being above its median; oi=open interest/1000; $volume$=volume/1000; and Σ^*, the risk neutral variance of the option payoff, defined in (27). Spreads are proportional, defined as spread $= 10000 \cdot (\text{Ask(Option)} - \text{Bid(Option)})/\text{Mid(Option)}$.

	put (0)	put (1)	put (2)	put (3)	put (4)
RA	-13.84*** (0.40)	6.28*** (0.72)			
$RA \cdot \mathbf{1}_{RA>q}$		-15.79*** (0.48)			
var				11.40*** (0.56)	154.50*** (17.86)
$var \cdot \mathbf{1}_{var>q}$					-142.36*** (17.76)
oi			-4.92*** (0.04)		
const	904.86*** (1.36)	870.65*** (1.71)	876.35*** (0.61)	858.52*** (0.62)	854.67*** (0.78)
R-squared	0.00	0.00	0.00	0.00	0.00
R-squared Adj.	0.00	0.00	0.00	0.00	0.00

IA.1.3 Regressions of call spreads with VIX as a proxy of risk aversion

Table 8: Results for a panel regression of levels of call bid-ask spreads on explanatory variables: VIX is the CBOE VIX level $\cdot$ 1000, and $1_{VIX>q}$ is the indicator of VIX being above its median; $var = 10000 \cdot \sigma^2 \cdot$ daysToExpiry/365, where σ is the rolling 20-day standard deviation of underlying returns, and $1_{var>q}$ is the indicator of var being above its median; oi=open interest/1000; $volume$=volume/1000; and Σ^*, the risk neutral variance of the call payoff, defined in (27). Spreads are proportional, defined as spread $= 10000 \cdot (\text{Ask(Option)} - \text{Bid(Option)})/\text{Mid(Option)}$.

	call (0)	call (1)	call (2)	call (3)	call (4)	call (5)	call (6)	call (7)
VIX	7.67***	10.32***				9.21***	6.86***	9.09***
	(0.07)	(0.19)				(0.19)	(0.08)	(0.08)
$VIX \cdot \mathbf{1}_{VIX>q}$		-1.89***				-1.48***		
		(0.13)				(0.13)		
var				43.53***	-190.14***	48.81**		
				(0.64)	(20.37)	(20.53)		
$var \cdot \mathbf{1}_{var>q}$					232.46***	-21.73		
					(20.26)	(20.43)		
$VIX \cdot \Sigma^*$								550.77***
								(6.59)
Σ^*								-42607.25***
								(253.22)
IV2							-149.11***	
							(5.05)	
VIX*IV2							3.49***	
							(0.11)	
oi			-2.59***			-0.56***		
			(0.05)			(0.06)		
volume						-13.28***		
						(0.17)		
const	803.68***	776.97***	957.27***	936.10***	942.39***	797.35***	827.72***	834.08***
	(1.49)	(2.35)	(0.70)	(0.71)	(0.89)	(2.47)	(1.72)	(1.60)
R-squared	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
R-squared Adj.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01

IA.1.4 Regressions of put spreads with VIX as a proxy of risk aversion

Table 9: Results for a panel regression of levels of put bid-ask spreads of explanatory variables: VIX is the CBOE VIX level $\cdot$ 1000, and $1_{VIX>q}$ is the indicator of VIX being above its median; $var = 10000 \cdot \sigma^2 \cdot$ daysToExpiry/365, where σ is the rolling 20-day standard deviation of underlying returns, and $1_{var>q}$ is the indicator of var being above its median; oi=open interest/1000; $volume$=volume/1000; and Σ^*, the risk neutral variance of the put payoff, defined in (27). Spreads are proportional, defined as spread $= 10000 \cdot (\text{Ask(Option)} - \text{Bid(Option)})/\text{Mid(Option)}$.

	put (0)	put (1)	put (2)	put (3)	put (4)	put (5)	put (6)	put (7)
VIX	-5.58***	5.11***				4.06***	-9.50***	-5.77***
	(0.06)	(0.17)				(0.17)	(0.07)	(0.07)
$VIX \cdot \mathbf{1}_{VIX>q}$		-7.61***				-7.06***		
		(0.11)				(0.11)		
var				11.40***	148.70***	-104.77***		
				(0.56)	(17.84)	(17.96)		
$var \cdot \mathbf{1}_{var>q}$					-136.59***	125.41***		
					(17.74)	(17.88)		
$VIX * \Sigma^*$								636.49***
								(4.21)
Σ^*								-38044.13***
								(182.91)
IV2							166.33***	
							(4.42)	
$VIX \cdot IV2$							2.20***	
							(0.10)	
oi			-4.92***			-3.01***		
			(0.04)			(0.05)		
volume						-8.96***		
						(0.17)		
const	968.81***	860.98***	877.36***	859.52***	855.83***	886.87***	1008.16***	1029.95***
	(1.31)	(2.06)	(0.61)	(0.62)	(0.78)	(2.16)	(1.50)	(1.43)
R-squared	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
R-squared Adj.	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01

IA.2 Computing Risk-Neutral Variance of Options Payoffs

For call and put options with strike K, the risk-neutral variances are given by

$$\begin{aligned} \text{var}^*[(S-K)^+] &= E^*[((S-K)^+)^2] - (E^*[(S-K)^+])^2 \\ \text{var}^*[(K-S)^+] &= E^*[((K-S)^+)^2] - (E^*[(K-S)^+])^2 \,. \end{aligned} \tag{IA.1}$$

Thus, to compute the risk-neutral variance one would need to observe prices of derivatives with payoffs $((S-K)^+)^2$ and $((K-S)^+)^2$, respectively. While such derivatives are not directly available for trading, their prices can be constructed synthetically using the Carr and Madan (1998) payoff decomposition formula.

Lemma IA.1 (Carr and Madan (1998) payoff decomposition formula)**.** *For any derivative with a payoff $f(S)$, we have*

$$f(S) \;=\; f(S_0) + f'(S_0)((S-S_0)^+ - (S_0-S)^+) + \int_{S_0}^{\infty} f''(k)(S-k)^+ dk + \int_0^{S_0} f''(k)(k-S)^+ dk. \tag{IA.2}$$

In the above formula S_0 is an arbitrary "split level," commonly chosen to be equal to the current price of the underlying asset. Let $C(k)$ and $P(k)$ denote the prices of call and put options with strike k, respectively. Then, (IA.2) implies the expressions for risk-neutral second moments of option payoffs summarized in the following proposition.

Proposition IA.1. *The risk-neutral variances of call and put payoffs are given by*

$$\operatorname{var}^*[(S-K)^+] = \begin{cases} (S_0-K)^2 + 2(S_0-K)\left(C(S_0)-P(S_0)\right) + 2\int_K^{S_0} P(k)dk + 2\int_{S_0}^{\infty} C(k)dk - C(K)^2, & K < S_0; \\ 2\int_K^{\infty} C(k)dk - C(K)^2, & K > S_0; \end{cases}$$

for calls and

$$\operatorname{var}^*[(K-S)^+] = \begin{cases} (K-S_0)^2 - 2(K-S_0)\left(C(S_0)-P(S_0)\right) + 2\int_0^{S_0} P(k)dk + 2\int_{S_0}^{K} C(k)dk - P(K)^2, & K > S_0 \\ 2\int_0^{K} P(k)dk - P(K)^2, & K < S_0 \end{cases}$$

for puts.

Using these expressions, we can now compute the required risk-neutral variances using formulas (27) and then test our model's predictions. We proceed as follows:

- We approximate the integrals above with the corresponding Riemann sums using available strikes. For example,

$$2\int_0^K P(k)dk \;\approx\; \sum_i 2P(K_i)(K_i - K_{i-1}),$$

  where K_i are available strikes.

- For the prices of options in the formula above we use mid-prices.

- Due to strike discreteness and large bid-ask spreads, the risk-neutral variances (27) sometimes happen to be negative. We keep only combinations (strike, expiration date) for which risk neutral variances of both put and call options are strictly positive.

- Since BA spreads are normalized by the stock price level, we normalize risk-neutral variance by the stock price squared.[46]

- We choose the split level S_0 to be equal to the current price of underlying.

[46] The reason is that the variance scales quadratically with the price.

IA.3 Contrasting to Competitive Benchmark

So far we have contrasted the results in our model to those in a Gaussian benchmark, thereby highlighting the role of higher cumulants. In this section, we highlight the role of market power by contrasting to a competitive equilibrium benchmark, where LPs take prices as given. We highlight four shortcomings of the competitive model: (i) the options prices are not affected by distribution of inventory among different traders; in particular, prices are unaffected by open interest (ii) the bid-ask spreads of options are not affected by the number L of LPs, (iii) the prices of options are not affected by LPs' inventories and by higher moments of options payoffs; and (iv) there are no (cross-)reversals. All these are testable predictions that can help distinguish a competitive model from the strategic one.

Formally, the competitive equilibrium is defined as follows.

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

We revisit the results of Proposition 2 in the competitive economy.

Proposition IA.2. *There exists a unique equilibrium in the competitive economy. The equilibrium inverse demand $I(q;x_0)$ is given by*

$$I(q;x_0) = I(q+x_0) = \nabla g(-\gamma(x_0+q)). \tag{IA.3}$$

The implication of the above proposition is that equilibrium quantities in the competitive economy are completely determined by LPs' final inventory x_0+q. Thus, for any combination of initial inventory x_0 and the size of the supply shock q, such that $x_0+q = \textit{const}$, the equilibrium quantities shall be the same. This is in contrast to the strategic economy; cf. Proposition 2.

IA.3.1 Option prices and open interest

Next, we show that in a frictionless model with competitive traders, the distribution of asset holdings across trader types is irrelevant: Only the aggregate supply (market portfolio) matters for equilibrium prices. Hence, since derivatives are in zero net supply, open interest will not matter in equilibrium, in contrast to empirical evidence.

We show the irrelevance of inventory distribution in the competitive economy in a general setting. The argument below is standard and only presented for the sake of completeness. We consider an economy with heterogenous LPs, $i \in \{1,..L\}$, who choose demand schedules $q^i(p)$ to maximize$f^i(x_0^i+q^i) - p^\top q^i$, taking prices as given. In addition, there are heterogenous LDs, $j \in \{1,..M\}$, who choose market orders q_j to maximize$f^j(x_0^j+q^j) - p^\top q^j$, taking prices as given. The total supply of assets is given by a vector Y.[47] We call the economy just described *heterogenous competitive economy.*

Proposition IA.3. *The prices in the heterogenous competitive economy are given by $\nabla U(Y)$, where Y is the aggregate supply of the securities and $U(x)$ is a certainty equivalent of the*

[47] The total supply might be uncertain. However, this is not necessary as the problem of multiplicity of equilibria (that required supply uncertainty in the strategic setting) does not arise in the competitive setting.

representative agent, determined as follows:

$$U(x) = \max_{x^i, x^j} \left\{ \sum_{i=1}^{N} f^i(x^i) + \sum_{j=1}^{M} f^j(x^j) \right\} \ s.t.\!: \ x = \sum_{i=1}^{N} x^i + \sum_{j=1}^{M} x^j.$$

Consequently, prices are unaffected by the distribution of asset holdings across traders x_0^i, $i \in \{1,..L\}$, and x_0^j, $j \in \{1,..M\}$.

The proposition above is intuitive. In the competitive economy, all traders can be aggregated and substituted with a representative trader holding the whole asset supply. Hence, the prices are only affected by total supply, and not by the inventory distribution across traders.

IA.3.2 Bid-ask spreads and competition among LPs

Next, we show that in the competitive equilibrium the bid-ask spread is unaffected by competition among LPs (measured by L) unlike in the strategic equilibrium.

Proposition IA.4. *The bid-ask spread can be written as*

- $BA_k = 2\frac{L-1}{L-2}\gamma\Sigma^*_{kk}$, *in the strategic equilibrium;*
- $BA_k = 2\gamma\Sigma^*_{kk}$, *in the competitive equilibrium.*

Moreover, BA does not depend on L in the competitive equilibrium and decreases in L in the strategic equilibrium.

The proposition above is intuitive. When L is larger in the strategic economy, the competition among LPs is fiercer, and the bid-ask spreads go down. In the competitive equilibrium, there is always perfect competition among LPs, so bid-ask spreads are unaffected by L.

IA.3.3 Derivatives in zero net supply

As we argue in the main body of the paper, our model is best suited for describing derivatives markets where higher-order moments and nonlinearities play a key role. Importantly, derivatives are always in zero net supply. Hence, when LPs accumulate an inventory of x_0, it means that LDs accumulated an inventory of $-Lx_0$. Hence, it is natural to assume that, on average, the supply of LDs satisfies $s \approx -Lx_0$. Motivated by this intuition, in the next result we assume that $s/L = -x_0 + \epsilon$, where ϵ is small and full-support uncertain.[48]

Proposition IA.5. *Suppose that $s/L = -x_0 + \epsilon$, where $\epsilon \sim N(0, \mathrm{diag}(\sigma_\epsilon))$. Then, in the limit as $\sigma_\epsilon \to 0$ we have*

- *Price of asset i in the competitive equilibrium is given by $E[\delta_i]$,*

[48] On any given day, average future inventory that LPs will hold during the whole life span of the derivative (until its expiration) exactly equals average future demand. Since most LDs close their positions on expiration date (to avoid transaction costs associated with physical settlement), the average demand of LDs is equal to minus their current inventory. Hence, $s/L = -x_0 + \epsilon$. Technically, the full-support ϵ will help to make sure that there is full-support uncertainty about the supply shock s.

- *Price of asset i in the strategic equilibrium are given by* $(L-1)\int_1^\infty \xi^{-L}\nabla_i g(-\gamma x_0(1-\xi))\,d\xi \;=\; \mathrm{E}[\delta_i] + \frac{1}{L-2}\gamma\mathrm{cov}(\delta_i, x_0^\top\delta) \;+\; \frac{1}{2}\frac{2}{(L-2)(L-3)}\gamma^2 coskew(\delta_i, x_0^\top\delta, x_0^\top\delta) + O\left(\gamma^3\right).$

Consequently, in the competitive economy, neither higher cumulants nor initial inventory x_0 affects prices, unlike in the strategic model.

IA.3.4 (Cross-)reversals

Another shortcoming of a competitive model is its inability to generate reversals. We first show that the strategic model does generate reversals. We will demostrate below that the demand reduction $\Lambda(q)q$ in (6) is associated with price reversals. Rostek and Weretka (2015a) show that this is the case in a dynamic CARA-normal model. Below we formulate a simple extension of our model, allowing us to verify such a result in a setting without normality.

To speak to price reversals, we need to define the prices after $t=0$. To do it, we add an additional trading period $t=1/2$. There is a supply shock s at $t=0$ (due to LDs' orders) and no such shock at $t=1/2$. As before, the LPs consume only at time $t=1$. We are looking for a symmetric subgame perfect Nash equilibrium in demand functions. We impose the same restrictions on the the demand functions as in Definition 2.

Given the supply shock of size s, denote equilibrium prices at $t=0$ and $t=1/2$, as $P_0(s)$ and $P_{1/2}(s)$. Define the immediate price reaction to the supply shock is $P_0(s)-P_0(0)$: the difference between the price with vs. without such shock. Part of such reaction may be reversed at $t=1/2$. We define reversal as the difference between the total price change $P_{1/2}(s)-P(0)$ and the immediate price reaction,

$$\mathrm{rev}(s) \equiv P_{1/2}(s) - P(0) - (P_0(s) - P_0(0)) = P_{1/2}(s) - P_0(s). \tag{IA.4}$$

Note that rev is a vector variable, tracking the changes in prices of all assets, caused by a supply shock s. Thus, our model implies a relationship between a price of asset i and a subsequent price of another asset j, which we call cross-reversals.

The following theorem provides an equilibrium in the extended model and gives the closed-form expression for cross-reversals.

Theorem IA.1 (Equilibrium in the extended model)**.** *There exists essentially unique equilibrium in the extended model.*[49] *Equilibrium inverse demand and price impact matrix at time 0, $I(q)$, and $\Lambda(q)$, are the same as in the baseline model and are given by* (15) *and* (17)*. Equilibrium price at $t=0$ is given by $P_0(s)=I(s/L)$. There is no trade at $t=1/2$, and $P_{1/2}=\nabla f(x_0+s/L)$. The price reversal is given by* $\mathrm{rev}(s)=\Lambda(s/L)s/L$.

The theorem above is intuitive. Consider time $t=1/2$. In the symmetric equilibrium, all LPs start with the same inventory and there is no supply shock at $t=1/2$. Hence, there will be no trade. The price at which LPs are indifferent between buying and selling (i.e., happy not to trade) is equal to their marginal certainty equivalent. Thus, $P_{1/2}=\nabla f(x_0+s/L)$. Now step back to $t=0$. The key step of the proof of the theorem shows that the traders do not have incentives to deviate to a $t=1/2$ subgame where their inventories are asymmetric (e.g.,

[49] Essentially unique means that there are multiple equilibrium demand functions implementing no-trade equilibrium at $t=1/2$. For all of them, prices $P_{1/2}$ are the same.

when a trader of interest adsorbs more than s/L while others adsorb less). In that case, LPs' tradeoff at $t = 0$ is unchanged since the continuation game after $t = 0$ does not yield any gains or losses (there is no trade at $t = 1/2$). Thus, $I(q)$ and $\Lambda(q)$ at $t = 0$ are the same as in the baseline model.

Some further insights can be obtained when we consider the supply shock that is small.

Proposition IA.6. *Suppose there is a supply shock $k \cdot s$, where k is a scalar. The immediate price reaction to this shock is given by $P_{0,j}(s) - P_{0,j}(0) = -\gamma k \frac{L-1}{L(L-2)} \mathrm{cov}^* \left(\delta^\top s, \delta_j\right) + O(k^2)$. The reversal is given by $P_{1/2,j}(s) - P_{0,j}(s) = \gamma k \frac{1}{L(l-2)} \mathrm{cov}^* \left(\delta^\top s, \delta_j\right) + O(k^2)$.*

The above proposition implies that supply shocks, even if not affecting the asset of interest, would contribute to negative auto-correlation in the returns of that asset. Indeed, the direction of price changes $P_{0,j}(s) - P_{0,j}(0)$ and $P_{1/2,j}(s) - P_{0,j}(s)$ are the opposites, which is true even if supply shock is zero for asset j. Our model also implies return correlations across assets, which we call cross-reversals. Indeed, if $\mathrm{cov}^* \left(\delta^\top s, \delta_i\right)$ and $\mathrm{cov}^* \left(\delta^\top s, \delta_j\right)$ are of the same (resp., opposite) signs, the price changes $P_{0,i}(s) - P_{0,i}(0)$ and $P_{1/2,j}(s) - P_{0,j}(s)$ are negatively (resp., positively) related.

We now show that the competitive model does not generate (cross-)reversals.

Proposition IA.7. *Consider an extended model of this section and suppose that LPs there take prices as given. In such a model we have $P_0(s) = P_{1/2}(s) = \nabla f(x_0 + s/L)$. Consequently, there are no price reversals.*

IA.4 A Model with Multiple Maturities

To be able to speak to empirical findings about options, we consider a simple modification of the model that allows for multiple maturities. There are N_t securities, which pay off only once, at time $t \in \mathcal{T} \equiv \{t_1, t_2, ..t_T\}$. Denote the total number of securities $N = \sum_{t \in \mathcal{T}} N_t$. The payoffs of all securities (maturity-t securities) are collected in a vector $\delta \in R^N$ (resp., $\delta_t \in R^{N_t}$). Traders can consume at times $t \in \mathcal{T}$ and have CARA utility with risk-aversion γ and time preference β. They can trade only at $t = 0$. Denote by $D^i(p) \in R^N$ (resp., $D^i(p,t) \in R^{N_t}$) the demand vector of trader i for all (resp., maturity-t securities). Analogously, denote by $p(D^i(p), D(p)) \in R^N$ (resp., $p_t(D^i(p), D(p)) \in R^{N_t}$) the vector of equilibrium prices of all securities (resp., maturity-t securities) given that trader i's demand is $D^i(p)$ and demands of all other traders are $D(p)$. Denote $s \in R^N$ the supply vector of all securities. Traders are endowed with $x_{0,t} \in R^{N_t}$, $t \in \mathcal{T}$ of maturity-t securities. The optimization problem for trader i can be written as

$$
\begin{gathered}
\max_{D^i(p,t), t \in \mathcal{T}} \sum_{t \in \mathcal{T}} E[-\exp(-\gamma c_t - \beta t)], \\
\text{s.t.} \quad c_t = \delta_t^\top (D_t^i(p) + x_{0,t}) - p_t(D^i(p), D(p))^\top D_t^i(p) \ \text{, and} \\
p(D^i(p), D(p)) : D^i(p) + (L-1) D(p) = s.
\end{gathered}
\tag{IA.5}
$$

As in the baseline model of Section 2, we focus on arbitrage-free symmetric Nash equilibria. Additionally, we look for equilibria where demands of maturity-t securities do not depend on prices of securities with different maturities $D(p,t) = D(p_t,t)$.[50]

The proposition below shows that the results from the baseline model continue to hold, for each separate maturity.

Proposition IA.8 (Closed-form solution, multiple maturities)**.** *There exists a unique equilibrium. For all $t \in \mathcal{T}$, the equilibrium inverse demand $I_t(q)$ and the price impact matrix $\Lambda_t(q)$ for securities with maturity t are given by, respectively:*

$$I(q,t) = (L-1)\int_1^\infty \xi^{-L}\nabla f_t(\xi q)\,d\xi \tag{IA.6}$$

$$= (L-1)\int_1^\infty \xi^{-L}\nabla g_t(-\gamma(x_{0,t}+\xi q))\,d\xi; \tag{IA.7}$$

$$\Lambda(q,t) = -\int_1^\infty \xi^{1-L}\nabla^2 f_t(\xi q)\,d\xi \tag{IA.8}$$

$$= \gamma\int_1^\infty \xi^{1-L}\nabla^2 g_t(-\gamma(x_{0,t}+\xi q))\,d\xi, \tag{IA.9}$$

where $g_t(\cdot)$ and $f_t(\cdot)$ denote, respectively, CGF and certainty equivalent of payoffs of securities maturing at t, $g_t(y) = \log E[exp(y^\top \delta_t)]$ and $f_t(q_t, x_{0,t}) = -\frac{1}{\gamma}g(-\gamma(x_{0,t}+q_t)))$.

IA.5 A Model with Price-elastic Supply

In this section, we consider an economy as in the main part of the paper but with (i) one risky asset and (ii) price-elastic supply $s(p;u) = u + S(p)$. We assume that $0 < S'(p) < \infty$: The supply is upward-sloping and not perfectly elastic. We assume that u is full-support uncertain. When $S(p) = 0$, the model reduces to the one studied in the main part of the paper.

A key result from the main part of the paper is that for large enough γ, the bid-ask spread is decreasing in γ. One could conjecture that this result may not hold true when the supply is price-elastic: In that case, the bid ask-spread will also be affected by (exogenous) elasticity of supply. We show below that the logic just described is not true. In the economy described above the bid-ask spread still decreases as γ gets large. We do all analysis numerically, as the setting considered here is much less tractable. However, the main takeaway appears to be extremely robust to various modifications of parameters. The intuition for why our main result still holds in this setting is as follows. In the extended model, the liquidity in the market comes from two sources: the price-elastic supply and the price-elastic demands of LPs. As long as the interaction between the liquidity in the supply $s(p,u)$ and γ is not the opposite of that between the liquidity in LPs' demand and γ (which is true here), our results still hold.

We start by deriving ODE characterizing equilibrium demands. Suppose that the equilibrium demand is $q(p)$ and trader i's demand is $q_i(p)$. The market-clearing can be written as

$$u = q_i(p) - S(p)/L + (L-1)(q(p) - S(p)/L).$$

[50]Note that this is not a restriction on the strategy space. The traders can condition on prices of other securities, but, in equilibrium, will not choose to do so. An alternative setting where markets for securities with different maturities are cleared separately will yield the same equilibrium as given in Proposition IA.8.

Denote

$$d_i(p) = q_i(p) - S(p)/L, \text{and } d(p) = q(p) - S(p)/L,$$

the "excess demands" of LPs. Denote $I(d)$ the inverse of $d(p)$.

Given the realization of inelastic demand u, the ex-post optimization problem can be written as

$$\max_{d_i} f(d_i + S(p(d_i; u))/L) - p(d_i, u)\,(d_i + S(p(d_i; u))/L)$$

$$s.t. : p(d_i; u) = I\left(\frac{u - d_i}{L - 1}\right).$$

For the realization $u = Ld$, the first order condition should hold for $d_i = d$:

$$(f'(d + S(I(d))/L) - I(d))\left(1 + \frac{1}{L} S'(I(d))\lambda(d)\right) = \lambda(d)\,(d + S(I(d))/L)\,, \tag{IA.10}$$

where

$$\lambda(d) = \frac{-I'(d)}{L - 1}. \tag{IA.11}$$

Thus, we arrive at the following proposition.

Proposition IA.9. *Suppose that the equilibrium exists. Then, the inverse demand $I(d)$ satisfies the ODE (IA.10) and (IA.11).*

We define the bid-ask spread analogously to the main paper, as the difference between price when $u = -\epsilon$ and when $u = \epsilon$, normalized by ϵ, where ϵ is small. It is straightforward to show that bid-ask spread is given by

$$\text{BA} = const \cdot \lambda(0),$$

where the constant does not depend on γ.

We then proceed numerically. We solve the ODE (IA.10) and (IA.11) with boundary condition $\lambda(\infty) = 0$, and then plot BA for different levels of γ. Figure IA.1 represents a typical picture. The dependence of BA on γ is hump-shaped, for various payoff distributions, supply specifications, and initial endowments x_0.

IA.6 A Quantitative Exercise

The aim of this appendix is to calibrate the model parameters to assess whether our main result, that bid-ask spreads might increase when risk-aversion γ increases, obtains for reasonable values of the parameter γ. To limit the number of parameters to be calibrated, we focus on a single-asset version of the model. We calibrate it to match key moments of the returns and bid-ask spreads of 1-month at-the-money options on S&P 500 index (the most popular options on the S&P).

We assume that the asset payoff is given by

$$\delta = \bar{\delta} + \sigma\hat{\delta}, \text{ where } \hat{\delta} \sim Beta(\alpha, \beta).$$

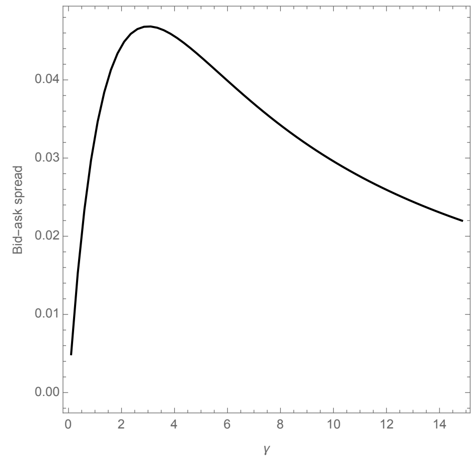

Figure IA.1: Bid-ask spread as a function of γ. We assume that $S(p) = p$, $\delta \sim \text{Beta}[0.2, 2]$, $L = 3$ and $x_0 = 1$.

We choose beta distribution, as it is flexible enough to allow us to match the first three moments of returns. The moments of δ can be computed in closed-form, as follows:

$$E[\delta] = \frac{\alpha\sigma}{\alpha+\beta} + \bar{\delta}, \tag{IA.12}$$

$$\text{var}[\delta] = \frac{\sigma^2}{(\alpha+\beta)^2}\frac{\alpha\beta}{\alpha+\beta+1}, \tag{IA.13}$$

$$\text{skewness}[\delta] = \frac{\text{third central moment}}{(\text{standard deviation})^3} = -\frac{2(\alpha-\beta)}{\sqrt{\frac{\alpha\beta}{\alpha+\beta+1}}(\alpha+\beta+2)}. \tag{IA.14}$$

Note that we do not use the normalization by (standard deviation)^3 for skewness in the main text.

There are 6 model parameters to calibrate: L, x_0, α, β, σ, and $\bar{\delta}$. Below we explain one-by-one how we calibrated these key model parameters.

Number of LPs. We set $L = 15$ following an industry report that estimates the number of market-makers between 5 and 15.[51] We chose the higher end of the range, because we focus on the most liquid segment of the options market.

Initial inventory. We do not have the data to estimate the parameter x_0 directly. However, the moments that we target depend on $x_0\sigma$ and do not depend on x_0 independently. Thus, we can normalize $x_0 = 1$ without loss of generality.

Parameters of the distribution. To calibrate the parameters α, β, σ, and $\bar{\delta}$, we target the three moments of option returns: skewness, Sharpe ratio, and volatility; in addition, we target the percentage bid ask-spreads. We use the estimates of skewness and volatility of annual

[51]Finextra, "Options Market Structure: Fragmented Reality", 16 May 2017.

returns on S&P 500 from Neuberger and Payne (2021): *skewness* $= \frac{\text{third central moment}}{(\text{standard deviation})^3} = -0.07$ and *volatility* $= 19.81\%$. We then compute the corresponding moments of option returns by treating the option as a levered portfolio of the underlying asset. The sample average portfolio weight of underlying asset in the replicating portfolio, i.e., the ratio of the dollar value of stocks in the replicating portfolio (Black-Scholes delta times the underlying price) to the price of the option is 43.5. Thus, the option volatility that we target is $43.5 \cdot 0.1981 = 8.6$. We use 5% as the estimate of the equity premium (Martin (2017)), which implies that the Sharpe ratio of option is approximately 0.25. For bid-ask spreads we target 0.5% (this is the average bid-ask spread for the S&P 500 options we focus on). The targeted moments are summarized in the table below.

targeted moment description	value
skewness	-0.07
volatility	8.6
Sharpe ratio	0.25
percentage bid-ask spread	0.5%

In our model, the above moments depend on γ. In our interpretation, the parameter γ changes over time. However, we need to assume some average level of γ when calibrating the above moments. We take the value of $\gamma = 30$ from Vayanos and Woolley (2013). It is likely a conservative estimate, as it is obtained for equity market and we have argued that the option market should have a higher γ. Our conclusions in this appendix are robust to other choices of "baseline" γ. The values of calibrated parameters are reported below.

calibrated parameter	value
α	12
β	10.1
$\bar{\delta}$	0.276
σ	24.26

Key result

Figure IA.2 shows that our key result, that bid-ask spreads decrease in risk aversion γ, holds for a wide range of levels of γ. It holds for any γ above 0.4. Note that the figure plots percentage bid-ask spreads (equal to BA/mid-price). When percentage bid-ask spreads decrease in γ so does the BA, since the mid-price decreases in γ.

IA.7 Unique Equilibrium for a Class of Unbounded-support Distributions

Consider the following class of distributions. Note that the Gaussian distribution is the special case.

Assumption IA.1. *The density $\eta(q)$ is such that there exists a positive definite matrix A and an $\alpha > 1$ and two positive constants $0 \leq C_1 \leq C_2$ such that*

$$C_1 e^{-\|Aq\|^\alpha} \ \leq \ \eta(q) \ \leq \ C_2 e^{-\|Aq\|^\alpha}.$$

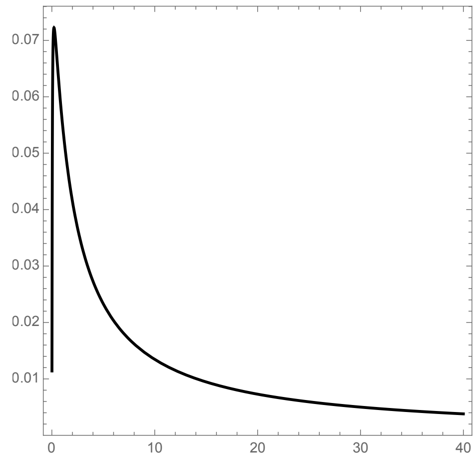

Figure IA.2: Percentage bid-ask spread as a function of γ in a calibrated model.

For such class of distributions, we prove the uniqueness of the limiting equilibria in the theorem below.

Theorem IA.2. *Expanding sets $A_n \subset \mathbb{R}^N$ are bounded sets such that $\mathbf{1}_{A_n} \to 1$ almost surely. Suppose that the density $\eta(q)$ satisfies Assumption IA.1. Then, there exists a unique regular equilibrium. That is, for any expanding sets A_n, their respective bounded support equilibria converge to the equilibrium of the non-truncated problem.*

IA.8 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 (IA.6) and (IA.8) converge to their corresponding counterparts in the Gaussian benchmark as the truncation bounds go to infinity. We start by deriving the Gaussian benchmark.

IA.8.1 CARA-normal benchmark

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

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

It follows that

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

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

$$I(q) = \ [\mu - \gamma\Sigma x_0] - \frac{L-1}{L-2}\gamma\Sigma q \tag{IA.15}$$

$$\Longrightarrow \ \Lambda(q) \equiv \ -\frac{1}{L-1}I'(q) \ = \ \frac{\gamma}{L-2}\Sigma\,. \tag{IA.16}$$

IA.8.2 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

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

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

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

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

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.

IA.8.3 CARA-normal benchmark as a limit II: 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

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

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

$$\left| e^{-\gamma(x_0+q)^\top y} \mathbb{1}_{I_b}(y) f_X(y) \frac{1}{\mu_b} \right| \;<\; e^{-\gamma(x_0+q)^\top y} f_X(y) \frac{1}{\mu_{b_0}}\,. \tag{IA.17}$$

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 IA.17 also implies that the BCT applies to $I(q)$ and $\Lambda(q)$:

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

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

IA.9 Proofs

IA.9.1 Proof of Lemma IA.1

We present the proof below for the paper to be self-contained.
Proof of Lemma IA.1. Start with Taylor's theorem with integral form of remainder

$$f(S) = f(S_0) + f'(S_0)(S - S_0) + \int_{S_0}^{S} (S-k) f''(k) dk.$$

First, note that for $S < S_0$, we have

$$\begin{aligned}\int_{S_0}^{S} (S-k) f''(k) d\xi &= \int_{S}^{S_0} (k-S) f''(k) d\xi \\ &= \int_{0}^{S_0} (k-S)^+ f''(k) d\xi.\end{aligned}$$

Second, for $S > S_0$, we have

$$\int_{S_0}^{S} (S-k) f''(k) dk = \int_{S_0}^{\infty} (S-k)^+ f''(k) dk.$$

Note that $\int_0^{S_0} (k-S)^+ f''(k) d\xi = 0$ for $S > S_0$ and $\int_{S_0}^{\infty} (S-k)^+ f''(k) dk = 0$ for $S < S_0$. Therefore,

$$\int_{S_0}^{S} (S-k) f''(k) d\xi = \int_{0}^{S_0} (k-S)^+ f''(k) d\xi + \int_{S_0}^{\infty} (S-k)^+ f''(k) dk.$$

Finally, note that

$$S - S_0 = (S - S_0)^+ - (S_0 - S)^+ .$$

Thus, we get

$$f(S) = f(S_0) + f'(S_0)\left((S-S_0)^+ - (S_0 - S)^+\right) + \int_{0}^{S_0} (k-S)^+ f''(k) dk + \int_{S_0}^{\infty} (S-k)^+ f''(k) dk.$$

■

IA.9.2 Proof of Proposition IA.1

Proof of Proposition IA.1. For calls, we use Carr-Madan formula (IA.2) with $f(S) = ((S-K)^+)^2$. We plug into (IA.2) $f'(S_0) = (S_0 - K)^+$ and $f''(k) = 2 \cdot 1(k > K)$.[52] We obtain

$$\begin{aligned} \left((S-K)^+\right)^2 = &\left((S_0-K)^+\right)^2 + 2(S_0-K)^+\left((S-S_0)^+ - (S_0-S)^+\right) + \\ &+ 2\int_0^{S_0} (k-S)^+ 1(k>K)dk + 2\int_{S_0}^{\infty} (S-k)^+ 1(k>K)dk. \end{aligned} \tag{IA.18}$$

We then apply risk-neutral expectation to the formula above. Noting that $E^*[(S-k)^+] = C(k)$ and $E^*[(k-S)^+] = P(k)$, we obtain

$$\begin{aligned} E^*\left[\left((S-K)^+\right)^2\right] = &\left((S_0-K)^+\right)^2 + 2(S_0-K)^+\left(C(S_0) - P(S_0)\right) + \\ &+ 2\int_0^{S_0} P(k)1(k>K)dk + 2\int_{S_0}^{\infty} C(k)1(k>K)dk. \end{aligned} \tag{IA.19}$$

The above can also be written as

$$E^*[(S-k)^+] = \begin{cases} (S_0-K)^2 + 2(S_0-K)\left(C(S_0)-P(S_0)\right) + 2\int_K^{S_0} P(k)dk + 2\int_{S_0}^{\infty} C(k)dk, & K < S_0 \\ 2\int_K^{\infty} C(k)dk, & K > S_0. \end{cases}$$

The formula for variance then follows.

For puts, we have

$$\begin{aligned} \left((K-S)^+\right)^2 = &\left((K-S_0)^+\right)^2 - 2(K-S_0)^+\left((S-S_0)^+ - (S_0-S)^+\right) + \\ &+ 2\int_0^{S_0} (k-S)^+ 1(k<K)dk + 2\int_{S_0}^{\infty} (S-k)^+ 1(k<K)dk. \end{aligned}$$

Applying risk-neutral expectations we obtain, similarly to calls,

$$E^*[(S-k)^+] = \begin{cases} (K-S_0)^2 - 2(K-S_0)\left(C(S_0)-P(S_0)\right) + 2\int_0^{S_0} P(k)dk + 2\int_{S_0}^{K} C(k)dk, & K > S_0 \\ 2\int_0^{K} P(k)dk, & K < S_0. \end{cases}$$

The formula for variance then follows. ■

IA.9.3 Proof of Proposition IA.2

Proof of Proposition IA.2. The proof is a simpler version of the corresponding proof in the non-competitive economy. We will rely on the optimality and consistency conditions. We first turn the optimization problem into the corresponding certainty-equivalent optimization problem:

$$\max_D f(q; x_0) - p^\top q\,.$$

[52]If a function $f(S)$ has a kink, then $f''(k)$ is the delta function at k times the jump of $f'(k)$. But for $f(S) = (S-K)^2$, we have $f'(K) = 0$, and hence no correction is needed.

The first-order condition in the competitive economy is

$$\nabla f(q; x_0) \; = \; P \equiv I(q; x_0) \, ,$$

which yields (IA.3). ■

IA.9.4 Proof of Proposition IA.3

Proof of Proposition IA.3. Take $Y = x$, $x^i = x_0^i + q^i$ and $x^j = x_0^j + q^j$. By Envelope Theorem, we have $\nabla U(x) = \nabla f^i(x^i) = \nabla f^j(x^j)$ for all $i \in \{1, ..L\}$ and $j \in \{1, ..M\}$. On the other hand, in the competitive economy, we have $p = \nabla f^i(x_0^j + q^j) = \nabla f^j(x_0^j + q^j)$. Hence, prices are unaffected by the distribution of asset holdings across traders. ■

IA.9.5 Proof of Proposition IA.4

Proof of Proposition IA.4. The first result appears in Proposition 3. In the competitive economy, we show in Proposition IA.2 that the inverse demand for asset k is given by $I_k = g_k(-\gamma(x_0 + q))$. Thus, the bid-ask spread is given by $\mathrm{BA}_k = \lim_{n_k \to 0} \frac{I_k(-n_k 1_k) - I_k(n_k 1_k)}{n_k} = 2\gamma g_{kk}(-\gamma x_0)$. ■

IA.9.6 Proof of Proposition IA.5

Proof of Proposition IA.5. The function g satisfies

$$\nabla g(q) = \frac{E[\delta \exp(q^\top \delta)]}{E[\exp(q^\top \delta)]} \, .$$

We start with the competitive case. It follows from equation (IA.3) that

$$\begin{aligned} I(q; x_0) = I(q + x_0) &= \nabla g(-\gamma(x_0 + q)) \\ &= \frac{E[\delta \exp(-\gamma(x_0 + q)^\top \delta)]}{E[\exp(-\gamma(x_0 + q)^\top \delta)]} \, . \end{aligned}$$

Thus,

$$\begin{aligned} I(-x_0 + \epsilon; x_0) = I(\epsilon) &= \nabla g(-\gamma(\epsilon)) \\ &= \frac{E[\delta \exp(-\gamma(\epsilon)^\top \delta)]}{E[\exp(-\gamma(\epsilon)^\top \delta)]} \, . \end{aligned}$$

The (sequence of) random variable ϵ converges to the zero vector *pointwise* as $\sigma_\epsilon \to 0$. It then follows that

$$\lim_{\sigma_\epsilon \to 0} P \; = \; \lim_{\sigma_\epsilon \to 0} I(-x_0 + \epsilon; x_0) \; = \; E[\delta] \, .$$

Next, we consider the strategic equilibrium. In this case, we have

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

Thus,

$$
\begin{aligned}
\lim_{\sigma_\epsilon\to 0} I(-x_0+\epsilon) &= (L-1)\lim_{\sigma_\epsilon\to 0}\int_1^\infty \xi^{-L}\frac{E[\delta\exp(-\gamma(\xi(-x_0+\epsilon)+x_0)\delta)]}{E[\exp(-\gamma(\xi(-x_0+\epsilon)+x_0)\delta)]}d\xi \\
&= (L-1)\int_1^\infty \xi^{-L}\lim_{\sigma_\epsilon\to 0}\frac{E[\delta\exp(-\gamma(\xi(-x_0+\epsilon)+x_0)\delta)]}{E[\exp(-\gamma(\xi(-x_0+\epsilon)+x_0)\delta)]}d\xi \\
&= (L-1)\int_1^\infty \xi^{-L}\frac{E[\delta\exp(-\gamma(\xi(-x_0)+x_0)\delta)]}{E[\exp(-\gamma(\xi(-x_0)+x_0)\delta)]}d\xi \\
&= (L-1)\int_1^\infty \xi^{-L}\nabla_i g(-\gamma x_0(1-\xi))\,d\xi\,,
\end{aligned}
$$

where the change of expectation and integration follows from Fubuni's Theorem and Lemma 6. Next, we derive the asymptotics in the limit as $\gamma\to 0$. Consider

$$
\frac{E[\delta\exp(-\gamma x_0(1-\xi)\delta)]}{E[\exp(-\gamma x_0(1-\xi)\delta)]} = \frac{\int\delta\exp\left(-\gamma x_0(1-\xi)\delta\right)\eta(\delta)d\delta}{\int\exp\left(-\gamma x_0(1-\xi)\delta\right)\eta(\delta)d\delta}\,.
$$

It follows from the Taylor series expansion that, in the limit as $\gamma\to 0$, we have

$$
\begin{aligned}
&\frac{\int\delta_k\exp\left(-\gamma x_0(1-\xi)\delta\right)\eta(\delta)d\delta}{\int\exp\left(-\gamma x_0(1-\xi)\delta\right)\eta(\delta)d\delta} \\
&\sim\ \frac{E[\delta]-\gamma(1-\xi)E[\delta\delta^\top x_0]+\frac{1}{2}\gamma^2(1-\xi)^2E[\delta(\delta^\top x_0)^2]}{1-\gamma(1-\xi)x_0^\top E[\delta]+\frac{1}{2}\gamma^2(1-\xi)^2x_0^\top E[\delta\delta^\top]x_0}+O\left(\gamma^3\right) \\
&\sim \mathrm{E}[\delta]+(1-\xi)\gamma\left(E[\delta]E[\delta^\top]-E[\delta\delta^\top]\right)x_0 \\
&\quad+\frac{1}{2}(1-\xi)^2\left(E[\delta(\delta^\top x_0)^2]-x_0^\top E[\delta\delta^\top]x_0E[\delta]+2x_0^\top E[\delta]\left[x_0^\top E[\delta]E[\delta]-E[\delta\delta^\top x_0]\right]\right)\gamma^2+O\left(\gamma^3\right) \\
&=\ \mathrm{E}[\delta]-(1-\xi)\gamma\Sigma x_0\ +\frac{1}{2}(1-\xi)^2\gamma^2 coskew(\delta,x_0^\top\delta,x_0^\top\delta)+O\left(\gamma^3\right)\,.
\end{aligned}
$$

Thus,

$$
\lim_{\sigma_\epsilon\to 0} I(-x_0+\epsilon)\sim\ \mathrm{E}[\delta]+\frac{1}{L-2}\gamma\Sigma x_0\ +\frac{1}{2}\frac{2}{(L-2)(L-3)}\gamma^2 coskew(\delta,x_0^\top\delta,x_0^\top\delta)+O\left(\gamma^3\right)\,.
$$

■

IA.9.7 Proof of Proposition IA.6

Proof of Proposition IA.6. Denote $q = s/L$. We have

$$P_j(s) = I_j(s/L) = (L-1)\int_1^{\infty} \xi^{-L} g_j(-\gamma(x_0 + \xi q))\, d\xi.$$

For the immediate price reaction, we have

$$\begin{aligned}
\lim_{k\to 0} \frac{P_j(k1_i) - P_j(0)}{k} &= \lim_{k\to 0}\left(\frac{L-1}{1}\int_1^{\infty} \xi^{-L}\left(\frac{g_j(-\gamma(x_0+\xi k/L1_i)) - g_j(-\gamma(x_0))}{k}\right) d\xi\right) \\
&= -\gamma \frac{L-1}{L}\int_1^{\infty} \xi^{1-L} g_{ji}(-\gamma x_0)\, d\xi \\
&= -\gamma \frac{L-1}{L\,(L-2)} g_{ji}(-\gamma x_0) \\
&= -\gamma \frac{L-1}{L\,(L-2)} \mathrm{cov}^*\left(\delta_i, \delta_j\right).
\end{aligned}$$

The last transition is true because

$$\nabla^2 g(y) = \; E\left[\delta\delta^\top \frac{e^{y^\top\delta}}{E[e^{y^\top\delta}]}\right] \; - \; E\left[\delta \frac{e^{y^\top\delta}}{E[e^{y^\top\delta}]}\right] E\left[\delta^\top \frac{e^{y^\top\delta}}{E[e^{y^\top\delta}]}\right],$$

so,

$$g_{ij}(y)\big|_{y=-\gamma x_0} = \; \mathrm{cov}^*(\delta_i, \delta_j).$$

For the reversal matrix, we have

$$R_{ij} = \lim_{k\to 0}\frac{\Lambda_{ij}(k1_i/L)k/L}{k} = 1/L\Lambda_{ij}(0).$$

Note that price impact matrix is given by

$$\begin{aligned}
\Lambda_{ij}(0) &= \frac{\gamma}{1}\int_1^{\infty} \xi^{1-L} g_{ij}(-\gamma x_0)\, d\xi \\
&= \frac{\gamma}{(L-2)} g_{ij}(-\gamma x_0) \\
&= \frac{\gamma}{(L-2)} \mathrm{cov}^*(\delta_i, \delta_j).
\end{aligned}$$

Hence, $R_{ij} = \frac{\gamma}{(L-2)L}\mathrm{cov}^*(\delta_i, \delta_j)$. ■

IA.9.8 Proof of Theorem IA.1

Proof of Theorem IA.1. The key is to show that the inverse demand solving

$$\nabla f(q) + \frac{1}{L-1}\nabla I^*(q) q = I^*(q) \tag{IA.20}$$

continues to be optimal in the extended model. We do so in several steps.

Given that other traders' equilibrium demands at $t = 0$ are given by $I^*(\cdot)$, the ex-post optimization problem of a trader i at $t = 0$ can be written as follows

$$\sup_{q_0} \left\{ \begin{array}{c} f(x_0 + q_0) - f(x_0) - I^* \left(\frac{s-q_0}{L-1}\right)^\top q_0 + \\ f\left(x_0 + q_0 + q^*_{1/2}\right) - f(x_0 + q_0) - p^{*\top}_{1/2} q^*_{1/2} \end{array} \right\}. \tag{IA.21}$$

Here we denote by $p^*_{1/2}$ and $q^*_{1/2}$ the equilibrium price and quantity in the $t = 1/2$ subgame following allocation of q_0 to trader i. It is sufficient to show that (IA.21) is maximized with $q_0 = s/L$ and $q^*_{1/2} = 0$.

Denote the first and the second line in (IA.21) as, respectively, A and B:

$$A(q_0) \equiv f(x_0 + q_0) - f(x_0) - I^* \left(\frac{s - q_0}{L-1}\right) q_0,$$

$$B(q_0, q_{1/2}) \equiv f\left(x_0 + q_0 + q_{1/2}\right) - f(x_0 + q_0) - p_{1/2}(q_0, q_{1/2}) q_{1/2}.$$

A and B can be interpreted as trading gain of a trader i at $t = 0$ and $t = 1/2$.

Denote

$$q^* \equiv s/L.$$

Step 1. For any $\Delta \neq 0$, we have

$$A(q^* + \Delta) - A(q^*) < f(q^* + \Delta) - f(q^*) + (L-1)\left(f\left(q^* - \frac{\Delta}{L-1}\right) - f(q^*)\right). \tag{IA.22}$$

This is proved in Lemma IA.2 (to follow). Here, we note the intuition. A change in the allocation to trader i from $q^*$to $q^* + \Delta$ implies a change of allocations from q^* to $q^* - \frac{\Delta}{L-1}$ for all other $L - 1$ LPs. The right-hand side of (IA.22) is then the change in welfare of all LPs. Equation (IA.22) then simply states that a change in the trading gain of a trader i is less than the change in the welfare of all traders.

Step 2. For any $\Delta \neq 0$, and for any $q_{1/2}$, we have

$$B(q^* + \Delta, q_{1/2}) - B(q^*, 0) < \begin{array}{c} f(q^*) - f(q^* + \Delta) + \\ (L-1)\left(f(q^*) - f\left(q^* - \frac{\Delta}{L-1}\right)\right) \end{array}. \tag{IA.23}$$

First note that since for any q_0, $B(q_0, 0) = 0$, we have $B(q^* + \Delta, q_{1/2}) - B(q^*, 0) = B(q^* + \Delta, q_{1/2}) - B(q^* + \Delta, 0)$. The term $B(q^* + \Delta, q_{1/2}) - B(q^* + \Delta, 0)$ is a change of the trading gain of trader i when he trades $q_{1/2}$ instead of 0 at $t = 1/2$. It must be less than the change in the welfare of all traders. Indeed, traders have an option not to trade, so the change in their trading gains cannot be negative. Thus, we write

$$\begin{aligned}
&B(q^*+\Delta, q_{1/2}) - B(q^*+\Delta, 0) \leq \\
&\left\{ \begin{array}{c} f\left(q^*+\Delta+q_{1/2}\right) - f\left(q^*+\Delta\right) + \\ (L-1)\left(f\left(q^* - \frac{\Delta}{L-1} - \frac{q_{1/2}}{L-1}\right) - f\left(q^* - \frac{\Delta}{L-1}\right)\right) \end{array} \right\} \leq \\
\sup_{q_{1/2}} &\left\{ \begin{array}{c} f\left(q^*+\Delta+q_{1/2}\right) - f\left(q^*+\Delta\right) + \\ (L-1)\left(f\left(q^* - \frac{\Delta}{L-1} - \frac{q_{1/2}}{L-1}\right) - f\left(q^* - \frac{\Delta}{L-1}\right)\right) \end{array} \right\} = \\
&f\left(q^*\right) - f\left(q^*+\Delta\right) + (L-1)\left(f\left(q^*\right) - f\left(q^* - \frac{\Delta}{L-1}\right)\right) \; .
\end{aligned}$$

The last equality is true because the aggregate welfare is maximized at the symmetric allocation q^*, i.e., when $q_{1/2} = -\Delta$.

Steps 1 and 2 imply that (IA.21) is indeed maximized with $q_0 = s/L$ and $q^*_{1/2} = 0$. This, in turn, implies that inverse demand solving (IA.20) continues to be optimal in the extended model.

It remains to determine the prices at $t = 1/2$. Since there is no trade, the prices must be equal to marginal utility, thus

$$p_{1/2} = \nabla f(x_0 + q^*).$$

■

Lemma IA.2. *For any $\Delta \neq 0$, we have*

$$A(q^*+\Delta) - A(q^*) < f\left(q^*+\Delta\right) - f\left(q^*\right) + (L-1)\left(f\left(q^* - \frac{\Delta}{L-1}\right) - f\left(q^*\right)\right). \qquad \text{(IA.24)}$$

Proof. Denote

$$\begin{aligned}
c(\Delta) &= \nabla\left(I^*\left(\frac{s-q}{L-1}\right)^{\top} q\right)\Bigg|_{q=q^*+\Delta} \qquad \text{(IA.25)}\\
&= I^*\left(q^* - \frac{\Delta}{L-1}\right) - \frac{1}{L-1}\nabla I^*\left(q^* - \frac{\Delta}{L-1}\right)\left(q^*+\Delta\right).
\end{aligned}$$

Then, the change in the trading gain of trader i due to deviation $q^* \to q^* + \Delta$ is

$$\begin{aligned}
A(q^*+\Delta) - A(q^*) &= \int_0^1 \Delta^{\top}\left(\nabla f\left(q^*+t\Delta\right) - c(t\Delta)\right) dt \\
&< \int_0^1 \Delta^{\top}\left(\nabla f\left(q^*+t\Delta\right) - \nabla f\left(q^* - t\frac{\Delta}{L-1}\right)\right) dt \\
&= f\left(q^*+\Delta\right) - f\left(q^*\right) + (L-1)\left(f\left(q^* - \frac{\Delta}{L-1}\right) - f\left(q^*\right)\right).
\end{aligned}$$

To get the first inequality we used that for any $\Delta \neq 0$ we have $\Delta^{\top}c(\Delta) > \Delta^{\top}\nabla f\left(q^* - \frac{\Delta}{L-1}\right)$., which is proved in Lemma IA.3. ■

Lemma IA.3. *For any $\Delta \neq 0$ we have $\Delta^\top c(\Delta) > \Delta^\top \nabla f\left(q^* - \frac{\Delta}{L-1}\right)$.*

Proof. Indeed, from (IA.20), we get $I^*\left(q^* - \frac{\Delta}{L-1}\right) = \nabla f(q^* - \frac{\Delta}{L-1}) + \frac{1}{L-1}\nabla I^*(q^* - \frac{\Delta}{L-1})\left(q^* - \frac{\Delta}{L-1}\right)$. Substitute this to (IA.25) to get

$$z(\Delta) \equiv c(\Delta) - \nabla f\left(q^* - \frac{\Delta}{L-1}\right) = -\frac{1}{L-1}\nabla I^*\left(q^* - \frac{\Delta}{L-1}\right)\left(\frac{L\Delta}{L-1}\right).$$

Multiply both parts of the above by $\Delta^\top$ and account for the fact that $\Delta I(\cdot)$ is negative-definite to get that for any $\Delta \neq 0$, $\Delta^\top z(\Delta) > 0$. The statement follows. ■

IA.9.9 Proof of Proposition IA.7

Proof of Proposition IA.7. In the period $t = 1/2$, LPs solve $\max_D f(x_0 + s/L + D) - P_{1/2}^\top D$. Plugging $D = 0$ (no trade) to the first-order condition, we get $P_{1/2}(s) = \nabla f(x_0 + s/L)$. In the period $t = 0$, LPs solve $\max_D f(x_0 + D) - P_0^\top D$. Plugging $D = s/L$ (symmetric equilibrium), to the first-order condition we get $P_0(s) = \nabla f(x_0 + s/L)$. ■

IA.9.10 Proof of Proposition IA.8

Proof of Proposition IA.8. The proposition follows because the optimization problem (IA.5) separates into maximizing each of the exp(·) terms separately:

$$\begin{aligned}
&\max_{D^i(p,t)} E[-\exp(-\gamma c_t - \beta t)], \\
\text{s.t.}\quad & c_t = (\delta_t - p_t(D^i(p), D(p)))^\top (D_t^i(p) + x_{0,t}) \text{ and} \\
& p(D^i(p_t), D(p_t)) : D^i(p_t, t) + (L-1)D(p_t, t) = s_t.
\end{aligned} \tag{IA.26}$$

That is, the optimization problem is reduced to (1). The analysis of the baseline model then implies the closed-form solutions.

■

IA.9.11 Proof of Theorem IA.2

Proof of Theorem IA.2. We have

$$g(\xi) = \log E[e^{\xi X}] = \log \int e^{\xi X}\eta(X)dX.$$

First, we note that

$$\nabla g(\xi) = \frac{\int X e^{\xi X}\eta(X)dX}{\int e^{\xi X}\eta(X)dX},$$

while

$$\nabla g_n(\xi) = \frac{\int_{A_n} X e^{\xi X}\eta(X)dX}{\int_{A_n} e^{\xi X}\eta(X)dX}.$$

Then, we note that the Lebesgue dominated convergence implies that for each ξ, we have $\nabla g_n(\xi) \to g(\xi)$. We also have

$$I(q) = (L-1)\int_1^\infty \xi^{-L}\nabla g(-\gamma(x_0+\xi q))\,d\xi,$$

while

$$I_n(q) = (L-1)\int_1^\infty \xi^{-L}\nabla g_n(-\gamma(x_0+\xi q))\,d\xi.$$

Below we find an integrable majorant for $\|\nabla g_n(\xi)\|$. The Lebesgue Dominated Convergence Theorem then implies that for each q, we have $I_n(q) \to I(q)$.

We have

$$\begin{aligned} &\|\int Xe^{\xi X}\eta(X)dX\| \le \int \|X\|e^{\xi X}\eta(X)dX \\ &\le \int \|X\|e^{\xi X}C_2 e^{\|AX\|^\alpha}dX \;=\; \{X = A^{-1}Y,\; dX \;=\; |\det A|dY\} \\ &= |\det A|\int \|A^{-1}Y\|e^{\xi A^{-1}Y}C_2 e^{\|Y\|^\alpha}dY \;\le\; |\det A|\|A^{-1}\|\int \|Y\|e^{\tilde{\xi}Y}C_2 e^{-\|Y\|^\alpha}dY, \end{aligned} \tag{IA.27}$$

where $\tilde{\xi} = A^{-1}\xi$. Let us now rotate the coordinates with an orthogonal matrix U such that $Y = UZ$ and $\tilde{\xi} = \|\tilde{\xi}\|Ue_1$, where $e_1 = (1,0,\cdots,0)$. Then,

$$\int \|Y\|e^{\tilde{\xi}Y}C_2 e^{-\|Y\|^\alpha}dY \;=\; \int \|Z\|e^{\|\tilde{\xi}\|Z_1}C_2 e^{-\|Z\|^\alpha}dZ\,.$$

Now, we make a transformation $Z = \|\tilde{\xi}\|^{1/(\alpha-1)}Q$. Then,

$$\int \|Z\|e^{\|\tilde{\xi}\|Z_1}C_2 e^{-\|Z\|^\alpha}dZ \;=\; (\|\tilde{\xi}\|^{1/(\alpha-1)})^{N+1}\int \|Q\|e^{\|\tilde{\xi}\|^{\frac{\alpha}{1+\alpha}}(Q_1-\|Q\|^\alpha)}C_2 dQ.$$

The same argument implies that

$$\int e^{\xi X}\eta(X)dX\| \le \int \|X\|e^{\xi X}\eta(X)dX \;\ge\; C_3(\|\tilde{\xi}\|^{1/(\alpha-1)})^N\int e^{\|\tilde{\xi}\|^{\frac{\alpha}{1+\alpha}}(Q_1-\|Q\|^\alpha)}C_2 dQ$$

for some constant $C_3 > 0$.

Thus, it remains to show that

$$\frac{\int \|Q\|e^{\|\tilde{\xi}\|^{\frac{\alpha}{1+\alpha}}(Q_1-\|Q\|^\alpha)}dQ}{\int e^{\|\tilde{\xi}\|^{\frac{\alpha}{1+\alpha}}(Q_1-\|Q\|^\alpha)}dQ}$$

stays uniformly bounded when $\|\tilde{\xi}\| \to \infty$. This follows from the standard Saddle Point Theorem (Lemma IA.4 below) because this quotient converges to a finite limit when $\|\tilde{\xi}\| \to \infty$: it converges to $\|Q_*\|$ where $Q_* = \arg\max(Q_1 - \|Q\|^\alpha)$.

Lemma IA.4. *Suppose that $g(x)$ is strictly concave, attains a global maximum at x_0, and is two-*

times continuously differentiable in a neighborhood of x_0*. Suppose also that* $\int(|f(x)|+1)e^{\gamma g(x)}dx < \infty$ *for some* $\gamma > 0$*. Then,*

$$\lim_{\gamma \to +\infty} \frac{\int f(x)e^{\gamma g(x)}dx}{\int e^{\gamma g(x)}dx} = f(x_0)\,.$$

Proof. The claim follows from classic results in Fedoryuk (1987).

■

■

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.

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.

Geert Bekaert, Eric C Engstrom, and Nancy R Xu. The time variation in risk appetite and uncertainty. Management Science, 2021.

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.

Melanie Cao and Jason Wei. Option market liquidity: Commonality and other characteristics. Journal of Financial Markets, 13(1):20–48, 2010.

Peter Carr and Dilip Madan. Towards a theory of volatility trading. Option Pricing, Interest Rates and Risk Management, Handbooks in Mathematical Finance, 22(7):458–476, 1998.

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

Hui Chen, Scott Joslin, and Sophie Xiaoyan Ni. Demand for crash insurance, intermediary constraints, and risk premia in financial markets. The Review of Financial Studies, 32(1): 228–265, 2019.

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

Carole Comerton-Forde, Terrence Hendershott, Charles M Jones, Pamela C Moulton, and Mark S Seasholes. Time variation in liquidity: The role of market-maker inventories and revenues. The journal of finance, 65(1):295–331, 2010.

Carole Comerton-Forde, Terrence Hendershott, Charles M. Jones, Pamela C. Moulton, and Mark S. Seasholes. Time variation in liquidity: The role of market-maker inventories and revenues. Journal of Finance, 65(1):295–331, 2010. URL https://EconPapers.repec.org/RePEc:bla:jfinan:v:65:y:2010:i:1:p:295-331.

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.

MV Fedoryuk. Asymptotics: integrals and series. Mathematical Reference Library, 1987.

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

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

Paul M Healy and Krishna G Palepu. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. Journal of accounting and economics, 31(1-3):405–440, 2001.

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.

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

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

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

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

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.

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

Hong Liu and Yajun Wang. Market making with asymmetric information and inventory risk. Journal of Economic Theory, 163:73–109, 2016.

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.

Dmitriy Muravyev and Neil D Pearson. Options trading costs are lower than you think. The Review of Financial Studies, 33(11):4973–5014, 2020.

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

Anthony Neuberger and Richard Payne. The skewness of the stock market over long horizons. The Review of Financial Studies, 34(3):1572–1616, 2021.

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.

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.

Albert N. Shiryaev. Probability. Springer-Verlag New York, 1996.

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.

Dimitri Vayanos and Paul Woolley. An institutional theory of momentum and reversal. The Review of Financial Studies, 26(5):1087–1145, 2013.

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.

Jason Wei and Jinguo Zheng. Trading activity and bid–ask spreads of individual equity options. Journal of Banking & Finance, 34(12):2897–2916, 2010.

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.