The London School of Economics and Political Science

THREE ESSAYS IN FINANCIAL ECONOMICS

Sergey Glebkin

A thesis submitted to the Department of Finance of the London School of Economics for the degree of Doctor of Philosophy, London, May 2016

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of 32,828 words.

I confirm that Chapter 3 is co-authored with Ji Shen.

Abstract

The thesis examines how different aspects of market quality are affected by imperfect competition. The first chapter presents a model of strategic liquidity provision in a uniform-price auction that does not require normally distributed asset payoffs. I propose a constructive solution method: finding the equilibrium reduces to solving a linear ODE. With non-normal payoffs, the price response becomes an asymmetric, non-linear function of order size: greater for buys than sells and concave (convex) for small sell (buy) orders when asset payoffs are positively skewed; concave for large sell (buy) orders when payoffs are bounded below (above). The model speaks to key empirical findings and provides new predictions concerning the shape of price response.

The second chapter analyses a market with large and small traders with different values. In such market illiquidity and information efficiency are complements. Policy measures promoting liquidity might be harmful for information efficiency and vice versa. An increase in risk-bearing capacity may harm liquidity. An increase in the precision of information may harm information efficiency. Increasing market power or breaking up a centralized market into two separate exchanges might improve welfare. Multiple equilibria, in which higher liquidity is associated with lower information efficiency, are possible.

The third chapter (co-authored with Ji Shen) studies OTC markets. Traders in a market a-la Duffie, Garleanu and Pedersen (2005) can search via Multilateral Trading Platform (MTP), querying n dealers and running first-price auction among them. Dealers have homogenous valuation for the asset, yet the distribution of bid and ask prices is non-degenerate: uncertainty about the number of competitor dealers responded induces mixed-strategy equilibrium. We provide testable implications linking skewness and dispersion of bid and ask prices to dealers response rate in the auctions.

Acknowledgments

I would like to thank my advisors, Dimitri Vayanos and Georgy Chabakauri for all the invaluable guidance and support I received. Their comments and suggestions improved this thesis a lot. Their personal example helped me to develop as a scholar.

I also benefited from numerous discussions with permanent and visiting faculty at the London School of Economics. I would like to thank Amil Dasgupta, Daniel Ferreira, Dong Lou, Peter Kondor, Albert (Pete) Kyle, Semyon Malamud, Igor Makarov, Ian Martin, Rohit Rahi, Anna Obizhaeva, Martin Oehmke, Marzena Rostek, Gyuri Venter, Kathy Yuan and Kostas Zachariadis. Peter Kondor and Igor Makarov provided vital help during my job market year. Marzena Rostek provided invaluable enthusiasm, support and feedback for the first chapter of the thesis.

The five years at LSE were such a joy, thanks to my fellow PhD students: Svetlana Bryzgalova, Huazhi Chen, Hoyong Choi, Shiyang Huang, Jason Donaldson, Jesus Gorrin John Kuong, Yiqing Lu, Milan Martinovic, Olga Obizhaeva, Giorgia Piacentino, Michael Punz, Ji Shen, Luana Zaccaria, Jing Zeng. Thanks for being around.

I would also like to acknowledge the financial support from the Economic and Social Research Council and the Paul Woolley Centre.

I am indebted to my parents, Anna and Vladimir, and to all my relatives and friends back in Russia for their support and faith in me. Thanks to Patrick and Beverley Keogh, my Great British Friends, for all their help.

I could not thank enough my wife Irina for being with me through all the ups and downs. Thank you for your love and kindness. Lastly, thanks to little Anna for sharing the joy of life and for all your numerous questions.

Contents

Chapter 1

Strategic Trading without Normality

1.1 Introduction

In many markets trade is dominated by large institutional investors (such as mutual and pension funds) whose trades can affect prices. These investors often trade strategically, taking their price impact into account.[1] Empirical evidence documents that prices react to orders of such investors in an asymmetric and non-linear way: purchases typically have greater price impact compared to sells and price response is a concave function of order size.[2]

Previous papers on strategic trading have often adopted a CARA-normal framework for tractability: traders have negative exponential (CARA) utility functions and asset payoffs are normally distributed. The CARA-normal models feature linear equilibria (in which the price is a linear function of order size and purchases and sells have the same price impact) which are hard to align with empirical evidence. Normality also implies that higher moments play no role which may not be true in practice, and that asset payoffs are not bounded which is unrealistic, e.g., due to limited liability. In this paper I present a tractable model of strategic trading that allows for general distribution of asset payoffs. I show that if asset payoffs are

[1]Some investors, such as J.P. Morgan or Citigroup, have in-house optimal execution desks which devise trading strategies to minimize price-impact costs. Other investors use the software and services provided by more specialized trading firms.

[2]Hausman et al. (1992), Almgren et al. (2005), Frazzini et al. (2014) find concave price reaction functions (absolute value of price change as a function of order size) for equities. Muraviev (2015) presents the evidence for options. He decomposes the price reaction function into inventory and information components and finds that both are concave. Regarding asymmetry, Saar (2001) summarizes the evidence that shows bigger price impact of buy orders compared to sell orders. However, Chiyachantana et al. (2004) link the asymmetry to the underlying market condition and find that in bullish markets buy orders have a bigger price impact than sells, while in the bearish markets sells have a higher price impact.

positively skewed and bounded the model can speak to key empirical findings concerning nonlinearity and asymmetry of price impact. The main technical challenge is that with non-normal distribution the traditional guess-and-verify approach is no longer applicable as it is not clear what should be the guess. I propose a constructive solution method that allows to overcome this difficulty and to solve the model in closed form for any distribution.

I assume that CARA traders exchange a risky for a riskless asset over one period. Traders have the same risk aversion coefficient and are symmetrically informed. Trading is structured as a uniform-price double auction: traders submit simultaneously demand functions, and all trades are executed at the price that clears the market. My main innovation relative to previous literature is to assume that the distribution of the risky asset payoff is completely general save for the technical restriction. The restriction requires that a *risk function*, a transformation of the cumulant generating function (CGF) which I introduce in this paper, exists.[3] This restriction holds for any distribution with finite support as well as for many infinite support distributions, including normal and mixture of normals. In addition to the symmetric CARA traders, there is a "block trader" who submits an exogenous market order of random size. Trade occurs because the CARA traders compete to absorb part of the block trader's order, hence providing liquidity to that trader. In most of the paper I assume that the block trader's order is independent of the asset payoff, and so the block trader is uninformed. In that setting the unique source of price impact is inventory risk. I also extend my model to allow the block trader's order to be correlated with the asset payoff. In that extension, price impact is driven by both inventory risk and asymmetric information. The model is similar to Kyle (1989), with the main simplification of absence of heterogeneity among strategic traders (both in terms of information and risk aversion) and the main generalization of allowing for non-normal payoffs.

In equilibrium, traders determine their optimal demand function knowing the demand functions of all other traders. I show that the optimization problem is equivalent to traders not knowing others' demand functions but knowing their own price impact (i.e., how their trade moves the price at the margin) for each order

[3] By definition, the cumulant generating function (CGF) is a log of a moment generating function of the distribution. Given the CGF $g(x)$ the risk function with a parameter a is defined as $\rho_a(x) = \int_0^1 g''(-t^{1-a}\gamma x)dt$.

size. This is an intuitive representation of the problem: real-world traders typically have a market impact model that is an input in their optimal execution algorithm. The equilibrium price impact function is pinned down by the requirement that it is consistent with the demand functions of the other traders. The consistency requirement yields a linear ordinary differential equation (ODE) that I use to compute the equilibrium price impact function in closed form for any probability distribution. I show that the properties of the price impact function can be derived by those of a risk function, which summarizes the probability distribution of the asset payoffs. I derive main properties of a risk function, in particular the ones related to comparative statics, and believe that these can be useful for future research.

I show that the ODE for the price impact function can have a continuum of solutions, even with a normal distribution. Thus, there is a continuum of equilibria. Equilibrium non-uniqueness can be attributed to the following complementarity: if strategic traders believe that the price impact is high, they provide less liquidity, which confirms higher equilibrium price impact. In Glebkin, Rostek and Yoon (2015) we study equilibrium uniqueness in demand functions. Applying the results from the latter paper, a unique equilibrium with bounded payoff can be pinned down by requiring prices to lie within asset payoff bounds. This requirement is intuitive: if the price is outside payoff bounds the block trader gets negative profit with certainty and hence should not trade. The requirement rules out equilibria in which CARA traders' price impact is too high, selecting the unique equilibrium. With unbounded payoff the equilibrium can be selected by requiring the prices to be close to that of an asset with an arbitrary close, but bounded payoff. This equilibrium corresponds to the linear one under normality.

Using the characterization of the price impact function, I examine the relationship between price and order size. When asset payoffs are positively skewed, small purchases have greater price impact compared to small sells. The intuition for the result can be seen by contrasting with the benchmark case where the asset payoff is normally distributed and hence the skewness is zero. Consider first sells by the block trader. The trader's counterparties, who buy from him, receive a positively skewed profit, which they like. Intuitively, positive skewness implies that positive surprises to profits are more likely than negative ones. As a result, traders require a lower premium for providing liquidity and the price reaction to the order is smaller than in the benchmark case. For purchases, the trader's counterparties, who sell to him, receive a negatively skewed profit and require a greater premium. The price reaction is greater than in the benchmark case. Consequently, with positive skewness small purchases have greater price impact compared to small sells. Similarly, when asset payoffs are negatively skewed sells have greater price impact compared to purchases.

The concavity of the price reaction function (absolute value of price change as a function of order size) for large orders arises when asset payoffs are bounded. If asset payoff is bounded below, e.g., by zero, the sell order cannot push the price below zero. Consequently, the price reaction function for sells is bounded above, which rules out convex shapes. Assuming a further mild restriction on payoff distribution I show that the price reaction function is concave for large sell orders.[4] Similarly, if payoff distribution is bounded above and satisfies the same restriction, the price reaction function is concave for large buy orders.

With positive skewness the price reaction function is concave for small sells and convex for small purchases. As noted above, with positive skewness the price reaction for purchases is smaller than that in a benchmark CARA-normal case, in which the price reaction linear. This "smaller than linear" price reaction generates a concave shape. For purchases, the price response is greater than that in a benchmark case, which generates convexity. Similarly, with negative skewness, the price reaction function is concave for small purchases and convex for small sells.

Summarizing, the baseline model can speak to two key empirical findings concerning the shape of the price impact: the asymmetry and the concavity of the price reaction function. With positively skewed asset payoffs the model predicts that price impact of (small) purchases is greater than that of (small) sells, which is in line with the evidence summarized by Saar (2001). With bounded asset payoffs the model predicts concave price reaction function for large orders, consistent with the evidence in Hausman et al. (1992), Almgren et al. (2005) and Frazzini et al. (2015). I also derive new predictions. The model implies that difference in curvatures of price reaction function for small purchases and sells is positive (negative) with positive (negative) skewness. This prediction can be tested in equities market: individual stocks have positively skewed returns, while returns on stock indices are negatively skewed (e.g. Chen et al. (2001)). The effects of payoff bounds can be examined in

[4]The restriction is that the third derivative of the CGF, $g'''(x)$, does not change sign for large enough x. In the numerical simulations (as well as analytic calculations, when direct calculation of CGF is possible) I was unable to find an example of a finite-support distribution for which this condition does not hold. Consequently, I believe that this restriction is mild.

options market. Payoffs of puts, unlike that of calls are bounded above. The model suggests that the price reaction to purchases should be more concave for puts rather than calls.

In the extension of the model I show that when the block trader possesses private information regarding the asset payoff, the price impact can be separated into an inventory risk part and asymmetric information one. The bounds of the asset payoff play the same role as in the case of no asymmetric information. Unlike that case, however, there are additional determinants of the curvature of price reaction function for small orders. In particular, the shape of the conditional moments (mean and variance of asset payoff conditional on order size of the block trader) as functions of order size also plays a role. For example, if (1) skewness is positive, (2) the conditional mean function is convex and (3) the conditional variance is decreasing, both inventory and asymmetric information components of the price reaction function are convex (concave) functions of order size for small buy (sell) orders. One would not capture these effects in the jointly normal setting because, in that case, the conditional mean is linear and the conditional variance is a constant.

In the extended setting, I also show that the slope of the information component of the price reaction function is diagnostic of the degree of informed trading. This slope is proportional to the slope of the conditional mean function. The latter shows how a marginal unit liquidated or purchased by a block trader affects expectation of liquidity providers regarding asset payoffs. Therefore, a higher slope of the price reaction function indicates a more informed block trader.

This paper is related to three broad strands of the literature: strategic trading, divisible good auctions, and models of asset trading without normality.

The literature on strategic trading dates back to Kyle (1985). The models typically feature asymmetric information and rely on CARA-normal assumptions for tractability.[5] The equilibria in those models are linear, which results in liquidity measures being linear in the trade size.

The representation of equilibrium in this paper and the corresponding intuition builds on the result of Rostek and Weretka (2015), who are the first to show that the Nash equilibrium in demand submission games can be represented through two conditions: the optimality of the bid given a price impact model and the consistency

[5] An incomplete list includes: Kyle (1985, 1989), Subrahmanyam (1991), Douglas and Viswanathan (1996), Vayanos (1999, 2001), DeMarzo and Urosevic (2006), Rostek and Weretka (2015). See also Brunnermeier (2001) or Biais et al. (2005) for a review.

of the model. A related paper by Weretka (2011) defines a new equilibrium concept in which traders are not price takers but slope takers. The consistency in that paper is a part of that equilibrium concept. In both Rostek and Weretka (2015) and Weretka (2011), however, the price impact is a constant, which is due to normality in the first paper and is a part of the equilibrium concept in the second paper.

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

The divisible good auctions literature commonly examines the two most popular auction formats, namely the discriminatory-price auction (DPA) and the uniform-price auction (UPA), that are used to distribute divisible goods such as government debt, electricity, spectrum and emission permits.[6] The key contribution of my paper relative to this literature is to examine nonlinear equilibrium in UPA, solve for equilibrium in closed form and to link the nonlinearities to the properties of asset payoff distribution.

Wang and Zender (2002) consider both UPA and DPA formats in a CARA-normal setting with random supply of the asset. They show that there is a continuum of equilibria when uniform-price auction format is adopted. My non-uniqueness result is thus related to the one in that paper. My main contribution relative to Wang and Zender (2002) is to consider a more general setting allowing, in particular, the asset payoff to be bounded. The latter enables me to use the selection argument resulting in a unique equilibrium.

DeMarzo and Skiadas (1998, 1999), Breon-Drish (2015), Chabakauri et al. (2015) and Albagli, Hellwig, and Tsyvinski (2011) study REE models without assuming normality. Palvolgyi and Venter (2015) demonstrate the existence of non-linear equilibria with discontinuous price function in a standard CARA-normal REE model a la

[6] Such papers include Wilson (1979), Klemperer and Meyer (1989), Wang and Zender (2002), Kremer and Nyborg (2004), Rostek and Weretka (2012) and Ausubel et al (2014). The papers focusing specifically on finance applications, such as Kyle (1989), Vayanos (1999) and Rostek and Weretka (2015) (which I mentioned in the strategic trading literature) also belong to this literature.

Grossman and Stiglitz (1980) and Hellwig (1980). The traders in those papers are competitive. Consequently, these papers abstract from price impact, which is the focus of my paper. Rochet and Vila (1994) analyze a model a la Kyle (1985) without normality and prove uniqueness of equilibrium.[7] Their uniqueness result is in contrast to the multiplicity result in my paper, which extends the Kyle (1989) model. Moreover, Rochet and Vila (1994) do not study how departures from normality affect the shapes of the price impact, which I do. Biais et al. (2000), Baruch (2005) and Back and Baruch (2013) study strategic liquidity provision without assuming normality. However, the liquidity providers in their models are risk-neutral. Consequently, there is no inventory risk, which is the main focus of my paper. Martin (2013) introduces the language of cumulant generating functions (CGF) in a Lucas tree asset pricing model with general (not log-normal) i.i.d. consumption growth. This paper applies the language of CGFs and demonstrates that it is useful in a strategic setting.

Finally, my paper speaks to a literature on optimal optimal dynamic execution algorithms under exogenous and non-constant price impact (see, e.g., Almgren et al. (2005)).[8] My paper complements this literature by providing equilibrium foundations for non-linear price functions.

The remainder of the paper is organized as follows. Section 2 presents the model, and Section 3 solves for equilibrium. Section 4 derives the model's predictions about the price impact measures and solves the model for specific probability distributions. Section 5 extends the model to the case where the block trader is informed. Section 6 links the theory to empirical evidence. Section 7 concludes. All proofs are in Appendix A. Appendix B collects the main properties of the risk functions.

1.2 The model

There are two time periods $t \in \{0, 1\}$; two assets, a stock and a bond; and $L > 2$ large traders. The bond earns zero net interest, without loss of generality, as the bond can be chosen as a numeraire. The stock is a claim on a terminal dividend δ

[7]Boulatov and Bernhardt (2015) demonstrate the uniqueness of a *robust* equilibrium in a Kyle (1983) paper. They also derive, using a different technique, an ODE for price impact in their model. Their model features risk-neutral market makers and normally distributed asset payoff.

[8]Other papers on optimal dynamics execution algorithms with exogenous price impact, which can be linear or non-linear, include Bertsimas and Lo (1998), Almgren and Chriss (2001), Huberman and Stanzl (2005) and Obizhaeva and Wang (2012)

characterized by the *cumulant generating function (CGF)*

$$g(x) \equiv \log E[\exp(x\delta)].$$

I assume that the CGF exists for all $x \in \mathbb{R}$.

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

$$g'(0) = E[\delta] \equiv \mu,\ g''(0) = Var[\delta] \equiv \sigma^2,$$

$$g'''(0) = \sigma^3 \cdot \text{skewness},\ g^{(4)}(0) = \sigma^4 \cdot \text{excess kurtosis}. \tag{1.2.1}$$

The stock is in price-inelastic supply s, which is uncertain for large traders. The supply s has a full support and is independent of a terminal dividend δ.[9] I interpret the supply as being provided by a block trader who trades with a market order. In this interpretation, the uncertainty about s is due to the uncertainty about the identity of the block trader. The uncertainty of the supply is important because without it there is a dramatic multiplicity of equilibria, as is common in the literature (see, e.g., Klemperer and Meyer (1989), Vayanos (1999)). The independence of the supply s and the terminal dividend δ implies that there is no information to be learned from observing s. With uninformative supply, its distribution does not affect the equilibrium. The independence assumption is relaxed in the section 5.

The large traders are identical and maximize expected utility from their terminal wealth W. I refer to a problem solved by the traders as problem $\mathcal{P}$

$$\max_{x(p)} E_{\delta,s}[-\exp(-\gamma W)],$$

$$s.t.: W = \hat{x}\delta - \hat{p}(\cdot)\hat{x}.$$

As can be seen from the above, all traders have CARA utility with risk aversion γ. They have no initial inventories of stocks or bonds.[10] Their strategy is a bid (demand schedule) $x(p)$: the quantity of the risky asset they wish to buy ($x > 0$) or sell ($x < 0$) at a price p. Traders are *large* in the sense that they can affect market clearing prices $\hat{p}(\cdot)$ and account for this effect. I use the notation $\hat{p}(\cdot)$ (not just $\hat{p}$)

[9] In what follows, I will abuse notation and denote the random variable and its realization by the same letter. E.g., s may denote the random supply or its particular realization.

[10] This assumption is made to simplify the exposition. One can easily relax it as long as initial inventories are symmetric across traders.

to emphasize the latter fact.[11]

The trading mechanism is a uniform-price double auction. I denote the outcomes of the auction (market clearing price and allocations) by hat. The quantity $\hat{x}$ allocated to a trader is his bid evaluated at the market clearing price $\hat{p}(\cdot)$, $\hat{x} = x(\hat{p}(\cdot))$. The market clearing price $\hat{p}(\cdot)$ is determined as follows. Given the bids of the traders $(x_i(p))_{i=\overline{1,L}}$ and a particular realization of s, the equilibrium price $\hat{p}(\cdot)$ is the one that clears the market

$$\sum_{i=1}^{L} x_i(p) = s.$$

If there are several market clearing prices, then the smallest of them is chosen. If there are no market clearing prices, then there is excess demand at all prices. The price is set to $p = \infty$ if the excess demand is positive and to $p = -\infty$ if it is negative. Such rules are the same as in Kyle (1989) and provide well-defined prices for all possible strategies of the traders. In equilibrium, the bids will be such that the finite market clearing price always exists and is unique.

Throughout the paper, I use the following notation. The *certainty equivalent of a position y in the risky asset* is denoted by $f(y)$. By definition, $f(y)$ solves

$$\exp(-\gamma f(y)) \equiv E_\delta[\exp(-\gamma y \delta)].$$

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

$$f(y) = -\frac{1}{\gamma} g(-\gamma y). \tag{1.2.2}$$

It can also be shown that the function $f(y)$ is strictly concave.[12]

I also introduce the *index of imperfect competition* $\nu \in (0; 1]$:

$$\nu = \frac{1}{L-2} \in (0; 1].$$

Higher values of the index correspond to a less competitive market, and the limiting case $\nu \to 0$ corresponds to perfect competition.

[11]To be more rigorous, $\hat{p}(\cdot)$ is not a scalar (as would be implied by notation $\hat{p}$) but a functional that maps bids of all traders onto market clearing prices.

[12]See, for example, Gromb and Vayanos (2002), Lemma 1. It also follows from a strict convexity of a cumulant-generating function (e.g., Billingsley (1995)).

1.3 Equilibrium

The equilibrium concept is a symmetric Nash equilibrium (simply *equilibrium* in what follows) and is formally defined below.

Definition 1. A bid $x(p)$ is a symmetric Nash equilibrium if for any $i = 1, 2, ..L$, given that the traders $j \neq i$ submit bids $x_j(p) = x(p)$, it is optimal for trader i to submit bid $x_i(p) = x(p)$.

I will focus on equilibria in which the bids are continuously differentiable. Before proceeding to equilibrium characterization, I define several objects that will be used in the following parts of the paper.

1.3.1 Residual supply, ex post maximization and price impact

Given other traders' equilibrium bids $(x_j(p))_{j\neq i}$ for a given realization of s, rewrite the market clearing condition as

$$x_i(p) = s - \sum_{j\neq i} x_j(p). \tag{1.3.1}$$

The right-hand side of the equation above,

$$R_i(p; s) = s - \sum_{j\neq i} x_j(p), \tag{1.3.2}$$

is called the *residual supply.* The subscript i indicates that this is a residual supply faced by the trader i.[13] The inverse function, $P_i(x; s)$, solving

$$x = s - \sum_{j\neq i} b_j(P_i(x; s)), \tag{1.3.3}$$

is called the *inverse residual supply.*[14] The locus

$$\mathcal{C}_i(s) = \left\{ (x, p) | \, x = s - \sum_{j\neq i} x_j(p) \right\}$$

[13]Because the equilibrium is symmetric, I will often omit the subscript i for the residual supply when it does not cause confusion.

[14]The equilibrium bids are shown to be strictly decreasing. Therefore, the inverse functions in this section are well defined.

is called the *residual supply curve.*

As follows from (1.3.1), from the perspective of a particular trader, the market clearing price and quantity are determined by the intersection of his bid and the residual supply curve. Each realization of s corresponds to a horizontal parallel shift of the residual supply curve $\mathcal{C}(s)$. There is a price-quantity pair $M^*(s) = (x^*(s), p^*(s))$ on $\mathcal{C}(s)$ that maximizes the utility of a trader. I call the point $M^*(s)$ the *optimal* point. If there exists a bid that intersects with each realization of the residual supply at the optimal point, then such a bid is ex post optimal, i.e., it produces optimal price and quantities *for any* realization of s . A bid given parametrically (as a function of s) by $(x^*(s), p^*(s))$ is clearly such a bid. This bid is clearly optimal ex ante and solves problem $\mathcal{P}$.

The process of finding the equilibrium bid can therefore be simplified to finding the locus of ex post optimal points $(x^*(s), p^*(s))$ corresponding to different realizations of s. The locus $(x^*(s), p^*(s))$ is a parametric representation of the bid. The latter problem, which I call *ex post optimization* and denote $\mathcal{P}_{EP}$, can be written as[15]

$$(x^*(s), p^*(s)) = \arg\max_{x,p} \{f(x) - p \cdot x\}, \tag{1.3.4}$$

$$\text{s.t.: } (x, p) \in \mathcal{C}(s).$$

The optimal point maximizes the certainty equivalent of a position x in risky asset $f(x)$ minus the costs of reaching this position, $p \cdot x$, on a given residual supply realization. Figure 1.3.1 provides an illustration of the ex post maximization procedure outlined above.

Apart from being helpful in ex post maximization, the residual supply is also a useful object because it allows one to define the *price impact.* I define it analogously to Kyle's lambda: it is the slope of the equilibrium inverse residual supply evaluated at the quantity allocated to the trader in equilibrium. More precisely, consider a profile of equilibrium bids of the traders. Denote by $s_i(x^*)$ the residual supply realization such that a trader i is allocated x^* given the profile of equilibrium bids (in the symmetric equilibrium, $s_i(x^*) = Lx^*$). The price impact is by definition the slope of $P_i(x, s_i(x^*))$ evaluated at x^*

[15] The idea of finding the equilibrium in games in which players submit demand schedules by means of ex post optimization is not new. Klemperer and Meyer (1989) and Kyle (1989) were among the first to develop it.

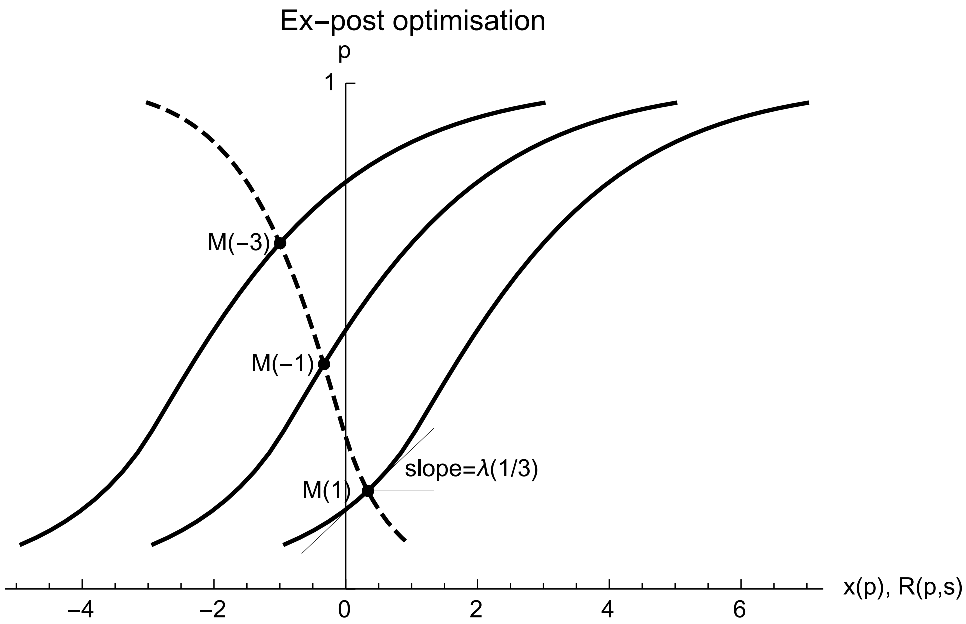

Figure 1.3.1: The figure shows realizations of the residual supply curve corresponding to the realizations of supply $s \in \{-3, -1, 1\}$ (thick black solid lines). On each curve, there is an optimal point $M(s)$ that is marked explicitly in the figure. The set of such optimal points represents the equilibrium bid (dashed line). The figure also shows the price impact at point $x = 1/3$: it is a slope of the inverse residual supply corresponding to $s = 1$ ($L = 3$ for the figure above).

$$\lambda_i\left(x^*\right) \equiv \left.\frac{\partial P_i\left(x, s_i\left(x^*\right)\right)}{\partial x}\right|_{x=x^*}. \tag{1.3.5}$$

I show that the equilibrium bids are strictly decreasing. This implies that the inverse residual supply is strictly increasing and there is always only one point at which it intersects with the equilibrium bid. Therefore, for a given x^*, there is only one realization of the inverse residual supply that intersects with the bid at x^* and only one corresponding value of the slope $\left.\frac{\partial P(x,s(x^*))}{\partial x}\right|_{x=x^*}$. Therefore, λ is a well-defined object: it is *a function* of x^* , and it *depends on* x^* *only* (and does not depend, for example, on s).

The price impact shows the equilibrium price sensitivity. For a given realization of s (producing x^* for a trader of interest), holding the bids of other traders fixed, if the trader of interest modifies his bid in a way that he is allocated $x^* + \Delta$ instead of x^* , the price will change by $\lambda(x^*) \cdot \Delta$.

1.3.2 Characterization of the equilibrium

I first derive the equilibrium characterization heuristically to show the intuition and then justify the derivation in the Theorem 1 below.

Consider first a competitive (price-taking) trader and his equilibrium bid. He solves

$$\max_{x} f(x) - px. \tag{1.3.6}$$

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

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

A strategic trader accounts for the fact that he can move prices, and his first-order condition will have a new term that is due to price impact

$$f'(x) - \frac{\partial p}{\partial x} x = p. \tag{1.3.7}$$

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

Suppose that each trader has a *model* $l(x) = \frac{\partial p}{\partial x}$ of his price impact. This function shows how much, at the margin, the trader of interest can move prices if he trades x. This model, together with first-order condition (1.3.7), determines his optimal (inverse) bid

$$f'(x) - l(x)x = p. \tag{1.3.8}$$

In a Nash equilibrium, the models cannot be arbitrary. The way to pin down the equilibrium price impact model is to require it to be *consistent* with equilibrium demands of the other traders.The *consistency* condition requires assumed price impact $l(x)$ to be equal to the equilibrium one, the slope of the equilibrium inverse residual supply, i.e., $l(x) = \lambda(x)$. Intuitively, an inconsistent price sensitivity model will produce suboptimal bids and, therefore, cannot be an equilibrium because the traders will have incentives to deviate.

Consistency implies an ODE for price impact function. In the symmetric equilibrium, there are $(L-1)$ identical bids contributing to the slope of the residual supply. The slope of the residual supply is thus

$$\text{slope of the residual supply} = -(L-1)\frac{1}{I'(x)}.$$

The minus is to account for the fact that the residual supply is upward-sloping while the bid is downward-sloping; I also exploit the fact that the slope of the bid is the reciprocal of the slope of its inverse. The slope of the *inverse* residual supply is the reciprocal of the above; therefore,

$$\lambda(x) = \frac{-1}{L-1} I'(x). \tag{1.3.9}$$

The inverse bid is given by (1.3.8), which, accounting for the consistency condition $l(x) = \lambda(x)$ becomes $I(x) = f'(x) - \lambda(x)x$. The slope of the inverse bid is thus

$$I'(x) = f''(x) - \lambda'(x)x - \lambda(x).$$

Combining the above and (1.3.9) yields the ODE

$$x\lambda'(x) = (L-2)\lambda(x) + f''(x). \tag{1.3.10}$$

The equilibrium bid should therefore satisfy two conditions: the *optimality* (equation (1.3.8)) and *consistency* of the model (equation (1.3.10)). The third condition was used implicitly: for the inverse bid and the inverse residual supply to exist, the bids should be *monotone*. The analysis in Appendix C shows that monotonicity should hold in equilibrium, as the second-order conditions are violated otherwise. The theorem below summarizes and justifies the above heuristic derivation.

Theorem 1. *The bid $b(p)$ is an equilibrium if, and only if, it satisfies the following three conditions:*

(1) The bid $b(p)$ is ***optimal*** *given a model $l(x)$ of price impact*

$$f'(x) - xl(x) = p \Rightarrow x = b(p). \tag{1.3.11}$$

(2) The assumed models of price impact are ***consistent****, i.e., $l(x) = \lambda(x)$. The latter condition is equivalent to the following ODE*

$$x\lambda'(x) = (L-2)\lambda(x) + f''(x). \tag{1.3.12}$$

(3) ***Monotonicity:*** $0 < \lambda(x) < \infty \, \forall x.$

I follow Rostek and Weretka (2015) in using the representation of equilibrium through optimality and consistency conditions because it captures nicely the decision-making of real traders. As in real life, traders have a market impact model and determine optimal bids (do the optimal trade execution) given this model. In a Nash equilibrium, the price impact model is specified by the consistency condition. The main contribution relative to Rostek and Weretka (2014) is to derive the consistency condition when price impact is a function, not a constant. The fixed point condition then results in an ODE, not an algebraic equation.

The above representation is also very useful for solving the model. To find an equilibrium bid, one has to find the equilibrium price impact, which satisfies a linear ODE (1.3.12). The latter is easy to solve with standard methods.

1.3.3 Solution

In this section, I provide the solution to the model. I solve for the equilibrium strategies of traders, their bids, and the corresponding equilibrium price impacts. Before doing so, I introduce the new object that will be helpful in the analysis that follows.

Risk function

I introduce the following transformation of the cumulant-generating function, which I call the *risk function.* The risk function with a parameter a is given by

$$\rho_a(x) = \int_0^1 g''(-t^{1-a}\gamma x)dt.$$

This function summarizes the relevant risk inherent in the distribution of the terminal payoff δ when the trader changes his position in the risky asset from zero to x.

As can be seen from the above, for the normal distribution for which all risk is summarized by the variance (σ^2), the risk function is equal to σ^2. It is also clear that the risk function evaluated at zero is equal to variance

$$\rho_a(0) = g''(0) = \sigma^2.$$

However, the main justification for the function $\rho_a(x)$ as a measure of risk comes, of course, from the equilibrium. We will see that the risk function appears in the expressions for equilibrium objects in a general case where variance is present in a CARA-normal case. See the remark below.

Remark 1. We will see that the results under a general distribution can be obtained from a CARA-normal case by substituting $\rho_a(x)$ instead of σ^2 as a measure of risk. The value of parameter a differs depending on where the risk function appears. The risk contributing to price impact function (arising due to strategic interactions) is measured by the risk function with the parameter a related to the degree of competition in the economy, $a = 1 + \nu$. In all other cases, the parameter a is equal to zero. In other words, one can obtain the results under the general distribution by substituting $\rho_{1+\nu}(x)$ instead of σ^2 when computing the price impact function and substituting $\rho_0(x)$ instead of σ^2 in all other cases. The shapes of liquidity measures and the important comparative static results will be determined by the properties of risk functions that I present in Appendix 1.9.

CARA-normal benchmark

I begin by investigating the CARA-normal version of my model. The results for this case are well known, and the following corollary to Theorem 1 summarizes them.

Corollary 1. *Suppose $\delta \sim N(\mu, \sigma^2)$. There exists a linear equilibrium, given by*

$$\lambda = \nu\gamma\sigma^2; \tag{1.3.13}$$

$$I(x) = \mu - \gamma\sigma^2 x - \lambda x. \tag{1.3.14}$$

Proof. In a linear equilibrium, the price impact is constant. Plugging $\lambda' = 0$ into (1.3.12) yields that $\lambda = -\nu f''(x)$. For a normal distribution, $f''(x) = -\gamma\sigma^2$, and hence (1.3.13) and (1.3.14) obtains. □

Equation (1.3.13) demonstrates that there are three sources generating the illiquidity (λ): imperfect competition (ν), limited risk-bearing capacity (γ) and the riskiness of the asset (σ^2). The higher the risk, the lower the risk-bearing capacity; the lower the competition, the higher the price impact.

Further observe that relative to the competitive case (for which $\nu = 0$), the traders reduce their bids: they bid smaller quantities for a given price. This is a consequence of imperfect competition among traders and is a standard result for divisible good auctions (see, e.g., Ausubel et al. (2014)).

General case

To construct an equilibrium in the general case, I divide the problem into two. I first find the equilibrium price impact. Once it is found, it is easy to find the equilibrium bid from the first-order condition (1.3.11).

According to Theorem 1, a function $\lambda(x)$ is an equilibrium price impact if and only if it satisfies the ODE (1.3.12) and $0 < \lambda(x) < \infty$. The equation (1.3.12) is easy to analyze: it is a linear ODE.

For $x \neq 0$, multiply both parts of equation (1.3.12) by the integrating factor x^{1-L} and rearrange to obtain

$$\left(x^{2-L}\lambda(x)\right)' = f''(x)x^{1-L}. \tag{1.3.15}$$

Integrating the above between x and x_0, where $x > x_0 > 0$, indicates that the solution can be written as

$$\lambda(x) = x^{L-2}\left(\lambda(x_0)x_0^{L-2} + \int_{x_0}^{x} f''(t)t^{1-L}dt\right). \tag{1.3.16}$$

An analogous equation can be written for the case in which $x < x_0 < 0$. Any solution, corresponding to different boundary conditions $\lambda(x_0)$, such that $0 < \lambda(x) < \infty$, will be an equilibrium price impact.

Two things are evident from equation (1.3.16). First, there might be many equilibrium price impact functions, as there might be many boundary conditions $\lambda(x_0)$ such that $0 < \lambda(x) < \infty$. Second, for the price impact function to exist for all real x (in particular, at infinity), some technical conditions regarding the certainty equivalent $f(x)$, or, equivalently, on CGF $g(x)$ have to be met for the integral $\int_{x_0}^{x} f''(t)t^{1-L}dt$ to converge as $x \to \infty$. Those technical conditions yield the restrictions on the CGF that need to be satisfied for the equilibrium to exist. I discuss the multiplicity and the technical conditions needed for the existence of equilibrium in a greater detail in Remarks 2 and 3. Proposition 1 below summarizes the solution.

Proposition 1. *(1) The equilibrium exists if and only if* $\rho_{1+\nu}(x) < \infty$. *In particular, it exists for any distribution with bounded support.*

(2) The equilibrium price impact function is given by

$$\lambda(x) = \nu\gamma\rho_{1+\nu}(x) + \mathbb{I}(x \geq 0)C^{+}x^{L-2} + \mathbb{I}(x < 0)C^{-} \cdot (-x)^{L-2},$$

for any $C^{+}, C^{-} \geq 0$.

(3) The equilibrium inverse bid is given by

$$I(x) = \mu - \gamma\rho_0(x)x - \lambda(x)x. \tag{1.3.17}$$

I comment on the technical conditions that need to be satisfied for the equilibrium to exist and on the multiplicity in the two remarks below.

Remark 2. According to Proposition 1, the equilibrium exists if and only if the risk function $\rho_{1+\nu}(x)$ is finite for every finite x. After some algebra, the latter function can be written as

$$\rho_{1+\nu}(x) = (L-2)\int_1^{\infty} g''(-\gamma y x)y^{1-L}dy.$$

For the integral to converge, the second derivative of the CGF should grow not too fast as $y \to \pm\infty$. An example of the distribution for which the equilibrium does not exist is a Poisson distribution for which the CGF and its second derivative grow exponentially.

However, given that equilibria exist for any distribution with bounded support, I believe that the technical conditions are not too restrictive. Indeed, in the real world, the payoff cannot be unbounded. There is a lower bound, which is due to limited liability, and there is an upper bound, which is due to the fact that the resources of are limited (and hence an asset cannot have an infinite payoff). It should also be noted that the technical conditions hold also for distributions with unbounded support: examples include (and are not restricted to) normal (the benchmark) and the mixture of normal distributions.

Remark 3. Despite the presence of uncertainty regarding the supply, there is nevertheless a multiplicity of equilibria. The mechanism behind the multiplicity is the following.

Technically, according to Theorem 1, the equilibrium lambda is a price impact model that is consistent. The consistency condition is equivalent to ODE (1.3.12). The standard result from the theory of differential equations implies that having obtained a solution to a linear ODE (1.3.12) another can be obtained by adding a solution of an homogenous ODE, in this case $x\lambda'(x) = (L-2)\lambda(x)$. The latter solution is given by $\mathtt{1}(x \geq 0)C^+x^{L-2} + \mathtt{1}(x < 0)C^-(-x)^{L-2}$. After filtering out solutions not satisfying condition (3) of Theorem 1, we are left with $C^+, C^- \geq 0$.

In economic terms, the multiplicity can be attributed to the following complementarity: if a trader believes that the price impact is high, he provides less liquidity, which implies a higher price impact for other traders. This, in turn, induces them to provide less liquidity, confirming the higher price impact for a trader of interest.

Note that the multiplicity is present even in the standard CARA-normal case. The multiplicity result in a CARA-normal case is known in the divisible goods auctions literature (see Wang and Zender (2002), Proposition 3.4). The non-uniqueness of equilibrium arises for a similar reason in Bhattacharya and Spiegel (1991). However, in the generalized setting of my paper a selection argument, exploiting the generality of distribution is possible.

In Glebkin, Rostek and Yoon (2015) we study the uniqueness of equilibrium in supply functions. Parts (1) and (2) of the following Proposition is from there and are included here for completeness. The Proposition demonstrates that there is only one equilibrium satisfying the following natural properties.

Proposition 2. *The equilibrium with $C^+ = C^- = 0$ is the unique equilibrium having the following properties:*

(1) Suppose that δ has a bounded support (a, b). The equilibrium with $C^+ = C^- = 0$ is the unique equilibrium with equilibrium prices within $[a, b]$.

(2) Suppose that δ has infinite support. Consider an asset with a payoff $\delta_n = \delta \cdot \mathbb{I}(\delta \in (a_n, b_n))$, where a_n and b_n are finite. Denote $p_n(s)$ the price of the asset that pays δ_n when the supply realization is s in the unique equilibrium in which $p_n(s)$ is within (a_n, b_n) for all s. Then $p_n(s)$ converges pointwise to $p(s)$ as $a_n \to -\infty$ and $b_n \to \infty$, where $p(s)$ is a price of the asset that pays δ in the equilibrium with $C^+ = C^- = 0$.

(3) The equilibrium with $C^+ = C^- = 0$ is the unique equilibrium in which $\lambda(s)$ converges to zero pointwise as either $\nu \to 0$ or $\gamma \to 0$.

In the CARA-normal model, the equilibrium with $C^+ = C^- = 0$ corresponds to the linear equilibrium of corollary 1.

The first property is intuitive: if the price is outside payoff bounds the block trader gets negative profit with certainty and hence should not trade. Block trader is unmodeled and is assumed to submit a price-inelastic bid: he is willing to pay any price. However, the latter is only consistent with rational behavior if prices outside the bounds of payoff never realize in equilibrium. Another piece of intuition is as follows. By assumption, there is a full support uncertainty regarding the supply, i.e., any quantity may be supplied with positive probability. Suppose, for example, that for some supply realization s the price is outside the payoff bounds. Then, the trader providing the supply will realize a negative profit with certainty, for any realization of δ. However, it is then not consistent to assume that quantity s may be supplied with positive probability.

The second property extends the first one to the case of infinite support. It highlights that price of an asset with arbitrary close, but bounded payoffs (for which the price can be uniquely pinned down requiring the property (1) to hold) is close to the price of an asset with unbounded payoff in the equilibrium with $C^+ = C^- = 0$.

The third property highlights that price impact vanishes as the sources of illiquidity disappear (the economy becomes perfectly competitive, $\nu \to 0$; or risk-bearing capacity becomes infinite, $\gamma \to 0$) only in the equilibrium with $C^+ = C^- = 0$.

Henceforth, I focus on the equilibrium with $C^+ = C^- = 0$ (and will simply refer to it as *equilibrium* hereafter). Although other equilibria may be of theoretical interest, they appear to have little empirical relevance. Indeed, unbounded prices are not observed for assets with bounded payoffs (e.g., negative prices on limited liability assets), and there is empirical evidence that competition reduces illiquidity (e.g., Kagel and Levin (2001)).

1.4 The shape of the price impact

The purpose of this section is to study how the prices in the imperfectly competitive market may be affected by a block sell or buy order s. This is measured by several liquidity measures that I describe below.

I will use the notation $\bar{s}$ for per capita supply

$$\bar{s} \equiv s/L.$$

A positive $\bar{s}$ corresponds to a sell order, whereas a negative $\bar{s}$ corresponds to a buy order.

Denote $p(s)$ as the equilibrium price when the supply is s. The function $p(s)$ is easy to find. Because each large trader is allocated s/L in the symmetric equilibrium, the equilibrium price can be determined from condition (1.3.11), which using the risk function, can be written as (cf. (1.3.17))

$$p(s) = \mu - \gamma\rho_0(\bar{s}) - \lambda(\bar{s})\bar{s}. \tag{1.4.1}$$

The price reaction function $\Pi(s) = |p(s)-\mu|$ is the absolute value of the difference between the equilibrium price when the supply is s and the equilibrium price when the supply is zero (which is equal to μ). It measures the total reaction of the equilibrium price to a block of size s. Equation (1.4.1) yields

$$\Pi(s) = \gamma\rho_0(\bar{s})|\bar{s}| + \lambda(\bar{s})|\bar{s}|. \tag{1.4.2}$$

The price reaction function can be decomposed into two parts. The first

$$\tau(\bar{s}) \equiv \lambda(\bar{s})|\bar{s}| = \nu\gamma\rho_{1+\nu}(\bar{s})|\bar{s}|, \tag{1.4.3}$$

arises due to strategic interactions. I will refer to it as a *strategic component* of price reaction function. It is (the absolute value of) the difference between price $p(s)$ and price $f'(\bar{s})$ that a competitive trader would pay.[16] In the limit, $\nu \to 0$, $s/L = const = \bar{s}$, corresponding to perfect competition, this strategic component disappears.

The second component

$$\pi(s) \equiv \gamma\rho_0(\bar{s})|\bar{s}|, \tag{1.4.4}$$

Is non-strategic. It is the equal to $|f'(s/L) - \mu|$, the price reaction in a perfectly competitive economy. I will refer to it as a *non-strategic component* of price reaction function.

[16]It is easy to show that $f'(\bar{s}) = \mu - \gamma\rho_0(\bar{s})$.

In a dynamic CARA-normal model, Rostek and Weretka (2014) demonstrate that the component of the price reaction due to strategic interactions is transitory. This is intuitive: the strategic component in a particular period (say t) arises because the traders account for their price impact in this particular period. The extent to which a trader can move prices in period t does not matter for him in periods $t+1$ and so on because the price at t is already realized. Therefore, there will be a price effect at time t and no effect at times $t+1$ and so forth. They also show that the non-strategic component, in contrast, is permanent. This is also intuitive. If a block trader sells in period t, he increases the inventory of traders who absorb this trade. They have greater inventories in periods t,$t+1,..$ and, being risk averse, require a price discount in those periods. The price effect lasts for subsequent periods and is thus permanent.

Given the above, one can interpret the strategic component $\tau(\bar{s})$ as a temporary price effect and the non-strategic component $\pi(\bar{s})$ as a permanent price effect. See also the remark below.

Remark 4. One can also justify the difference between the temporary and the permanent price effects as follows. I allow traders to trade in two additional periods, $t=-1/2$ and $t=1/2$. However, given the complexity of a dynamic problem without normality, I assume that in each period traders behave myopically and do not foresee the possibility to trade in the future. There is a block sale s at $t=0$, and the supply at $t=-1/2$, $1/2$ is zero. As before, the traders consume only at time $t=1$.

At time $t=-1/2$, there will be no trading, as traders are symmetric and there is no supply provided. The price at which the traders having no initial endowments will neither buy nor sell is

$$p_{-1/2}=\mu.$$

The price at $t=0$ is already found and is given by

$$p_0=\mu-\gamma\rho_0(\bar{s})\bar{s}-\lambda(\bar{s})\bar{s}.$$

At time $t=1/2$, there will also be no trading due to symmetry and the absence of supply. However, each trader now has an endowment $\bar{s}$; therefore, the price at which traders will be indifferent between buying and selling is

$$p_{1/2} = f'(\bar{s}) = \mu - \gamma\rho_0(\bar{s})\bar{s}.$$

The immediate price reaction to block selling is thus $p_0 - p_{-1/2} = \gamma\rho_0(\bar{s})\bar{s} + \lambda(\bar{s})\bar{s}$; however, in the longer term, the strategic component disappears, $p_{-1/2} - p_{1/2} = \gamma\rho_0(\bar{s})\bar{s}$. The strategic component is therefore temporary, while the non-strategic component is permanent.

1.4.1 Theoretical results

In this section, I investigate the shapes of liquidity measures (price reaction function, and its' two components), focusing on the monotonicity, convexity and asymmetry of those functions. These properties have been the focus of empirical studies.

The expressions for the price reaction function and its' components are given by (1.4.2), (1.4.3) and (1.4.4), respectively. It then follows that for a normal distribution, the three functions are linear in the absolute value of the order size $|\bar{s}|$, and are also symmetric:

$$\Pi(s) = \gamma(1+\nu)\sigma^2 \cdot |\bar{s}|,\ \tau(s) = \gamma\nu\sigma^2 \cdot |\bar{s}| \text{ and } \pi(s) = \gamma\sigma^2 \cdot |\bar{s}|. \tag{1.4.5}$$

The analytical results in the general case can be obtained in two limits: when the order size is small ($|\bar{s}| \to 0$) and when it is large ($|\bar{s}| \to \infty$). Those two limits are associated with the two forces that play no role under normality. Higher moments play a role in the $|\bar{s}| \to 0$ limit, whereas the bounds of the support are important in the $|\bar{s}| \to \infty$ limit. The two limits, and the two respective forces, are examined separately below.

Proposition 3. *The price reaction function and its' non-strategic component are strictly increasing in the absolute value of the order size $|\bar{s}|$. Suppose that δ has bounded support $\delta \in [a, b]$. The following is true in the $|\bar{s}| \to \infty$ limit:*

buy order	sell order
$\Pi(-\infty) = b - \mu$	$\Pi(+\infty) = \mu - a$
$\pi(-\infty) = b - \mu$	$\pi(+\infty) = \mu - a$
$\tau(-\infty) = 0$	$\tau(+\infty) = 0$

Consequently, there exists $x_1 > 0$ such that the following is true for $s : |s| > x_1$:

Liquidity measure $L(s)$	$\text{sign}(L(s) - L(-s))$
$\Pi(s)$	$\text{sign}((\mu - a) - (b - \mu))$
$\pi(s)$	$\text{sign}((\mu - a) - (b - \mu))$

If $g'''(x)$ does not change sign given a large enough x, the price reaction function $\Pi(x)$ is concave for a large enough x.

It is intuitive that the price reaction function and the its' non-strategic component are monotone: larger orders have larger price effects. Finite limits at infinity and monotonicity imply that they have a horizontal asymptote $b - \mu$ ($\mu - a$) at minus (plus) infinity. This is also intuitive: the price reaction $\Pi(\cdot)$ cannot be higher than $b - \mu$ ($\mu - a$) for a very large buy (sell) order because otherwise the price itself would be higher than b (lower than a) and the block trader should not trade. The non-strategic part $\pi(\cdot)$ is a price reaction in a competitive economy; therefore, the latter argument also applies to it.

If the support is unbounded, the price reaction and its' non-strategic component would also be unbounded. The bounds of the distribution represent a force that bends those functions preventing violation of the bounds. The functions therefore cannot be convex. Moreover, a concave shape is typically observed. This is illustrated in figure 1.4.1, where I compare the price reaction function in the normal and truncated normal cases. Under the condition that the third derivative of the CGF does not change sign given a large enough x, it is possible to prove that the price reaction function is concave for large order sizes.[17]

The strategic component has to decrease to zero at infinity. This is because $\Pi(s) = \pi(s) + \tau(s)$ and the limits of $\Pi(s)$ and $\pi(s)$ coincide. Intuitively, the non-strategic component already drives prices toward the bounds of the payoff's support; therefore, if the non-strategic component is not zero, the price bounds will be violated. Therefore, the force bending the price reaction function and the non-strategic component is even stronger, making the strategic component non-monotone. Due to the presence of this force, the function $\tau(s)$ cannot be convex. If attention is restricted to the monotone part of $\tau(s)$, its shape also appears concave, as illustrated by the left pahnel of Figure 1.4.2.

[17] I could not find an example of a distribution with bounded support for which this condition does not hold. I believe that this condition is not too restrictive.

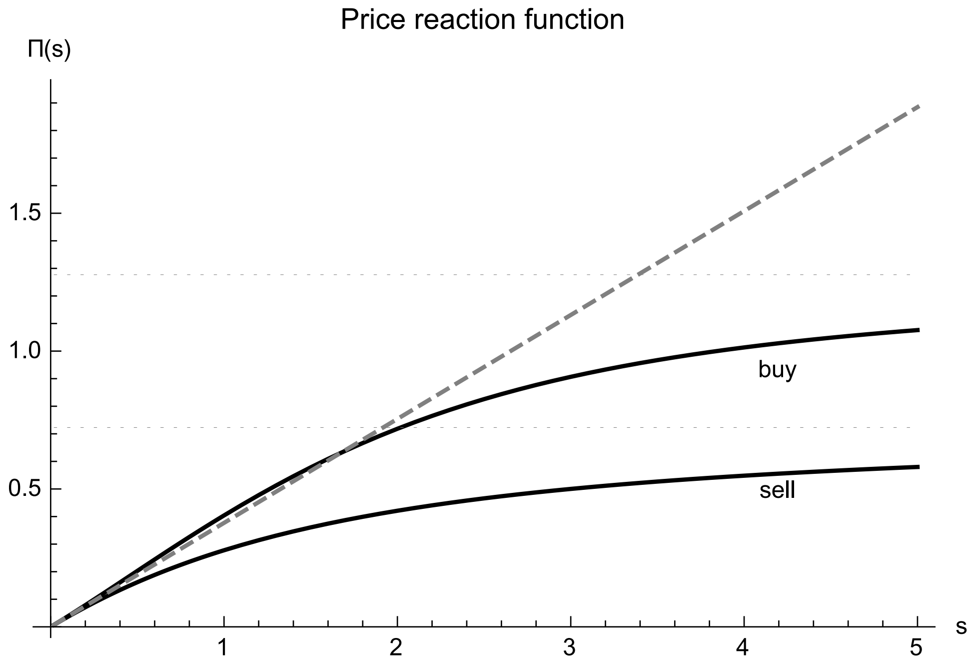

Figure 1.4.1: If the payoff is bounded, and hence the price of the asset and the price reaction function should also be bounded. The figure represents the price reaction function for the case of a normal distribution $N(0,1)$ truncated to the segment$[0,2]$ (solid lines with buy and sell orders labeled accordingly). It compares the latter to the price reaction function for the case of a normal distribution (with mean 0.72 and variance 0.25 equal to that of the truncated normal distribution) represented by the dashed line. For a normal distribution, the function is linear, unbounded and the same for buy and sell orders. For the truncated normal distribution, the function is asymmetric and is bounded by the horizontal dotted lines. The combination of boundedness and monotonicity produces the concave-looking shape. The shape is not exactly concave: the price reaction function for buy orders is above the tangent line (and hence convex) in the neighborhood of zero, consistent with Proposition 4.

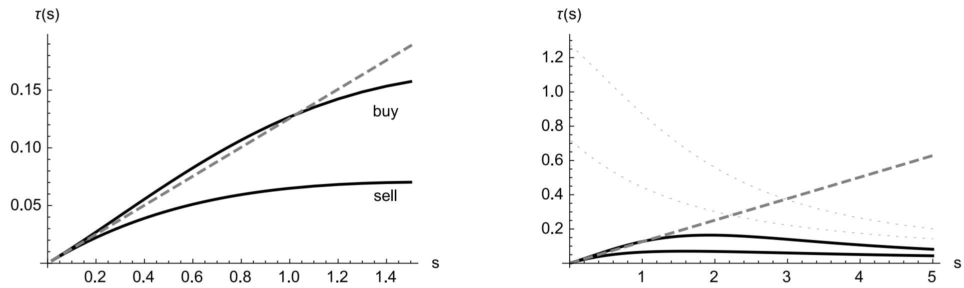

Figure 1.4.2: The bounds of the payoff produce a strategic component of price impact with a concave-alike shape if the order size is not very large (left panel). However, the force that shaped the price reaction function is even stronger for its' strategic component: it is bounded not by a horizontal line but by $\mu - a - \pi(s)$ $(b - \mu - \pi(s))$ for sell (buy) orders represented by a dotted line (right panel). The plots are for a normal distribution $N(0,1)$ truncated to the segment$[0,2]$

The asymmetry of the price reaction function at infinity obtains if $\mu \neq \frac{b-a}{2}$, i.e., if the mean of the distribution does not coincide with the middle of its support. Intuitively, because the shape of the price impact is linked to the shape of the distribution, asymmetry in the one leads to the asymmetry in the other.

In the proposition below I investigate the shape of liquidity measures for small order size.

Proposition 4. *Suppose that $L > 3$. Denote the skewness by κ_3. There exists $x_0 > 0$ such that the following is true :*

1) Local convexity and monotonicity:

Liquidity measure $L(s)$	sign($L'(\|s\|)$)	sign($L''(\|s\|)$)
$\Pi(s)$	$+^*$	-sign($\kappa_3 \cdot s$)
$\pi(s)$	$+^*$	-sign($\kappa_3 \cdot s$)
$\tau(s)$	$+$	-sign($\kappa_3 \cdot s$)

for all y if marked by $*$, and for $s : |s| < x_0$ otherwise.

2) Local asymmetry:

$$\text{sign}(\Pi(s) - \Pi(-s)) = -\text{sign}(\kappa_3)\, \forall s : |s| < x_0,$$

$$\text{sign}(\pi(s) - \pi(-s)) = -\text{sign}(\kappa_3)\, \forall s : |s| < x_0.$$

3) *Comparative statics:* $|\Pi(s) - \Pi(-s)|$ and $|\Pi''(s)|$ are increasing in ν, γ and $|\kappa_3|$ for $s : |s| < x_0$.

When asset payoffs are positively skewed, the price reaction function is a convex (concave) function of order size for small buy (sell) orders. Similarly, when asset payoffs are negatively skewed, the price reaction function is a concave (convex) function of order size for small buy (sell) orders.

Consider the case of a sell order and positive skewness. To understand the intuition, consider a benchmark economy in which higher moments play no role. It is identical to the initial economy, except that the asset's payoff is normally distributed with mean and variance equal to that in the initial economy. In the benchmark economy, the price reaction function is linear. This linear function is represented by a dashed line in the figure 1.4.2. Moreover, for a very small (infinitesimal) order size, the role of higher moments is negligible; therefore, the price impact function in the benchmark economy should be arbitrarily close to that in the initial economy in the neighborhood of zero. Consequently, the line representing the price impact in the benchmark economy is tangent to the price reaction in the initial economy (which the left panel of figure 1.4.2 illustrates). A concave function lies below its tangent line, and hence the remaining question is why the price impact in the benchmark economy is greater. The intuition is simple: with a positively skewed payoff, the trading profit of the investors accommodating the sale order is also positively skewed (i.e., they occasionally receive large positive surprises to their profits). Consequently, they require less price compensation relative to the case of zero skewness. Similarly, for small purchases, the shape of the price reation function is convex with positive skewness.

The above discussion also implies that with non-zero skewness the price reaction function is asymmetric for small orders. As noted above, with positive skewness the price reaction function for purchases is convex, and therefore lies above its' tangent line. It is concave for sell orders, and therefore lies below the tangent line and, consequently, below the price reaction function for purchases. Hence, with positive skewness the price reacts stronger to small purchases rather than sells. Similarly, with negative skewness the price reacts stronger to sells. Left panel of the figure 1.4.2 provides an illustration for the case of positive skewness.

1.4.2 An example

In this section, I examine numerically the case of a δ distributed according to the mixture of normal distributions. This example illustrates that, in theory, the shapes of the price impact can be quite rich. This case has at least two natural interpretations.

The first interpretation is as follows. Suppose that δ is a claim to cash flows generated by a firm. Suppose also that between time 0 and time 1, there is a corporate event (occurring with probability p) that may increase or decrease δ, also making it more or less risky. Suppose for simplicity that conditional on the outcome of this event, the distribution of δ is normal. The resulting distribution is then the mixture of normals

$$F_\delta(x) = p \cdot \Phi(x|\mu_1, \sigma_1) + (1-p) \cdot \Phi(x|\mu_2, \sigma_2), \tag{1.4.6}$$

where F_δ denotes the CDF of the dividend δ and $\Phi(x|\mu_i, \sigma_i)$ denotes the CDF of a normal distribution with mean μ_i and variance σ_i^2. The event can, for example, be:

- default, in which case p is the probability of default, μ_1 is a mean repayment in the event of default and $\sigma_1 \to 0$ if this repayment is certain.

- appointment of a new CEO, in which case p is the probability that a search for a new CEO is successful. If the new is CEO is better than the old one, one may expect $\mu_1 > \mu_2$ and $\sigma_1 < \sigma_2$.

The second interpretation is as follows. Suppose that all traders have a prior belief $\delta \sim N(\mu, \sigma)$ and receive the same signal ι concerning the asset. There is, however, uncertainty with respect to whether the signal is informative, in the spirit of Banerjee and Green (2014). For example, the signal might be

$$\iota = \begin{cases} \delta + \epsilon, & \text{with probability } p, \\ \epsilon, & \text{with probability } 1-p, \end{cases}$$

where δ and $\epsilon \sim N(0, \sigma_\epsilon^2)$ are independent. The posterior distribution will then be a mixture of normals (1.4.6) with $\mu_2 = \mu$, $\sigma_2^2 = \sigma^2$ and $\mu_1 = \mu + \beta(\iota - \mu)$, $\sigma_1^2 = \sigma^2 - \beta^2(\sigma^2 + \sigma_\epsilon^2)$, and $\beta = \frac{\sigma^2}{\sigma^2 + \sigma_\epsilon^2}$.

Figure 1.4.3 represents the risk functions $\rho_{1+\nu}(x)$ and $\rho_0(x)$. Recall that $\rho_{1+\nu}(|\bar{s}|)$ and $\rho_0(|\bar{s}|)$ are proportional to the per unit strategic and non-strategic components of price reaction function: e.g., for $s > 0$ we have $\tau(s)/\bar{s} = \nu\gamma\rho_{1+\nu}(\bar{s})$ and $\pi(s)/\bar{s} = \gamma\rho_0(\bar{s})$[18].

The intuition behind the shapes represented in figure 1.4.3 is the following. When the order size is small, a normal distribution with lower variance (denote it σ_1^2) dominates. The risk is thus small and is close to σ_1^2. For a large-sized order, the normal distribution with higher variance (denote it σ_2^2) dominates and the risk is close to σ_2^2. In between, the uncertainty regarding the variance increases risk and there are spikes in the risk functions. Therefore, if one aims to execute a small order, he or she should be optimistic about risk and assume that the level of risk is close

[18]The general formula is $\tau(|s|)/|\bar{s}| = \nu\gamma\rho_{1+\nu}(|\bar{s}|)$ and $\pi(|s|)/|\bar{s}| = \gamma\rho_0(|\bar{s}|)$.

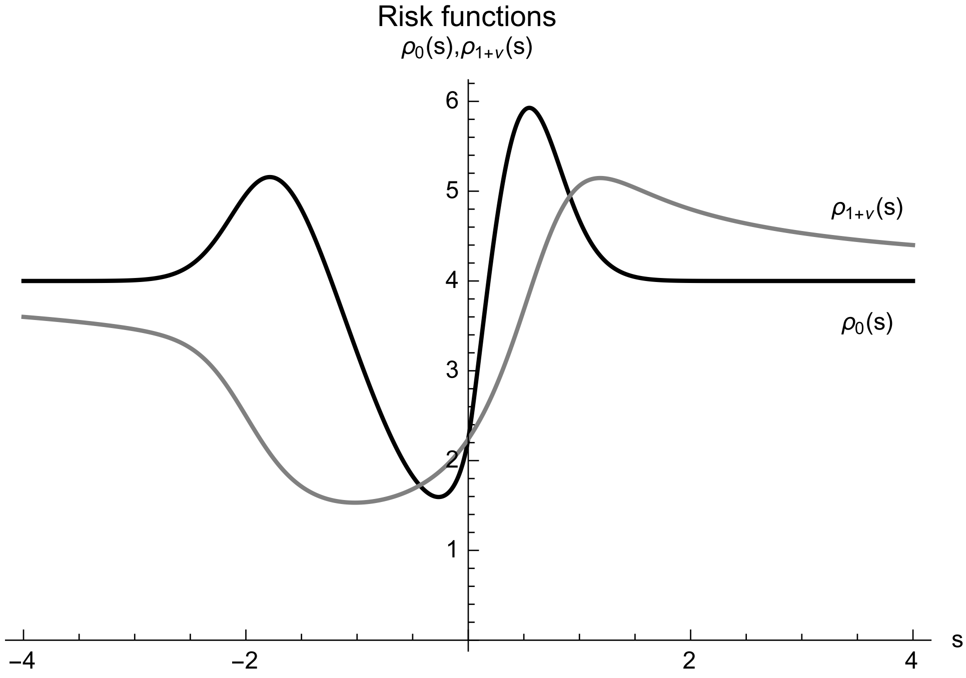

Figure 1.4.3: Mixture of normals $N(0,1)$ with probability 0.8 and $N(-2,2)$ with probability 0.2. For small order sizes, the normal with the lower variance of 1 dominates and the risk is close to 1. For large order sizes, the normal with the higher variance of 4 dominates and the risk is close to 4. In between, the risk spikes above 4 due to uncertainty regarding the variance.

to σ_1^2. If the order is very large, one should be conservative about risk and assume that the risk is close to σ_2^2. If the order is neither very large nor very small, one should be pessimistic, as the uncertainty makes the risk higher than σ_2^2.

1.5 Informed block trader

In this section, I relax the assumption that the block seller is uninformed, i.e., that s and δ are independent. The conditional distribution of δ given s is characterized by the conditional CGF

$$g(y,s) = \ln E[\exp(y\delta)|s].$$

As before, I assume that the distribution of s has full support. The conditional CGF and the conditional certainty equivalent, defined as $f(y,s)$, that solves

$$\exp(-\gamma f(y,s)) \equiv E[\exp(-\gamma y\delta)|s],$$

are related as follows

$$f(y,s) = -\frac{1}{\gamma} g(-\gamma y, s).$$

As before, my focus is liquidity: I will investigate how price reaction function and its' components depend on the size of supply s. I begin by deriving heuristically the equilibrium representation.

The approach to solving the problem is still ex post optimization, but now the residual supply curve realization reveals s and thus provides information on the terminal payoff. The ex post maximization problem is then written as

$$(x^*(s), p^*(s)) = \arg\max_{x,p} \{f(x, s) - p \cdot x\},$$

$$\text{s.t.: } (x, p) \in \mathcal{C}(s).$$

The first-order condition in the ex post maximization problem is given by

$$f_1(x, s) - l(x)x = p,$$

where $l(x) = \frac{\partial p}{\partial x}$ is the price sensitivity model. The above first-order condition makes it possible to determine the optimal x for a given s. In the symmetric equilibrium, $x = s/L$ should be optimal. Equivalently, x is optimal when $s = Lx$ is realized. The above can thus be rewritten as follows

$$f_1(x, Lx) - l(x)x = p = I(x). \tag{1.5.1}$$

As before, the way to specify the price sensitivity model is to require it to be *consistent*: the assumed price sensitivity should be equal to the equilibrium sensitivity. Consistency is equivalent to

$$l(x) = \lambda(x) = -\frac{1}{L-1} I'(x).$$

Combining the above and (1.5.1) yields the ODE for price impact function

$$x\lambda'(x) = (L-2)\lambda(x) + h(x),$$

where

$$h(x) \equiv \frac{d}{dx} f_1(x, Lx) = f_{11}(x, Lx) + Lf_{12}(x, Lx) = -\gamma g_{11}(-\gamma x, Lx) + Lg_{12}(-\gamma x, Lx). \tag{1.5.2}$$

To proceed further, I impose the following assumption.

Assumption 1. $h(x) < 0$.

The above assumption states that the supply distribution is such that the equilibrium marginal certainty equivalent $f_1(y|Ly)$ is decreasing in the equilibrium quantity y. Given that $g_{11}(x|y) > 0$ Assumption 1 is satisfied if a stronger assumption holds[19]

Assumption 1s. $f_{12}(x|s) < 0 \ \forall x, s$.

Assumption 1s says that higher supply realization is bad news for large traders. This is intuitive: supply is high when the informed traders sell a great deal, which is the case when they receive bad news about the asset.

The following theorem generalizes Theorem 1 for the setting with informed block traders.

Theorem 2. *If Assumption 1 holds, bid $b(p)$ is an equilibrium if, and only if, it satisfies the following three conditions:*

*(1) Bid $b(p)$ is **optimal** given a model $l(x)$ of price sensitivity*

$$f'(x, Lx) - xl(x) = p \Rightarrow x = b(p). \tag{1.5.3}$$

*(2) The assumed models of price sensitivity are **consistent**, i.e., the assumed models are equal to equilibrium price sensitivity, $l(x) = \lambda(x)$. The latter condition is equivalent to the following ODE*

$$x\lambda'(x) = (L-2)\lambda(x) + h(x). \tag{1.5.4}$$

*(3) **Monotonicity**: $0 < \lambda(x) < \infty \, \forall x$.*

The analysis in the main part of the paper can be generalized to the case of asymmetric information by substituting $h(x)$ for $f''(x)$. I define the *augmented risk function* $\varphi_a(x)$, which generalizes the risk function in the baseline model, as follows

[19]The conditional CGF $g(x|y)$ can be proven to be convex in x analogously to the case of the unconditional CGF, using Cauchy-Schwartz inequality.

$$
\begin{aligned}
\varphi_a(x) &\equiv -1/\gamma \int_0^1 h(t^{1-a}x)dt = \\
&= -1/\gamma \int_0^1 \left(f_{11}(t^{1-a}x, Lt^{1-a}x) + Lf_{12}(t^{1-a}x, Lt^{1-a}x)\right) dt \\
&= \underbrace{\int_0^1 g_{11}(-t^{1-a}\gamma x, Lt^{1-a}x)dt}_{\text{risk function } \rho_a(x)} \underbrace{-L/\gamma \int_0^1 g_{12}(-t^{1-a}\gamma x, Lt^{1-a}x)dt}_{\text{information function } \iota_a(x)} \\
&= \rho_a(x) + \iota_a(x).
\end{aligned}
$$

The augmented risk function is defined analogously to the risk function in the baseline setting with the sensitivity of marginal utility $f''(x)$ substituted by $h(x)$. Because the sensitivity of the marginal utility has two components, the sensitivity to information and the sensitivity to the quantity (due to risk aversion), the augmented risk function can be decomposed into two parts: inventory risk (simply risk hereafter) and information.

The risk function is familiar: it measures the sensitivity of marginal certainty equivalent $f_1(x, s)$ to the size of the position in the risky asset x. The latter sensitivity is not zero because the asset is risky and the traders are risk averse.

The information function measures the sensitivity of marginal certainty equivalent $f_1(x, s)$ to the news provided by the supply s. A higher equilibrium quantity x of the risky asset indicates higher supply provided by the informed traders, which, in turn indicates bad news about the terminal payoff δ under Assumption 1s. Therefore, the information function summarizes the *winner's curse* faced by the large traders: if they are allocated a great deal of supply in equilibrium, this means that the suppliers sold a great deal, which indicates that the asset is of poor quality.

It is instructive to consider the following example.

Example 1. Suppose that δ and s are jointly normal. Then, by projection theorem

$$g(x, s) = (\bar{x} + \beta(s - E[s]))x + 1/2\sigma^2_{\delta|s}x^2,$$

where $\beta = \frac{cov(x,y)}{var(y)}$, and $\sigma^2_{\delta|s}$ is a conditional variance of δ given s. The risk function and the information function are given by

$$\rho_a(x) = \sigma^2_{\delta|s},\ \iota_a(x) = \beta.$$

The above indicates that in the CARA-normal case, both components of the risk function are constants. The risk function is a conditional volatility. The information function is the sensitivity of the conditional mean to the signal.

The analysis in the main part of the paper can be generalized to the case considered in this section by substituting the augmented risk function for the risk function. Following the steps of Proposition 1, one obtains the results for the case considered here. It will be clear from Proposition 6 that the solution with $C^{+} = C^{-} = 0$ is the unique one satisfying prices being in $[a, b]$ for the bounded payoff $\delta \in [a, b]$, thus providing grounds for focusing on this equilibrium here.

As in the augmented risk function, all liquidity measures now separate into two components that are denoted by superscripts ρ (risk component) and ι (information component).

Proposition 5. *Suppose that Assumption 1s holds.*

(1) The equilibrium with asymmetric information exists if and only if $\varphi_{1+\nu}(x) < \infty$.

(2) The equilibrium price impact function is decomposed into risk and information components and is given by

$$\lambda(x) = \nu\gamma\varphi_{1+\nu}(x) = \nu\gamma\rho_{1+\nu}(x) + \nu\gamma\iota_{1+\nu}(x),$$

(3) The equilibrium price is given by

$$p = \mu - \gamma\varphi_0(\overline{s})\overline{s} - \lambda(\overline{s})\overline{s}. \tag{1.5.5}$$

The strategic and non-strategic components of price reaction function are decomposed into risk and information parts

$$\pi(\mathrm{s}) = \gamma\phi_0(\overline{\mathrm{s}})\overline{\mathrm{s}} = \gamma\rho_0(\overline{\mathrm{s}})|\overline{\mathrm{s}}| + \gamma|\iota_0(\overline{\mathrm{s}})\overline{\mathrm{s}}| \equiv \pi^{\rho}(\mathrm{s}) + \pi^{\iota}(\mathrm{s}),$$

$$\tau(\mathrm{s}) = \nu\gamma\phi_{1+\nu}(\overline{s})\overline{s} = \nu\gamma\rho_{1+\nu}(\overline{s})|\overline{s}| + \nu\gamma|\iota_{1+\nu}(\overline{s})\overline{s}| \equiv \tau^{\rho}(s) + \tau^{\iota}(s).$$

The results of the main part of the paper are nicely generalized: one has to augment the risk function with the information function to obtain equilibrium objects in the informed block trader case. The price reaction function is nicely decomposed into two components. The risk component is due to the risk aversion of the traders, and the information component is due to the winner's curse. The non-strategic component arises because traders account for their own risk aversion and the winner's curse to which they are exposed. The strategic component arises because a trader strategically recognizes that others are risk averse and exposed to the winner's curse, which affects his price sensitivity.

Below, I investigate the shape of the price reaction function in two limiting cases: $s \to \infty$ and $s \to 0$. I first examine the $s \to \infty$ limit and introduce the following assumption.

Assumption 2. The support of the conditional distribution of δ given s is the same as the support of the unconditional distribution of δ.

The above assumption states that the traders cannot improve their knowledge about the support of the terminal payoff by observing the supply. This assumption holds, for example, when the supply can be decomposed into information and noise, $s = f(\delta) + \epsilon$ and the noise ϵ has a full support[20].

Proposition 6. *Suppose that δ has a bounded support and that Assumption 1s holds. Then the price reaction function and its' non-strategic component are strictly increasing in the absolute value of the order size $|\bar{s}|$. Suppose that δ has bounded support $\delta \in [a, b]$ and Assumption 2 holds . The following is true in the $|\bar{s}| \to \infty$ limit:*

buy order	*sell order*
$\Pi(-\infty) = b - \mu$	$\Pi(+\infty) = \mu - a$
$\pi(-\infty) = b - \mu$	$\pi(+\infty) = \mu - a$
$\tau(-\infty) = 0$	$\tau(+\infty) = 0$

The proposition above confirms the robustness of the results of Proposition 3 for the case in which the block trader may be informed. As long as the available information does not allow to improve one's knowledge of the support of the distribution of δ, the boundedness of δ works in the the same way as in the uninformed case. It represents a force that bends down functions $\Pi(\cdot)$ and $\tau(\cdot)$. The functions therefore cannot be convex - exactly as in the baseline setting.

Below, I investigate the shape of the price impact in the $s \to 0$ limit. Before doing so I introduce the following notation. The conditional mean, variance and

[20] Indeed, $f_{\delta|s} = \frac{f_{s|\delta} \cdot f_\delta}{f_s}$, and $f_{s|\delta}$ and f_s have full support.

skewness of δ given s are all functions of s and are denoted by $\mu(s)$, $\sigma^2(s)$ and $\kappa_3(s)$, respectively.

Proposition 7. *Suppose that $L > 3$ and Assumption 1s holds. There exists $x_0 > 0$ such that the following is true.*

1) Local convexity and monotonicity:

Liquidity measure $L(s)$	$\text{sign}(L'(\|s\|))$	$\text{sign}(L''(\|s\|))$
$\Pi^\rho(s)$	$+$	$\text{sign}\big(\big((\sigma^2)'(0) - \gamma/L \cdot \kappa_3(0)\big) \cdot s\big)$
$\Pi^\iota(s)$	$+$	$\text{sign}\big(\big((\sigma^2)'(0) - L/\gamma \cdot \mu''(0)\big) \cdot s\big)$
$\pi^\rho(s)$	$+^*$	$\text{sign}\big(\big((\sigma^2)'(0) - \gamma/L \cdot \kappa_3(0)\big) \cdot s\big)$
$\pi^\iota(s)$	$+^*$	$\text{sign}\big(\big((\sigma^2)'(0) - L/\gamma \cdot \mu''(0)\big) \cdot s\big)$
$\tau^\rho(s)$	$+$	$\text{sign}\big(\big((\sigma^2)'(0) - \gamma/L \cdot \kappa_3(0)\big) \cdot s\big)$
$\tau^\iota(s)$	$+$	$\text{sign}\big(\big((\sigma^2)'(0) - L/\gamma \cdot \mu''(0)\big) \cdot s\big)$

for all y if marked by $*$, and for $s : |s| < x_0$ otherwise.

2) Local price reaction function asymmetry:

$$\text{sign}(\Pi(s) - \Pi(-s)) = \text{sign}\left(2\gamma L \left(\sigma^2\right)'(0) - \gamma^2\kappa_3(0) - L^2\mu''(0)\right) \ \forall s : |s| < x_0.$$

One additional force determines the shape of the liquidity measures relative to the baseline model. In the setting considered here, the moments are *functions* of the supply s. In addition to higher moments, the shape of the conditional moments (as functions of the size of the supply s) also plays a role. In particular, the convexity of the information component of the liquidity measures is also driven by the convexity of the conditional mean function, whereas the convexity of the risk component is driven by the slope of the conditional variance. Note that one would not capture those effects in the jointly normal setting (see, e.g., example 1) because in that case the conditional mean is linear (convexity is zero) and the conditional variance is a constant (the slope is zero).

If one can separate the the inventory risk and the information components of the price impact (Muraviev (2015) provides an excellent example of how one can do so), then one can extract the information about how an order affects the expectations of traders accommodating it. The proposition below shows that the slope of the information component of the price impact is proportional to the slope of the conditional mean $\mu(s)$.

Proposition 8. *The slope of the conditional mean function and the slope of the information component of the price impact are related as follows:* $\mu'(0) = -\frac{(\Pi^\iota)'(0)}{L(1+\nu)}$.

The slope of the conditional mean function shows how a marginal unit liquidated by a block trader affects the expectations of liquidity providers regarding the asset's payoff. Consequently, all else being equal, in markets with greater slope of $\Pi^\iota(\cdot)$, a marginal liquidated unit affects the expectation more, which is indicative of a more informed block trader.

1.6 Empirical evidence and testable predictions

I begin by summarizing the key findings concerning the shape of the price impact. These are as follows:

1. The price reaction function is a concave function of the order size.[21]

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

Proposition 3 implies that with bounded asset payoff, the price reaction function is concave for a large order size, providing the explanation for the first finding. Propositions 3 and 7 link the asymmetry of the permanent price impact to asymmetry and skewness of the distribution. The model predicts that with positive skewness (a natural property for stocks at the individual level (e.g. Chen et al. (2001))) price impact of small purchases is greater than that of small sells, consistent with evidence summarized by Saar (2001).

While Saar (2001) summarizes the evidence of a greater price impact of purchases compared to sells, Chiyachantana et al. (2004) find that the asymmetry of the permanent price impact is linked to the underlying market condition, and show, in particular, that in the bearish markets sells have a higher price impact. My model links the local asymmetry to skewness, and predicts that when skewness is negative small sells have a higher price impact compared to small purchases. Perez-Quiros and Timmermann (2001) present the evidence that skewness varies with the

[21]Equities: Hausman et al. (1992), Almgren et al. (2005), Frazzini et al. (2014). Options: Muraviev (2015).

[22]Saar (2001) summarizes the evidence that shows bigger price impact of buy orders compared to sell orders. However, Chiyachantana et al. (2004) link the asymmetry to the underlying market condition and find that in bullish markets buy orders have a bigger price impact than sells, while in the bearish markets sells have a higher price impact.

underlying market condition. The model predictions are therefore in line with these findings.

While empirical papers typically find concave price reaction functions, there is some conflicting evidence for price impact of purchases. E.g., Keim and Maddavan (1996) and Holthausen, Leftwich and Mayers (1990) find the convex price reaction function for purchases. This is in line with the prediction of my model: with positive skewness the price reaction function for small purchases is convex. Thus, model findings provide a way of reconciling the conflicting evidence.

Proposition 8 demonstrates that the slope of the asymmetric information component of the price reaction function is linked to the slope of the conditional mean function (i.e., the extent to which, on the margin, the order affects the expectations of liquidity providers about the asset's payoff). Muraviev (2015) presents a methodology that makes it possible to separate the price impact into inventory and information components. The idea is that, when assets are traded on multiple exchanges, the information spreads among traders on different exchanges, while the inventory risk is only accommodated by the traders on the exchange in which the block trader executes the order. Consequently, one can estimate the shape of the information component of the price impact, in particular its slope, which is diagnostic of the informativeness of block trades and can be useful in detecting informed trading.

My model identifies two forces that shape the price reaction function. For small order sizes, its' curvature is linked to skewness. The model implies that the difference in curvatures of price reaction function for purchases and sells is positive with positive skewness and is negative with negative skewness. One can test it in equities market. At the individual level stock returns are positively skewed, whereas at the aggregate level the skewness is negative. Therefore, the difference in curvatures of the price reaction function for purchases and sells should be positive for individual stocks but negative for ETFs.

For large order sizes, the bounds of the payoff represent another force affecting the curvature of price reaction function. One can examine the role of this force in the options market. For put options, the upside is limited, while this is not the case for calls. Consequently, for buy orders, the price reaction function should be more concave for puts relative to calls. Muraviev (2015) finds that the price reaction functions in options markets are concave, but, unfortunately, he did not estimate them separately for calls and puts.

One can also examine the role of payoff bounds and skewness by running cross-sectional regressions, in which one estimates the non-linear model of price reaction function also controling for skewness and “distance to the bound”, maximal payoff minus current price for purchases and current price minus minimal payoff for sells.

The model also implies that the curvature and asymmetry is more pronounced when market is less competitive and when the risk bearing capacity is smaller. Following Nagel (2012) the changes in risk bearing capacity can be proxied by VIX. Consequently, when VIX is high, the asymmetry and curvature of price response should be higher. The variation in competitiveness can be observed by comparing after hours and regular hours market. There are less market makers in the after hours market and, consequently, the asymmetry and difference in curvatures should be higher. At a higher frequency, the competitiveness of the market can be proxied by Herfindahl indices, as in Hasbrouck (2015).

1.7 Conclusion

This paper presents a tractable model of strategic trading without normality. It develops a methodology that makes it possible to solve for the equilibrium in a constructive way, which allows one to uncover the multiplicity of equilibria. Closed-form solutions are helpful in selecting the unique equilibrium. The paper also demonstrates that the two forces absent under the normal distribution, the boundedness of the support and higher moments, play an important role in determining the shape of liquidity measures, such as price reaction function and its' components.

This paper focuses on the implications of departures from normality to shape of price impact function and shuts down several channels that might be worth studying. First, as all large traders are symmetric, there is no risk sharing between them. Second, as all traders are informed symmetrically, no information aggregation is taking place. Finally, as the model is static, I am unable to study how nonlinearities in the shape of the liquidity measures affect optimal order break up and the dynamics of the price impact.

Extending the model in any of the above directions is promising and is left for the future work. Below, I comment on the potential difficulties that one may encounter in investigating in those directions.

Solving the model with heterogeneous traders to study risk sharing is challenging because the symmetry of equilibrium generating the tractability in this paper will no longer be present.

Incorporating asymmetric information into the model is challenging because without normality, the bids are not linear and the uncertainty faced by each trader (which comes from the signals that other traders receive and the supply) does not aggregate in a scalar parameter shifting the residual supply curve, and one cannot use ex post maximization techniques.

The dynamic extension is challenging for a reason similar to the heterogeneous case. One has to find a value function, which will depend on endowments of other traders. In particular, one should find the value function for the case in which the endowments are asymmetric.

Despite the above mentioned difficulties, I believe that it still might be possible to approach the above questions, perhaps with the help of numerical techniques.

1.8 Omitted proofs.

Proof. (Theorem 1) Consider a particular trader and fix the strategies of all other traders to be the equilibrium bid $x(p)$. Denote the inverse of $x(p)$ by $I(x)$. It is proved in the Appendix 1.10 that the equilibrium bids are strictly decreasing and have a finite slope. Given this the inverse residual supply and the inverse bid are both well-defined objects. The inverse residual supply is given by

$$P(x; s) = I\left(\frac{s - x}{L - 1}\right). \tag{1.8.1}$$

The ex-post maximization problem $\mathcal{P}_{EP}$ can be written as

$$\max_{x,p} f(x) - p \cdot x$$
$$s.t.: \; p = P(x; s).$$

Substituting the constraint into the objective and taking the first order condition with respect to x yields the following equation determining optimal quantity $x^*(s)$ on the residual supply curve for a given realization of s

$$f'(x^*) - x^* \underbrace{\frac{\partial}{\partial x} P(x;s)|_{x=x^*}}_{=\lambda(x^*)} - P(x^*;s) = 0. \tag{1.8.2}$$

Since $P(x^*(s);s)$ is the optimal price p^* corresponding to x^* the above becomes the expression (1.3.11) for the inverse bid

$$f'(x^*) - \lambda(x^*)x^* = p^* = I(x^*).$$

Differentiating the above with respect to x^* and applying the link (1.3.9) between the price impact function and the slope of the inverse bid in the symmetric equilibrium I get the ODE (1.5.4)

$$x\lambda'(x) = (L-2)\lambda(x) + f''(x). \tag{1.8.3}$$

The Proposition C1 implies that $0 < \lambda(x) < \infty$. The only thing that left is to check the second order conditions.

Second order conditions.

I will verify that 1) the first order condition gives unique candidate $x^*(s) = \frac{s}{L}$ and 2) the second order conditions are satisfied for this x^*.

First, plugging (1.8.1) into (1.8.2) one rewrites the first order condition (1.8.2) as

$$f'(x) + \frac{x}{L-1} I'\left(\frac{s-x}{L-1}\right) = I\left(\frac{s-x}{L-1}\right). \tag{1.8.4}$$

In the symmetric equilibrium $x = s/L$ should be optimal, so plugging it into (1.8.4) I find the ODE for the inverse bid (I denote $y = s/L$)

$$f'(y) + \frac{y}{L-1} I'(y) = I(y). \tag{1.8.5}$$

To show 1) I need to demonstrate that given that inverse bid satisfies (1.8.5) the unique solution to (1.8.4) is $x^* = s/L$.

Define $\xi = \frac{s-x}{L-1}$ and rewrite (1.8.4) as

$$f'(x) + \frac{x}{L-1} I'(\xi) = I(\xi).$$

For a given ξ the function on the left hand side of the above decreases in x whereas the right hand side does not depend on x. Therefore for each ξ there is at most one $x^*(\xi)$ solving the above. Since $x^*(\xi) = \xi$ is a solution (the above becomes (1.8.5)) we find that x^* solving $x^* = \frac{s-x^*}{L-1}$ is the unique optimal quantity. Therefore $x^* = s/L$.

To check second order conditions at $x^* = s/L$ differentiate (1.8.4) with respect to x to get the expression for the second derivative of the objective function

$$\begin{aligned}
\text{second derivative} &= f''(x) - \frac{2}{L-1} I'\left(\frac{s-x}{L-1}\right) - \frac{x}{(L-1)^2} I''\left(\frac{s-x}{L-1}\right) \\
&= f''\left(x^*\right) - \frac{2}{L-1} I'\left(x^*\right) - \frac{x^*}{(L-1)^2} I''\left(x^*\right) \\
&= f''\left(x^*\right) - \lambda\left(x^*\right) + \frac{x^* \lambda'\left(x^*\right) - (L-1)\lambda\left(x^*\right)}{(L-1)} \\
&= \left(f''\left(x^*\right) - \lambda\left(x^*\right)\right)\left(1 + \frac{1}{L-1}\right) < 0,
\end{aligned}$$

where the second line substitutes $x^* = s/L$, the third line uses (1.3.9), the fourth line substitutes (1.8.3). Since the second derivative is negative, the second order conditions are satisfied. □

Proof. (Proposition 1) (1) The equilibrium exists iff there is a solution to (1.3.12) satisfying $0 < \lambda(x) < \infty$. It follows from (1.3.15) that

$$x^{2-L}\lambda(x) = x_0^{2-L}\lambda(x_0) + \int_{x_0}^{x} f''(t)t^{1-L}dt. \tag{1.8.6}$$

Consider the case $x \geq x_0 > 0$. The condition $0 < \lambda(x) < \infty$ should hold for all x, therefore $\int_{x_0}^{x} f''(t)t^{1-L}dt$ has to be a bounded function of x for a given x_0. Indeed, we have that $-x_0^{2-L}\lambda(x_0) < \int_{x_0}^{x} f''(t)t^{1-L}dt$ from $\lambda(x) > 0$ and $\int_{x_0}^{x} f''(t)t^{1-L}dt < 0$ from $f''(t) < 0$.

Consider a function $\phi(x) = \int_{x_0}^{x} f''(t)t^{1-L}dt$. It is a decreasing function, which according to the above should be bounded. The latter condition is equivalent to $\lim_{x\to\infty} \phi(x) < \infty$.Thus we have that

$$\lim_{x\to\infty} \int_{x_0}^{x} f''(t)t^{1-L}dt < \infty,$$

which is equivalent to

$$\lim_{x\to\infty} \underbrace{x_0^{L-2} \int_{x_0}^{x} f''(t)t^{1-L}dt}_{(2-L)\rho_{1+\nu}(x_0)/\gamma} < \infty, .$$

The above implies that $\rho_{1+\nu}(x_0) < \infty$ for any given $x_0 > 0$. We can show that $\rho_{1+\nu}(x_0) < \infty\ \forall x_0 < 0$ analogously. The fact that $\rho_{1+\nu}(0) < \infty$ follows from $\rho_{1+\nu}(0) = \sigma^2$.

The fact that $\rho_{1+\nu}(x_0) < \infty$ for any distribution with bounded support follows from the Fact 5.

(2) I first find a particular solution and then find the general solution by adding the solution of the homogenous equation. Consider the solution satisfying $\lambda(x) = o\left(x^{L-2}\right)$. Taking the limit in the (1.8.6) as $x \to \infty$we get that this solution is given by

$$\lambda(x_0) = \nu\gamma\rho_{1+\nu}(x_0). \tag{1.8.7}$$

It follows from the above that $\frac{\lambda(x_0)}{x_0^{L-2}} = -\int_{x_0}^{\infty} f''(t)t^{1-L}dt$. The limit of the latter expression as $x_0 \to \infty$ is zero, so $\lambda(x)$ is indeed $o\left(x^{L-2}\right)$.

The general solution to the ODE (1.3.12) can be written as a particular solution plus a general solution of a homogenous ODE. Hence one can write

$$\lambda(x) = \nu\gamma\rho_{1+\nu}(x) + \mathbb{1}(x \geq 0)C^+x^{L-2} + \mathbb{1}(x < 0)C^-\left(-x\right)^{L-2}.$$

Since $\rho_{1+\nu}(x) = o\left(x^{L-2}\right)$ only solutions with $C^+, C^- \geq 0$ will satisfy $\lambda(x) > 0$. They all will satisfy $\lambda(x) < \infty$ because $\rho_{1+\nu}(x) < \infty$.

(3) The equilibrium inverse bid follows directly from the condition (2) of the Theorem 1. □

Proof. (Proposition 2) When the supply realization is s the equilibrium price in the case $C^+ = C^- = 0$ is

$$p = I(s/L) = g'(-\gamma s/L) - \nu\gamma\rho_{1+\nu}(s/L)s/L.$$

The bid is strictly decreasing and it follows from Results B2 and B4 that its limit as $x \to -\infty (x \to +\infty)$ is equal to $b(a)$. The inverse bid and prices are thus within $[a, b]$. If $C^+ > 0$ ($C^- > 0$) then the inverse bid is less than a (greater than b) for high enough (low enough) s due to unbounded power terms $C^{+}x^{L-2}$ (respectively $C^{-}x^{L-2}$). This fact implies that the equilibrium satisfying property (1) is unique.

To prove (2) note that one can write

$$p_n(s) = \int_1^{\infty} z^{-L} g_n'(-\gamma z \cdot s/L) dz,$$

where $g_n(t)$ is a CGF of an asset that pays δ_n. Interchanging the limit and integration (which is possible due to Dominated Convergence Theorem) one gets

$$p_n(s) \to \int_1^{\infty} z^{-L} \lim_{\substack{a_n \to -\infty \\ b_n \to +\infty}} g_n'(-\gamma z \cdot s/L) \cdot dz = \int_1^{\infty} z^{-L} g'(-\gamma z \cdot s/L) \cdot dz,$$

where $g(t)$ is a CGF of an asset that pays δ. The last equality follows from the application of the Dominated Convergence Theorem.

The third property follows directly by taking the limits in the results of Proposition 1. $\square$

Proof. (Proposition 3) The monotonicity of price reaction function follows from the monotonicity of equilibrium bid (recall that $p(s) = I(s/L)$). The monotonicity of permanent price impact in $|\bar{s}|$ follows from

$$\pi(s) = |f'(\bar{s}) - \mu|, \tag{1.8.8}$$

eq. (1.2.2) and the convexity of the CGF (Fact 1 in the Appendix 1.9).

That $\pi(-\infty) = b - \mu$ and $\pi(+\infty) = \mu - a$ follows from (1.8.8), (1.2.2) and Fact 2.

The limits $\tau(-\infty) = 0$ and $\tau(+\infty) = 0$ follow from (1.4.3) and the Fact 5.

Since $\Pi(\cdot) = \pi(\cdot) + \tau(\cdot)$, one gets $\Pi(-\infty) = b - \mu$ and $\Pi(+\infty) = \mu - a$ immediately from the above two limits.

It can be shown that the inverse bid is given by $I(x) = \int_1^{\infty} z^{-L} f'(zx) dz$. The convexity of the price reaction function at x is the same as the convexity of the inverse bid x/L. The latter is given by

$$I''(x) = \int_1^{\infty} z^{2-L} f'''(zx) dz.$$

For large enough x, the term $f'''(zx)$, $z \geq 1$ is either positive or negative, and hence the bid is either concave or convex. The convex shape is ruled out by the fact that the inverse bid is bounded. □

Proof. (Proposition 4) The global monotonicity of the functions $\Pi(\cdot)$ and $\tau(\cdot)$ was established in the Proposition 3 above. To prove that $\tau(|s|)$ is locally increasing I calculate its' derivative at zero. For positive s the function $\tau(|s|)$ can be written as (cf. (1.4.3)):

$$\tau(|s|) = \tau(s) = \nu\gamma\rho_{1+\nu}(\bar{s})\bar{s}.$$

Calculating the derivative with respect to $\bar{s}$ at zero one gets (cf. Fact 4) $\nu\gamma\rho_{1+\nu}(0) = \nu\gamma\sigma^2 > 0$. For negative s we have $\tau(|s|) = \tau(-s) = -\nu\gamma\rho_{1+\nu}(-\bar{s})\bar{s}$ and the calculation leads, analogously $\tau' = \nu\gamma\sigma^2 > 0$. Calculating the second derivative of $\tau(\cdot)$ at zero one gets $\nu\gamma\rho'_{1+\nu}(0) = -\nu\gamma\kappa_3\sigma^3$, where I've used Fact 4 and (1.2.1). For negative s the calculation is analogous and yields $\nu\gamma\kappa_3\sigma^3$. We thus get $\tau''(0+) = -\text{sign}(\kappa_3)$ while $\tau''(0-) = \text{sign}(\kappa_3)$. The convexity of price reaction function and the permanent price impact is considered analogously.

To prove the result about asymmetry note that the permanent price impact function for buy and sell orders have common slope at the origin, but different convexity. The asymmetry then follows.

The comparative statics follow from the following approximation, which one can obtain using Fact 3 and Fact 4: $x\rho_a(x) = \sigma^2 x - \frac{2\alpha}{1+\alpha}\gamma g'''(0)x^2 + o\left(x^2\right)$. □

Proof. (Theorem 2) Consider trader i and let his bid be $x_i(p)$. Fix the strategies of all other traders to be equilibrium bid $x(p)$. The residual supply in that case is given by

$$R(p; s) = s - (L-1)x(p). \tag{1.8.9}$$

The ex-post maximization problem $\mathcal{P}_{EP}$ can be written as

$$\max_{x,p} f(x, s) - p \cdot x$$

$$s.t.: \; x = R(p; s).$$

Substituting the constraint into the objective and taking the first order condition with respect to p yields the following equation determining optimal price $p^*(s)$ on the residual supply curve for a given realization of s

$$f_1(R(p^*,s),s)R_p(p^*,s) - R(p^*,s) - p^* \cdot R_p(p^*;s) = 0. \tag{1.8.10}$$

Given that the second-order conditions are satisfied the optimal price $p^*(s)$ is unique. The corresponding optimal quantity $x^*(s)$ is given by $x^*(s) = R(p^*;s)$. In the symmetric equilibrium $x^*(s) = x(p^*(s)) = s/L$. It follows that $R(p^*;s) = b(p^*(s))$. Substituting it to (1.8.10), noting that $R_p(p^*;s) = -(L-1)b'(p^*(s))$ (follows directly from (1.8.9)) and denoting $p = p^*(s)$ we rewrite (1.8.10), after some rearrangement, as

$$b'(p) = \frac{b(p)}{(L-1)(p - f_1(b(p), Lb(p)))} \equiv \phi(p,b). \tag{1.8.11}$$

Under assumption 1, $\psi(x) = f_1(x, Lx)$ is strictly decreasing in x. The analysis of the symmetric info case is applicable substituting $\psi(x)$ instead of $f'(x)$. In particular the ODE for price impact function is obtained by substituting $\psi'(x)$ instead of $f''(x)$ in the ODE (1.3.12) the analysis of the second order conditions can be done via the analogous substitution. □

Proof. (Proposition 5) The proof follows the same steps as the proof of Proposition 1 writing $h(x)$ instead of $f''(x)$ everywhere. □

Proof. (*Proposition 6)* The monotonicity of price reaction function follows from the monotonicity of equilibrium bid. The monotonicity of the permanent price impact follows from $\pi(s) = |f_1(\bar{s}, L\bar{s}) - \mu|$ and the fact that the function $f_1(\bar{s}, L\bar{s})$ is monotone in $\bar{s}$ provided that Assumption 1s holds.

To get the limits of $\pi(\cdot)$ at infinity note two facts. First, for any fixed y the we have$a \leq f_1(x,y) \leq b$ (monotonicity of $f_1(x,y)$ in x and the Fact 2). Second, Assumption 1s implies that for any $y \geq 0$ $f_1(x,y) \leq f_1(x,0)$. Combining the two facts we get that $a \leq f_1(x, Lx) \leq f_1(x,0)$. Taking the limit as $x \to +\infty$ using Fact 2 and applying Squeezing Theorem we get that $\lim_{x\to+\infty} f_1(x,Lx) = a$ from which $\pi(+\infty) = \mu - a$ follows. The limit at minus infinity can be derived analogously.

To get the limits of price reaction function at infinity note that it can be written as $\Pi(s) = |p(s) - \mu| = \left|(L-2)\int_1^\infty f_1(\bar{s}t, L\bar{s}t)t^{1-L}dt - \mu\right|$ (This can be most easily derived by solving the ODE $xI'(x) = (L-2)I(x) - f_1(x,Lx)$ for the inverse bid. The ODE can be obtained from 1.5.4 using $\lambda(x) = -\frac{1}{L-2}I'(x)$.). Now calculate the limit of $(L-2)\int_1^\infty f_1(xt, Lxt)t^{1-L}dt$ as $x \to \infty$. Using the Monotone Convergence Theorem and $\lim_{x\to+\infty} f_1(x, Lx) = a$ derived above we can write $\lim_{x\to+\infty}(L-2)\int_1^\infty f_1(xt, Lxt)t^{1-L}dt = (L-2)\int_1^\infty \lim_{x\to+\infty} f_1(xt, Lxt)t^{1-L}dt = a(L-2)\int_1^\infty t^{1-L}dt = a$. The limit at minus infinity can be derived analogously.

The limit of $\tau(\cdot)$ at infinity follows from $\tau(\cdot) = \Pi(\cdot) - \pi(\cdot)$ and the limits derived above. □

Proof. (Proposition 7 and 8) I start by deriving the results for the permanent price impact. The risk part for $s > 0$ can be written as

$$\pi^\rho(s) = \gamma\bar{s}\int_0^1 g_{11}(-\gamma t\bar{s}, Lt\bar{s})dt = \gamma\int_0^{\bar{s}} g_{11}(-\gamma y, Ly)dy.$$

Its derivative is thus given by $\gamma g_{11}(-\gamma\bar{s}, L\bar{s}) > 0$ and the second derivative at zero with respect to $|\bar{s}|$ is given by $-\gamma^2 g_{111}(0,0) + \gamma L g_{112}(0,0) = -\gamma^2 k_3(0) + \gamma L\left(\sigma^2\right)'(0)$ for $s > 0$ with the sign flipped for $s < 0$.

The information part can analogously be written as

$$\pi^\iota(s) = -L\int_0^{\bar{s}} g_{12}(-\gamma y, Ly)dy.$$

Its' first derivative is given by $-L \cdot g_{12}(-\gamma\bar{s}, L\bar{s}) > 0$ (Assumption 1s) (equal to $-L\mu'(0)$ at zero) and the second derivative at zero is $\gamma L\sigma'(0) - L^2\mu''(0)$ for $s > 0$ with the sign flipped for $s < 0$.

The risk part of the strategic component of the price reaction function can be written as

$$\tau^\rho(s) = \nu\gamma\bar{s}\int_0^1 g_{11}(-t^{-\nu}\gamma\bar{s}, Lt^{-\nu}\bar{s})dt.$$

It's first derivative at zero is given by $g_{11}(0,0) = \nu\gamma\sigma^2(0) > 0$ the second derivative at zero is given by $2\nu\gamma\int_0^1 t^{-\nu}\left(Lg_{112}(0,0) - \gamma g_{111}(0,0)\right)dt = \frac{2\nu}{1-\nu}\gamma L(\left(\sigma^2\right)'(0) - \gamma/L \cdot \kappa_3(0))$ for $s > 0$ with the sign flipped for $s < 0$.

The information part of the strategic component of the price reaction function can be written as

$$\tau^\iota(s) = -\nu L\bar{s}\int_0^1 g_{12}(-t^{-\nu}\gamma\bar{s}, Lt^{-\nu}\bar{s})dt.$$

Its' first derivative at zero is given by $-\nu Lg_{12}(0,0) > 0$ (equal to $-\nu L\mu'(0)$ at zero). The second derivative at zero is given by $-2\nu L\int_0^1 t^{-\nu}\left(Lg_{122}(0,0) - \gamma g_{112}(0,0)\right)dt = -\frac{2\nu}{1-\nu}\gamma L(L/\gamma \cdot \mu''(0) - \left(\sigma^2\right)'(0))$ for $s > 0$ with the sign flipped for $s < 0$.

The results for the price reaction function are obtained combining the results for $\pi(\cdot)$ and $\tau(\cdot)$. In particular, the slope of the price reaction function is given by $-L(1+\nu)\mu'(0)$. $\square$

1.9 Some properties of CGFs and the risk function

I start by outlining some properties of CGFs. They are certainly known but I present them with proofs to provide a self-contained treatment of the results.

1.9.1 Properties of CGFs

Fact 1. *The CGF of a non-degenerate distribution is strictly convex.*

Proof. Differentiating the definition of the CGF twice one gets

$$g''(x) = \frac{E[\delta^2 \exp(\delta x)]E[\exp(\delta x)] - E[\delta \exp(\delta x)]^2}{E[\exp(\delta x)]^2}.$$

The sign of $g''(x)$ is equal to the sign of $E[\delta^2 \exp(\delta x)]E[\exp(\delta x)] - E[\delta \exp(\delta x)]^2$. To complete the proof apply the Cauchy-Schwartz inequality (stating that $E[XY]^2 < E\left[X^2\right] E\left[Y^2\right]$ for linearly independent random variables X and Y) to the random variables $X = \delta \exp(\delta x/2)$ and $Y = \exp(\delta x/2)$. $\square$

The next result is particularly important to analyze the case of δ having bounded support.

Fact 2. *Suppose that the support of δ is (a, b) (with a and b possibly infinite). The first derivative of the CGF, $g'(x)$ is increasing and*

$$\lim_{x\to-\infty} g'(x) = a \text{ and } \lim_{x\to+\infty} g'(x) = b.$$

This result relies on the following two Lemmas.

Lemma 1. *Suppose that the support of δ is (a, b) (with a and b possibly infinite). Consider $c : a < c < b$. Then $E[\delta e^{x\delta}\mathbb{I}(\delta < c)] = O\left(e^{cx}\right)$* as $x \to +\infty$.

Proof. We need to show that $\exists x_0$, M such that $\forall x > x_0$

$$|E[\delta e^{x\delta}\mathbb{I}(\delta < c)]| < Me^{xc}. \tag{1.9.1}$$

Without loss of generality consider $x > 0$. Rewrite $|E[\delta e^{x\delta}\mathbb{I}(\delta < c)]|$ as

$$\begin{aligned} |E[\delta e^{x\delta}\mathbb{I}(\delta &< c)]| = Pr(\delta < c) \cdot |E[\delta e^{x\delta}|\delta < c]| \\ &< |c|E[e^{x\delta}|\delta < c] \\ &< |c|e^{xc}. \end{aligned}$$

It is clear that (1.9.1) holds with $M = |c|$ and $x_0 = 0$. □

Lemma 2. *Suppose that the support of δ is (a, b) (with a and b possibly infinite). Consider $c : a < c < b$. There exists $K > 0$ and $\theta > c$ such that*

$$E[e^{x\delta}\mathbb{I}(\delta \geq c)] \geq Ke^{\theta x}. \tag{1.9.2}$$

Proof. Rewrite $E[e^{x\delta}\mathbb{I}(\delta \geq c)]$ as

$$E[e^{x\delta}\mathbb{I}(\delta \geq c)] = Pr(\delta \geq c) \cdot E[e^{x\delta}|\delta \geq c].$$

By Jensen's inequality $E[e^{x\delta}|\delta \geq c] \geq e^{x\theta}$, where $\theta = E[\delta|\delta \geq c] > c$. Clearly, (1.9.2) holds with $K = Pr(\delta < c)$ and $\theta = E[\delta|\delta \geq c]$. □

I am now ready to provide the prof for the Fact 2.

Proof. (Fact 2) That $g'(x)$ is increasing follows directly from the Fact *1*.

I only prove that $\lim_{x\to+\infty} g'(x) = b$, since the second limit is analogous.

Consider arbitrary $c : a < c < b$ and rewrite

$$\begin{aligned} g'(x) &= \frac{E[\delta e^{x\delta}]}{E[e^{x\delta}]} = \frac{E[\delta e^{x\delta}\mathbb{I}(\delta < c)]}{E[e^{x\delta}\mathbb{I}(\delta < c) + E[e^{x\delta}\mathbb{I}(\delta \geq c)]} \\ &+ \frac{E[\delta e^{x\delta}\mathbb{I}(\delta \geq c)]}{E[e^{x\delta}\mathbb{I}(\delta < c) + E[e^{x\delta}\mathbb{I}(\delta \geq c)]}, \end{aligned}$$

Consider the first term. It follows from Lemmas 1 and 2 that there exist K, $M > 0$ and $\theta > c$ such that

$$\left|\frac{E[\delta e^{x\delta}\mathbb{I}(\delta < c)]}{E[e^{x\delta}\mathbb{I}(\delta < c) + E[e^{x\delta}\mathbb{I}(\delta \geq c)]}\right| \leq \frac{Me^{xc}}{Ke^{x\theta}} \to 0 \text{ as } x \to \infty.$$

It then follows from a Squeezing Theorem that the limit of the first term is zero. The latter implies that

$$\lim_{x\to\infty} g'(x) = \lim_{x\to\infty} \frac{E[\delta e^{x\delta}\mathbb{I}(\delta \geq c)]}{E[e^{x\delta}\mathbb{I}(\delta < c) + E[e^{x\delta}\mathbb{I}(\delta \geq c)]}.$$

The above and the fact that $cE[e^{x\delta}\mathbb{I}(\delta \geq c] < E[\delta e^{x\delta}\mathbb{I}(\delta \geq c] < bE[e^{x\delta}\mathbb{I}(\delta \geq c]$ imply that

$$c \leq \lim_{x\to\infty} g'(x) \leq b.$$

Because c is arbitrary and the limit cannot depend on c, we find that the limit is b. □

1.9.2 Properties of the risk function

Now I present the properties of the risk function $\rho_a(x)$. In all derivations I will assume that $x \geq 0$. The relevant derivations for the case $x \leq 0$ can be obtained analogously.

I will denote

$$\alpha = \frac{1}{1-a}.$$

I will also maintain the assumption

$$a \geq 0,\ a \neq 1$$

throughout this section which clearly holds for the cases $a = 0$ and $a = 1 + \nu$ which are of interest for the results of the paper. The case $0 \leq a < 1$ corresponds to $\alpha \geq 1$, whereas the case $a > 1$ corresponds to $\alpha < 0$.

Fact 3. *The risk function $\rho_a(x)$ satisfies the ODE*

$$x\rho_a'(x) = -\alpha\rho_a(x) + \alpha g''(-\gamma x). \tag{1.9.3}$$

Proof. Write the definition of the risk function

$$\rho_a(x) = \int_0^1 g''(-t^{1-a}\gamma x)dt.$$

Make a change of variable $t = (y/x)^{\alpha}$, $dt = \alpha\,(y/x)^{\alpha-1}\,\frac{dy}{x}$. The above becomes

$$\rho_a(x) = \begin{cases} \alpha x^{-\alpha} \int_0^x g''(-\gamma y) y^{\alpha-1} dy, & \alpha \geq 1 \\ -\alpha x^{-\alpha} \int_x^{\infty} g''(-\gamma y) y^{\alpha-1} dy, & \alpha < 0 \end{cases} \tag{1.9.4}$$

Differentiating the above with respect to x one gets (1.9.3). □

Fact 4. Provided that $\alpha \neq -n$, the n-th derivative of $\rho_a(x)$ is continuous at zero and is given by

$$\rho_a^{(n)}(0) = \frac{(-\gamma)^n \alpha}{n+\alpha} g^{(n+2)}(0).$$

Proof. Rewrite (1.9.3) as

$$\rho_a'(x) = -\alpha \left(\frac{\rho_a(x) - \sigma^2}{x} + \gamma \frac{g''(-\gamma x) - \sigma^2}{-\gamma x} \right).$$

Calculating the limit of the above as $x \to 0$ we get

$$\rho_a'(0) = -\alpha \rho_a'(0) - \alpha \gamma g'''(0).$$

The above implies

$$\rho_a'(0) = \frac{-\gamma \alpha}{1+\alpha} g'''(0).$$

So if $\alpha \neq -1$ $\rho_a'(x)$ is continuous at zero.

To get the second derivative at zero differentiate (1.9.3)

$$x \rho_a''(x) = -(1+\alpha)\rho_a'(x) - \gamma \alpha g'''(-\gamma x).$$

Rewrite the above as

$$\rho_a''(x) = -(1+\alpha) \left(\frac{\rho_a'(x) - \rho_a'(0)}{x} - \frac{\gamma^2 \alpha}{1+\alpha} \frac{g'''(-\gamma x) - g'''(0)}{-\gamma x} \right).$$

Taking the limit as $x \to 0$ we get

$$\rho_a''(0) = -(1+\alpha)\rho_a''(0) + \gamma^2 \alpha g^{(4)}(0).$$

We get

$$\rho_a''(0) = \frac{\gamma^2 \alpha}{2+\alpha} g^{(4)}(0).$$

Again, if $\alpha \neq -2$ $\rho_a''(x)$ is continuous at zero. One can get the general formula by induction:

$$\rho_a^{(n)}(0) = \frac{(-\gamma)^n \alpha}{n+\alpha} g^{(n+2)}(0).$$

□

Fact 5. *Suppose δ has bounded support $[a, b]$. Then for $a > 1$: $\rho_a(x) < \infty$ and*

$$\rho_a(x) = o(1/x) \text{ as } x \to \pm\infty.$$

Proof. First I prove that $\rho_a(x) < \infty$. According to (1.9.4) for $a > 1$ (i.e. $\alpha < 0$) we can write the risk function as

$$\rho_a(x) = -\alpha x^{-\alpha} \int_x^\infty g''(-\gamma y) y^{\alpha-1} dy.$$

Integrating by parts one gets

$$\rho_a(x) = \frac{-\alpha}{\gamma x}\left(g'(-\gamma x) - (\alpha-1)\, x^{1-\alpha} \int_x^\infty g'(-\gamma y) y^{\alpha-2} dy\right). \tag{1.9.5}$$

Rearranging the above I get

$$\rho_a(x) = \frac{\alpha(\alpha-1)}{\gamma}\left(\int_1^\infty \left(\frac{g'(-\gamma x) - g'(-\gamma x t)}{x}\right) t^{\alpha-2} dt\right).$$

Since $0 < g'(-\gamma x) - g'(-\gamma x t) < b - a$ (Fact 1 and Fact 2) we have that $\rho_a(x) < -\frac{\alpha}{\gamma}\frac{b-a}{x} < \infty$ for $x \neq 0$. To calculate $\rho_a(x)$ at zero we use the Dominated Convergence Theorem to interchange the limit and integration and L'Hopital's rule to calculate $\lim_{x\to 0} \frac{g'(-\gamma x) - g'(-\gamma x t)}{x} = -\gamma\sigma^2(1-t)$. Calculating $\alpha(\alpha-1)\left(\int_1^\infty \sigma^2\,(t-1)\, t^{\alpha-2} dt\right)$ yields σ^2 which is finite (since by assumption the CGF exists for all $x \in \mathbb{R}$.

Now I prove that $\rho_a(x) = o(1/x)$ *as* $x \to +\infty$.

We can write $\lim_{x\to\infty} x\rho_a(x) = \frac{\alpha(\alpha-1)}{\gamma}\left(\int_1^\infty \lim_{x\to\infty}\left(g'(-\gamma x) - g'(-\gamma x t)\right) t^{\alpha-2} dt\right) = 0$, where the Dominated Convergence Theorem was used to interchange limit and integration and Fact 2 to calculate $\lim_{x\to\infty} g'(-\gamma x) = \lim_{x\to\infty} g'(-\gamma x t) = a$. □

Fact 6. *The risk function satisfies the following comparative statics results:*

1) $\rho_a(x)$ is homogenous of degree 0 in $(\gamma, 1/x)$;

2) *For* $a' > a > 1$ *and* $a' < a < 1$ $1/\alpha' \cdot \rho_{a'}(x) < 1/\alpha \cdot \rho_a(x)$, *where* $\alpha' = \frac{1}{1-a'}$ *and* $\alpha = \frac{1}{1-a}$;

3) *For two distributions* A *and* B *satisfying* $g''_A(t) > g''_B(t)$ $\rho_a^A(x) > \rho_a^B(x)$.

Proof. Write the risk function as

$$\rho_a(x) = \begin{cases} \alpha \int_0^1 g''(-\gamma y x) y^{\alpha-1} dy & , \text{ if } \alpha > 0; \\ -\alpha \int_1^\infty g''(-\gamma y x) y^{\alpha-1} dy & , \text{ if } \alpha < 0. \end{cases} \tag{1.9.6}$$

The results are easy to verify from the above. □

1.10 Equilibrium bids are strictly decreasing and have a finite slope

Since we do not know whether the residual supply is invertible, we write the residual supply curve as

$$x = R(p; s).$$

The residual supply is given by

$$R(p; s) = s - (L-1)b(p), \tag{1.10.1}$$

where $b(p)$ denotes the equilibrium bid.

The ex-post maximization problem $\mathcal{P}_{EP}$ can be written as

$$\max_{x,p} f(x) - p \cdot x$$

$$s.t.: \ x = R(p; s).$$

Substituting the constraint into the objective and taking the first order condition with respect to p yields the following equation determining optimal price $p^*(s)$ on the residual supply curve for a given realization of s

$$f'\left(R\left(p^*; s\right)\right) R_p\left(p^*; s\right) - R\left(p^*; s\right) - p^* \cdot R_p\left(p^*; s\right) = 0. \tag{1.10.2}$$

The above equation determines the optimal price $p^*(s)$ for a given realization of s. The corresponding optimal quantity x^* is given by $x^* = R(p^*; s)$. In the

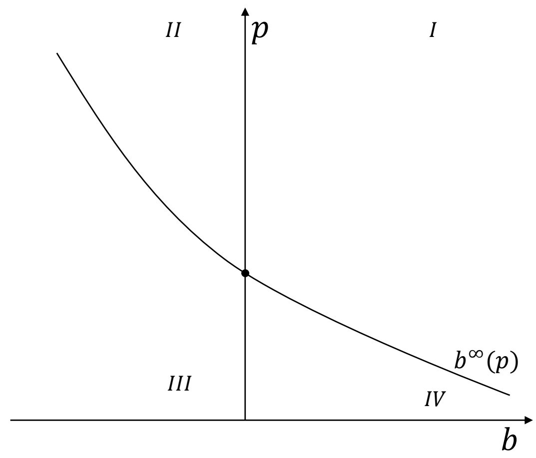

Figure 1.10.1: Possible trajectories originating in the II and IV quadrants.

symmetric equilibrium $x^* = b(p^*)$ (all traders get the same quantity). It follows that $R(p^*; s) = b(p^*)$. Substituting it to (1.10.2), noting that $R_p(p^*; s) = -(L-1)b'(p^*)$ (follows directly from (1.10.1)) and denoting $p = p^*(s)$ we rewrite (1.10.2), after some rearrangement, as

$$b'(p) = \frac{b}{(L-1)(p - f'(b))}. \tag{1.10.3}$$

The nonlinear ordinary differential equation above is a necessary condition for equilibrium bid. The analysis that follows is analogous to Klemperer and Meyer (1989).

The lines $b = 0$ and $b = b^{\infty}(p)$ solving $p - f'(b^{\infty}) = 0$ divide the (b, p) plane into four quadrants numbered as in the figure below.

Proposition C1. *The equilibrium bid $b(p)$ satisfies$-\infty < b'(p) < 0$.*

Proof. The condition $-\infty < b'(p) < 0$ is equivalent to saying that the bid lays within the second and fourth quadrants.

Suppose, on the contrary, that the equilibrium bid passes through the first quadrant. Then the price the trader will pay for some realizations of the supply will be above marginal utility $f'(x)$ which cannot be optimal because in that case the trader is strictly better off by submitting $f'(\cdot)$ for the corresponding supply realisations. The bid that passes through the first quadrant cannot be optimal for the same reason. The bid that satisfies $b'(p) = \infty$ is the bid that intersects with the

$b = b^{\infty}(p)$ locus. Such a bid eventually reaches either first or third quadrant and so cannot be optimal. Finally, as follows from (1.10.3) the only bid for which $b'(p) = 0$ is $b = 0$ which is not optimal since the consumer surplus is zero and the deviation to any bid that passes through second and fourth quadrants is profitable. $\square$

Chapter 2

Liquidity vs Information Efficiency

2.1 Introduction

In many modern markets, traders are heterogeneous along the following two dimensions. The first dimension is the price impact: there are large traders who are able to move prices and small traders whose effect on prices is negligible. For example, in financial markets there is evidence that large institutional investors (such as hedge, mutual and pension funds) have considerable price impact[1]. No such evidence exists for retail investors and smaller funds, and anecdotally, price impact is not an issue for these types of investors. The second dimension is heterogeneity in values. For example, in financial markets institutional and retail investors may have different values of an asset due to different trading needs or tax or risk-management considerations[2]. In this paper, I present a model that captures this heterogeneity and show that such heterogeneity has unexpected consequences for liquidity, information efficiency and welfare.

I consider a centralized market (modeled as a uniform-price double auction) populated by *large* and *small* traders. To capture the heterogeneity in price impacts, I assume that there is a countable number of large traders, whereas small traders form a continuum. The traders within each group are identical. I employ a linear-normal setting: traders have linear-quadratic objectives, and their values are distributed normally. To capture the second dimension of heterogeneity, I assume that the val-

[1]See, e.g., Chan and Lakonishok (1995), Keim and Madhavan (1995), Korajczyk and Sadka (2004), among others.

[2]Fund flows and fund managers' compensation relative to a benchmark can be conceptualized as endowment shocks. These endowment shocks create hedging needs that are specific to institutional investors. See Vayanos and Woolley (2013) for a treatment of the effect of fund flows. See Basak and Pavlova (2012) and Cuoco and Kaniel (2011) for a treatment of benchmarking.

ues of large and small traders are imperfectly correlated. For simplicity, I assume that the large traders know their value. However, for the information efficiency to play a role, the information concerning small traders value is dispersed among them. I show that the model provides a natural framework for considering asset, commodities, foreign exchange and product markets.

In my first set of results, I consider the interaction among liquidity, information efficiency and welfare. First, I show that a tension between liquidity and information efficiency might arise: policy measures intended to promote liquidity might be harmful for information efficiency and vice versa and changes in the market environment (such as risk-bearing capacity, number of large traders, information precision) can shift liquidity and information efficiency in opposite directions. Second, I show that a shock to the economic environment that has a positive direct effect on liquidity (an increase in risk-bearing capacity) may have a negative overall effect on liquidity (*liquidity paradox*). This is possible because the shock has a positive effect on information efficiency and there is a tension between the two. Similarly, a positive shock to information efficiency (an increase in the precision of the signals) might have a negative overall effect on information efficiency (*information aggregation paradox*). Third, when there is more competition between large traders, welfare might be lower. Moreover, all traders, even small ones, can be worse off as a result of more competition. This is possible because competition has negative effects on information efficiency. For a similar reason, breaking up a centralized market into two separate exchanges might improve welfare.

The above results are a consequence of an equilibrium mechanism that features a complementarity between illiquidity (price impact) and information efficiency. The mechanism is represented in Figure 2.1.1. A belief that the market is less liquid induces large investors to trade less aggressively (their demand is less sensitive to their information). It makes the price relatively less (more) sensitive to the values of large (small) traders. From the perspective of small traders, the price is more informative. Therefore, they provide less liquidity: if someone is buying and driving up the price, small traders are less willing to sell (decrease their demands) because they partly attribute higher prices to stronger fundamentals. In other words, when prices are more informative, small traders are less price-elastic. The latter confirms lower liquidity.

A direct consequence of such complementarity is the possibility of multiple equi-

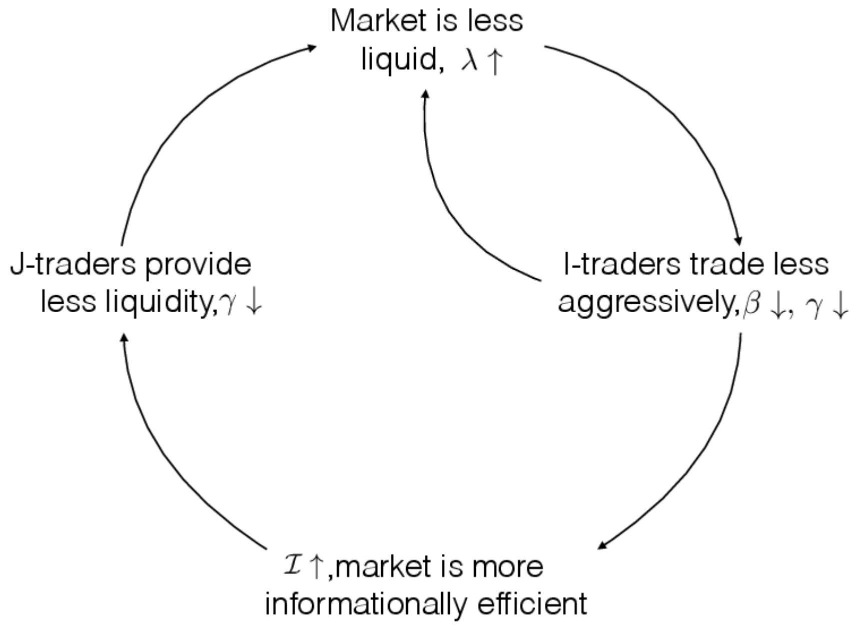

Figure 2.1.1: Equilibrium mechanism. I-traders are large. J-traders are small.

libria driven by self-fulfilling beliefs concerning liquidity or information efficiency. I provide the sufficient conditions for the multiplicity to emerge. I show that the equilibria can be ranked in terms of liquidity and information efficiency and that the rankings are the opposite of one another: the equilibria with higher liquidity feature lower information efficiency and vice versa. I also provide a sufficient condition under which the equilibria can be ranked in terms of welfare: if the price does not provide much incremental information to the traders, the equilibria with higher liquidity are those with greater welfare.

I also explore the implications of the mechanism for market crashes. I understand the latter either as a switch between the two equilibria with different price levels or as a large change in price caused by a small change in the economic environment. The latter is possible because the complementarities provide a natural amplification mechanism. I show that, depending on whether the large traders are on the buy or sell side of the market, there are two scenarios consistent with a market crash, which differ in the behavior of information efficiency, liquidity, volatility, and trading volume. Under the sufficient condition that price does not provide substantial incremental information, welfare decreases in only one scenario. Correspondingly, only one scenario suggests policy intervention.

I consider the implications of the model and empirical evidence in Section 2.7. Briefly, I consider two episodes that affected commodities markets, the 2008 boom/bust and the 2014 crash in oil prices, through the lens of the model. I emphasize the role of two forces: (1) informational frictions and (2) the market power of oil producers that is endogenously amplified because of (1). In asset markets, I seek evidence supporting the model's prediction that in a more liquid market, the price is more (less) reflective of the values of large (small) traders. I find suggestive evidence in the on-the-run treasury bonds and equity markets. I also discuss the policy implications concerning the effects of high-frequency traders and commodity index traders in the in asset and commodities markets and discuss the effects of competition on welfare.

On a technical side, I demonstrate how to perform a stability analysis in a strategic trading model with heterogeneous traders. The key idea is to represent the equilibrium as a fixed point that determines market liquidity. Given their beliefs concerning market liquidity, traders choose their demand schedules. In equilibrium, liquidity (which is determined by the slopes of the traders' demands) should be equal to assumed liquidity. The stability of equilibrium is associated with the stability of this fixed point.[3] This representation also allows me to characterize quantitatively the amplification through an *illiquidity multiplier*.[4]

Related literature This paper is related to two strands of literature: strategic trading/supply function equilibria and rational expectations models featuring multiple equilibria.

The first strand of literature can be further divided into two subgroups: the models with common values (Kyle (1989), Pagano (1989), Vayanos (1999), Rostek and Weretka (2015), and Malamud and Rostek (2015)) and the models with private values (Vives (2011), Rostek and Weretka (2012, 2014), Du and Zhu (2015), Kyle, Obizhaeva and Wang (2015), and Babus and Kondor (2015)). Technically, the common value models obviously lack the heterogeneity in trader' values, which I capture in my model. More important, given common values, the interaction between liquidity and information efficiency is in the opposite direction. In common value models, the price reflects traders' information and noise. If traders believe that the market is more liquid, they trade more aggressively on their information,

[3]The representation simplifies the stability analysis significantly, as mapping liquidity onto itself entails mapping $\mathbb{R}$ onto $\mathbb{R}$, whereas the best response mapping is $\mathbb{R}^4$ onto $\mathbb{R}^4$ in my model.

[4]The notion of an illiquidity multiplier is from Cespa and Foucault (2014). The idea of representing the equilibrium as a fixed point determining the price impact is from Weretka (2011) and Rostek and Weretka (2015).

and information efficiency improves. Consequently, the complementarity between illiquidity and information efficiency does not arise.

The private values model of Vives (2011), Rostek and Weretka (2012, 2014), Du and Zhu (2015) and Kyle, Obizhaeva and Wang (2015) capture the heterogeneity in traders' values; however, they focus on symmetric settings and there is no heterogeneity in price impact.[5] As a result, traders' behavior is affected by liquidity in a symmetric way, and the price reflects the same combination of their signals. Consequently, the complementarity uncovered in this paper does not arise. Babus and Kondor (2015) include the two dimensions of heterogeneity in their model. However, they focus on the over-the-counter markets, and the complementarity does not arise because of the bilateral interactions among the large traders.

The multiplicity of equilibria in REE models can arise for two reasons. First, due to demand nonlinearities, there can be multiple market-clearing prices (e.g., Gennotte and Leland (1990); Barlevy and Veronesi (2003); Yuan (2005)). In contrast, the equilibrium in this model is linear, and consequently, the market-clearing price is always unique.

The equilibrium multiplicity in this paper arises due to strategic complementarities, similar to Ganguli and Yang (2009), Goldstein, Li and Yang (2013), Cespa and Focault (2014), Cespa and Vives (2015), Rohi and Zigrand (2015), Huang (2015) and Bing et al. (2016). In these papers, the traders take prices as given, whereas the traders in my model account for their influence on prices. This difference is not merely technical: strategic behavior on the part of large traders is an integral component of the mechanism generating the complementarity in this paper. Moreover, price-taking behavior implies that traders regard the market as perfectly liquid; therefore, as my focus in the paper is liquidity, assuming the strategic behavior is desirable.

Through their focus on liquidity, the two most closely related papers among the above REE models with complementarities are Cespa and Foucault (2014) and Cespa and Vives (2015). These models also feature multiple equilibria that differ in liquidity and information efficiency. However, in these papers, the equilibria with higher liquidity are also those with higher information efficiency, which highlights the complementarity between *liquidity* and information efficiency (versus comple-

[5]Technically, there are small traders in Vives (2011). However, their behavior is not affected by either liquidity or information. The model predictions are the same if instead of small traders the model postulates an exogenously postulated demand curve.

mentarity between *illiquidity* and information efficiency in this paper).

The tension between liquidity and information efficiency can manifest as a comparative statics result in some other settings. For example, in Subrahamnyam (1991), increasing the variance of noise trading can increase liquidity but decrease information efficiency. However, in my paper this tension manifests through the potential coexistence of high liquidity/low efficiency and low liquidity/high efficiency equilibria. Bing et al. (2016) demonstrate that there might be a tension between the liquidity and information efficiency if noise traders chase liquidity: improvement in liquidity attracts more noise traders and may therefore harm the information efficiency. In my paper it is more aggresive trading, not the entry of traders, that have adverse effects on information efficiency.

The information aggregation paradox is reminiscent of the results of Banerjee et al. (2015), who show that reducing the cost of information acquisition (and, therefore, increasing signal precision in equilibrium) may not increase information efficiency. In their model, the traders may acquire information on asset fundamentals or on noise. They show that lowering the cost of information concerning the fundamentals may, under certain conditions, induce traders to learn more about noise. As a result, information efficiency may decrease. This mechanism differs from that in my paper, whereby more precise information improves liquidity and induces large traders to trade more, which is harmful for the price inference of small traders and, consequently, for information efficiency.

Rostek and Weretka (2014) show that increasing market size (the number of traders) does not necessarily increase welfare. They consider an equicommonal auction: a market with large traders who are heterogenous in their values, such that the average correlation of the value of each trader with the values of others is the same for all traders. They attribute the reduction in welfare to a decrease in gains from trade: in larger equicommonal markets, the values of traders are more aligned and, correspondingly, the gains from trade are lower. This mechanism is therefore different from that presented here, which emphasizes the negative externality that increased competition has on information efficiency.[6]

[6]In an equicommonal auction, the price always reflects the average of traders' signals. In contrast, in my model, the price is less (more) reflective of the value of small (large) traders when the competition among large traders increases.

2.2 The model

Consider a market for a divisible good in which two groups of agents, I and J, are trading. There are $N > 1$ of I-traders, indexed by $i \in I \equiv \{1, 2, ..N\}$, and there is a unit continuum of J-traders, indexed by $j \in J \equiv [0, 1]$. The traders within each group $k \in \{I, J\}$ are identical, and their preferences are given by a quasilinear-quadratic function

$$u_k = (v_k - p)\, x - \frac{w_k x^2}{2}, \tag{2.2.1}$$

where $w_k > 0$ is a constant, and the values $v_k \sim N\left(\bar{v}_k, \frac{1}{\tau_k}\right)$ are jointly normally distributed with

$$\mathrm{corr}(v_I, v_J) = \rho \in (-1, 1). \tag{2.2.2}$$

The information structure is as follows. The I-traders know their value, but it is not known to J investors. The $J-$investors have dispersed information about their value. Each $j \in J$ receives a signal

$$s_j = v_J + \epsilon_j, \tag{2.2.3}$$

where $\epsilon_j \sim N\left(0, \frac{1}{\tau_s}\right)$, and for any $j, k \in J$, such that $k \neq j$ the noise ϵ_j is independent of v_I, v_J and ϵ_k. The parameter τ_s measures the precision of the signal. The information structure can be summarized by the information sets

$$\mathcal{F}_i = \{v_I\},\ \mathcal{F}_j = \{s_j\},\ \forall i \in I,\ j \in J.$$

In equilibrium, traders will also learn from prices.

The market is modeled as a uniform-price double auction. Each trader k submits his net demand schedule $x_k(p)$: $x_k(p) > 0$ ($x_k(p) < 0$) corresponds to a buy (sell) order. The market-clearing price p^* is such that the net aggregate demand is zero

$$\sum_{i=1}^{N} x_i\left(p^*\right) + \int_0^1 x_j\left(p^*\right) dj = 0. \tag{2.2.4}$$

In equilibrium, each trader is allocated

$$x_k^* = x_k\left(p^*\right).$$

The equilibrium concept is a symmetric linear Bayesian Nash Equilibrium (henceforth, equilibrium). A symmetric linear equilibrium is an equilibrium in which traders $i \in I$ and $j \in J$ have the following demand schedules

$$x_i = \alpha + \beta \cdot v_I - \gamma \cdot p \text{ and } x_j = \alpha_J + \beta_J \cdot s_j - \gamma_J \cdot p. \tag{2.2.5}$$

2.2.1 Examples

Below, I show that the model presented above provides a natural framework for considering at least four types of markets.

1. Securities markets.

In this example, the good being traded is a financial asset, such as a bond or stock. The I-traders are institutional investors. In the model, their distinguishing features are that they are large (can affect prices), and sophisticated/informed (know their value). It is therefore natural to interpret them in this manner.[7] The J-traders can be interpreted as retail investors.

The preference specification (2.2.1) is common in the securities markets context.[8] The quadratic component $\frac{w_k x^2}{2}$ represents an inventory cost that may come from the regulatory capital requirements, collateral requirements or risk-management considerations.[9] The difference in the values of I- and J-traders may be due to the following reasons. The first is that along with the common value component v, representing the fundamental value of the security, investors may also care about a private value u_k, such that

$$v_k = v + u_k,\ k \in \{I, J\}.$$

The private values u_k, which differ between the two groups, may be due to different tax or risk-management considerations.[10] Assuming that v, u_I and u_J are normally distributed and not perfectly correlated, we obtain the setup with imperfectly correlated values described in the section above.

[7] There is a vast empirical literature demonstrating that institutional investors have price impact and that the costs associated with it are considerable. Examples include Chan and Lakonishok (1995), Keim and Madhavan (1995), and Korajczyk and Sadka (2004), among others. Anecdotally, institutional investors are more informed because they have more resources to support a larger research division, pay for relevant data streams, etc. Hendershott et al. (2015) present empirical evidence supporting this point.

[8] E.g., Vives (2011), Rostek and Weretka (2012) and Du and Zhu (2015).

[9] See Du and Zhu(2015), Section 2.1 for a discussion.

[10] See, e.g., Duffie, Garleanu and Pedersen (2005) or Du and Zhu (2015) for a discussion of private values in the context of financial markets.

An alternative explanation is that the difference in v_I and v_J may represent uncertainty concerning the endowment shocks. Suppose that both types of investors care about the fundamental value of the security v. Suppose that I-investors receive a (normally distributed) endowment shock e that is known to them but unknown to J-investors. The J-investors receive no endowment shocks. In that case, the preference relation of I-investors can be written as $(v-p)\,x - \frac{w_I(e+x)^2}{2}$, or, dropping the constant $\frac{w_I e^2}{2}$,

$$(v - w_I e - p)\,x - \frac{w_I x^2}{2}.$$

Denoting $v_I \equiv v - w_I e$, the above becomes

$$(v_I - p)\,x - \frac{w_I x^2}{2},$$

which is consistent with the specification (2.2.1). Moreover, as long as e is not perfectly correlated with v, $v_J = v$ and $v_I = v - w_I e$ are imperfectly correlated, which is consistent with the setup described above.

2. Commodities or intermediate good markets.

In this example, the good being traded is a commodity, such as crude oil or aluminum. More generally, imagine any intermediate good, i.e., one that is an output for some firms while being an input for the others. The I-traders are commodity *producers*. The J-traders are *firms*, buying the commodity to produce the final good.

Commodity producers have a production technology characterized by a convex cost function

$$c \cdot y + \frac{w_I}{2} y^2, \tag{2.2.6}$$

where $c \sim N\left(\bar{c}, \frac{1}{\tau_I}\right)$ is a cost shock, which is known to producers but not to firms. The latter assumption captures that the producers are better informed about their own production technology. Producers are risk neutral and maximize their profit

$$p \cdot y - \left(cy + \frac{w_I}{2} y^2\right). \tag{2.2.7}$$

Note that in the above, y is the amount of commodity sold, i.e., the net supply. The net demand of producers is $x = -y$. With this change of variable, the above becomes

$$(c-p)\,x-\frac{w_I}{2}x^2, \tag{2.2.8}$$

which is consistent with (2.2.1) with a value v_I equal to the cost shock c.

Firms $j\in[0,1]$ have a production technology characterized by a concave production function

$$Y(x)\equiv a\cdot x-\frac{w_J}{2}x^2. \tag{2.2.9}$$

In the above, $a\sim N\left(\bar{a},\frac{1}{\tau_a}\right)$ is a productivity shock. The latter shock is common to all firms. The firms have dispersed information concerning a. In particular, each firm j is endowed with a signal

$$s_j=a+\epsilon_j,$$

where $\epsilon_j\sim N\left(0,\frac{1}{\tau_s}\right)$ and $\forall j,k$, such that $k\neq j$ the noise ϵ_j is independent of c, a and ϵ_k. Following Sockin and Xiong (2015), the productivity shock can be interpreted as the strength of the economy. Firms are risk neutral and maximize their profit

$$p_g\left(a\cdot x-\frac{w_J}{2}x^2\right)-p\cdot x, \tag{2.2.10}$$

where $p_g=1$ is the price of the final good (endogenized below) and p is the price of the commodity. The above is consistent with (2.2.1) with the value v_J equal to the productivity shock a.

I close the model and assume that the final good is sold to consumers $l\in[0,1]$, who have a linear Marshallian utility function over consumption of the final good z and residual money $m=m_0-p_gz$

$$u_l(z,m)=z+m_0-p_gz,$$

where m_0 is the endowment of money that each consumer has. The fact that there is a continuum of them implies that they are price takers. Therefore the price of the final good is equal to the marginal utility and, indeed, $p_g=1$.

The setting considered in this example is a natural framework to study commodities markets. The linear-quadratic specification of the cost and production functions is common in the commodities literature.[11] The information structure with a cost shock known to producers but not to firms and firms having dispersed information

[11]E.g., Grossman (1977), Kyle (1984), Stein (1987), Goldstein and Yang (2015).

regarding the strength of the economy is the same as in Sockin and Xiong (2015), with an additional generality of allowing for correlation between c and a. The setting of this example can be considered a generalization of Sockin and Xiong (2015), in which I allow producers to have market power.[12]

3. Product markets.

This example is similar to the previous one, but the good being traded is a final good. The $I-$traders are the producers of that good. They have a cost function (2.2.6), with the same assumptions regarding the cost shocks distribution. The only difference is that J-traders are now consumers with concave utility

$$v_J x - \frac{w_J x^2}{2} - px.$$

The parameter v_J is interpreted as a quality of the product, and the consumers have dispersed information on quality in the form of signals (2.2.3). With cost c and quality v_J being imperfectly correlated, the example conforms to the setting presented in the section above.

4. Foreign exchange markets.

In this example, the good being traded is foreign currency. Suppose that the home currency is the pound and the foreign currency is the dollar. The price p is how many pounds one dollar is worth. The I-traders are exporters. The J-traders are importers. Exporters receive dollars from selling their goods. Importers need to buy dollars to purchase raw materials abroad. The supply and demand from those two groups determine the exchange rate.

The price of the good that the I-traders produce and export is denominated in dollars, and the exporters have no ability to influence it. Normalize it to one. Assume that the cost of production of y units of the export good is given by (2.2.6). The revenue from selling y units of the good is y dollars and $p \cdot y$ pounds. Therefore, the profit from a sale of y units (corresponding to the net demand of $x = -y$) is given by (2.2.8), just as in Example 2.

[12]The market power of producers is clearly relevant in commodities markets. E.g., in the crude oil market, OPEC accounts for more than 40% of world production (OPEC statistical bulletin (2015)); in the aluminum market, the 6 largest producers account for over 40% of world production (Nappi (2013)). Such concentration should not be surprising, and the possible reasons for it are twofold. First, for the energy and metals commodity classes, commodity-producing firms are typically monopolies in their home countries. Because there are few large commodity-producing countries, there are few large producers in the world. Second, even if there are many producers in a country (which is the case for agricultural commodities, for example) their actions in the global market are nevertheless orchestrated by their home governments through export quotas and tariffs.

The J-traders need to import raw materials, the price of which is denominated in dollars and normalized to one, similar to the above. The cost of buying x units of raw materials is therefore x dollars and $p \cdot x$ pounds. With x units of raw materials, the importers can produce $Y(x)$ units of the good, where the production function $Y(x)$ is given by (2.2.9). The price of the good that the importers produce is denominated in pounds and normalized to one. The profit from selling x units of the good is therefore given by (2.2.10), just as in Example 2. The mapping to the general framework can therefore be established in the same way as in Example 2.

2.3 Equilibrium

In this section, I characterize the equilibrium in the model. I restrict myself to the case

$$\rho \geq 0. \tag{2.3.1}$$

This is a reasonable assumption in the securities,[13] commodities[14] and product markets.[15] However, my main motivation to introduce it is to simplify exposition. The model with negative correlation is still tractable but exhibits additional complementarities. To focus on the main mechanism, I consider the case (2.3.1). Theorem 3 characterizes the equilibria.

Theorem 3. *There exists at least one equilibrium. The closed-form expressions, up to a solution of a sextic equation, for the equilibrium coefficients* $(\alpha, \beta, \gamma, \alpha_J, \beta_J, \gamma_J)$ *are given by the equations (2.10.24-2.10.29) in the Appendix. The equilibrium is unique if*

$$\tau_I < \overline{\tau}_1. \tag{2.3.2}$$

Suppose that

$$w_I < \overline{w}, \;\; N > 4. \tag{2.3.3}$$

Then there exist thresholds $\underline{\tau_2}$ *and* $\overline{\tau_2}$ *such that* $\overline{\tau}_1 < \underline{\tau_2} < \overline{\tau_2}$*, and there are at least three equilibria if*

[13]Indeed, under the traditional pure common value setup, the correlation is equal to one. If the departure from the pure common values is not too substantial, the correlation should still be positive.

[14]It is more general than the assumption of zero correlation of demand and supply shocks in Sockin and Xiong (2015) and Goldstein and Yang (2015).

[15]Indeed, the fact that a particular product is more expensive to produce is usually associated with that product having better quality.

$$\underline{\tau_2} < \tau_I < \overline{\tau_2}. \tag{2.3.4}$$

The closed-form expressions for the thresholds are given by equations (2.10.45, 2.10.53-2.10.55) in the Appendix.

I present a detailed proof of the above theorem in the Appendix. Below I provide the most important steps.

Consider I-traders. They choose their demand schedules to maximize $(v_I - p)\, x - \frac{w_I}{2} x^2$. The first-order condition is given by

$$v_I - p - x \frac{\partial p}{\partial x} - w_I x = 0,$$

where the third term $x \frac{\partial p}{\partial x}$ reflects the fact that the I-traders realize that they can move prices.

In equilibrium, the price sensitivity $\frac{\partial p}{\partial x}$ is given by the slope of the inverse residual supply (Kyle's lambda) $\frac{\partial p}{\partial x} = \lambda$, where

$$\frac{1}{\lambda} = (N-1)\gamma + \gamma_J. \tag{2.3.5}$$

The above expression is intuitive: $1/\lambda$ is the slope of the (direct) residual supply function, and there are $(N-1)$ I-traders with supply elasticity γ and a unit mass of J-traders with demand elasticity γ_J contributing to it.

In what follows, I will refer to λ as a *price impact* and $1/\lambda$ as *liquidity*. Equation (2.3.5) provides the first takeaway: liquidity is directly related to price elasticities γ and γ_J. This enables me to use the following language: *if a trader increases (decreases) his price elasticity*, I say that *he provides more (less) liquidity*.

The above implies that the demand of the I-traders is given by

$$x_i = \frac{1}{w_I + \lambda}(v_I - p), \tag{2.3.6}$$

from which it follows, in particular, that

$$\beta = \gamma = \frac{1}{w_I + \lambda} > 0, \tag{2.3.7}$$

where λ is given by (2.3.5). The above equation provides the second takeaway: a higher price impact implies that I-traders trade less aggressively (β is lower) and provide less liquidity (γ is lower).

As there is a continuum of J-traders, they cannot move prices. Their optimization problem is given by

$$\max_x \left(E[v_J|s_j,p]-p\right)x-\frac{w_Jx^2}{2},$$

implying an optimal demand of

$$x_j=\frac{1}{w_J}\left(E\left[v_J|s_j,p\right]-p\right). \tag{2.3.8}$$

It remains to understand the inference problem of the J-traders. In a linear equilibrium given by (2.2.5), the equilibrium price function is

$$p=\frac{1}{\Gamma}\left(N\beta v_I+\beta_J v_J\right)+c_p, \tag{2.3.9}$$

where $\Gamma=N\gamma+\gamma_J$ is a price elasticity of aggregate demand and c_p is a constant.[16]

The values v_I and v_J are positively correlated, and hence without loss of generality, we may assume that

$$v_I=A+Bv_J+C\epsilon,$$

where $\epsilon\sim N(0,1)$ is independent of v_J and A, $B\geq 0$ and $C>0$ are some constants.[17]

Substituting the above into (2.3.9), one can see that the price is informationally equivalent to the following sufficient statistic

$$\pi\equiv\frac{\Gamma p}{N\beta B+\beta_J}+\text{const}=v_J+\frac{N\beta C}{N\beta B+\beta_J}\epsilon,$$

where in the above and in what follows, I denote by const non-stochastic terms. Because ϵ and v_J are independent, the sufficient statistic π is an unbiased signal of v_J. This signal has a precision

$$\tau_\pi\equiv\mathrm{Var}[\pi|v_J]^{-1}=\frac{1}{C^2}\left(B+\frac{\beta_J}{N\beta}\right)^2>\left(\frac{B}{C}\right)^2. \tag{2.3.10}$$

[16]The exact value is $c_p=\frac{N\alpha+\alpha_J}{\Gamma}$.

[17]It is easy to express A, B and C through the parameters of the model. One can find $B=\rho\sqrt{\frac{\tau_J}{\tau_I}}$, $C=\sqrt{\frac{1-\rho^2}{\tau_I}}$ and $A=\bar{v}_I-B\bar{v}_J$.

From the Projection Theorem, the ex-post precision of v_J, measuring how much the J-traders can learn about their values, is

$$\tau = \text{Var}[v_J|s_j, p]^{-1} = \tau_J + \tau_s + \tau_\pi.$$

Define *information efficiency* as:[18]

$$\mathcal{I} \equiv \frac{\text{Var}(v_J)}{\text{Var}(v_J|s_j, p)} = \frac{\tau_J + \tau_s + \tau_\pi}{\tau_J}.$$

This reveals the third takeaway: less aggressive trading by I-traders (lower β) makes the price more informative for J-traders (greater τ_π). Because the J-traders are the only ones who learn, the information efficiency of the market improves ($\mathcal{I}$ increases).

From the Projection Theorem, one can compute

$$\begin{aligned} E\left[v_J|s_j, p\right] &= \frac{\tau_s}{\tau} s_j + \frac{\tau_\pi}{\tau}\pi + \text{const} \\ &= \frac{\tau_s}{\tau} s_j + \frac{\tau_\pi}{\tau}\frac{\Gamma p}{(N\beta B + \beta_J)} + \text{const}. \end{aligned}$$

Substituting the above into (2.3.8) and comparing to (2.2.5) yields

$$\beta_J = \frac{1}{w_J}\frac{\tau_s}{\tau} > 0, \tag{2.3.11}$$

and, after some rearrangement,

$$\gamma_J = \frac{1}{w_J} - \frac{\Gamma}{\tau_s}\sqrt{\tau_\pi}\left(\sqrt{\tau_\pi} - \frac{B}{C}\right). \tag{2.3.12}$$

Intuitively, there are two effects determining the elasticity γ_J. The first is the expenditure effect: for a higher price, a trader would demand less because a higher price implies higher expenditure $p \cdot x$ from buying x units of the good. This effect corresponds to the first term in (2.3.12). The second is the information effect: a higher price may also signal a higher value of v_J, and a trader might wish to buy more for a higher price. The information effect therefore has the opposite sign and corresponds to the second term in (2.3.12). Intuitively, this effect is stronger the more informative the price is. This is why, as can be seen from (2.3.12), the price elasticity γ_J is decreasing in τ_π. This observation provides the last takeaway: greater

[18] Intuitively, $\mathcal{I}$ measures the reduction in variance due to learning. As the I-traders know their value perfectly well, they do not contribute to $\mathcal{I}$.

price informativeness (higher τ_π) induces J-traders to provide less liquidity (decrease γ_J).

2.4 Strategic complementarities and multiplicity of equilibria

The four takeaways from the above section are the basis for the strategic complementarities in the model and the driver of the multiplicity of equilibria. The complementarities are represented in Figure 2.4.1, which depicts two feedback loops. The smaller one corresponds to complementarities *within* I-traders. If market is less liquid (λ is higher), I-traders provide less liquidity (γ is lower, cf. (2.3.7)). This, in turn, confirms a higher price impact (cf. (2.3.5)).

The larger loop corresponds to the complementarities *between* I- and J-traders. A higher price impact implies that I-traders are less aggressive (β is lower, cf. (2.3.7)). This implies that the price is more informative for J-traders (τ is higher, cf. (2.3.10)) and the market is more information efficient (as only J-traders learn from prices). Because the price is more informative, J-traders provide less liquidity (γ_J is lower): they are less willing to decrease their demand if the price increases because an increase in price may signal stronger fundamentals. This is confirmed by equation (2.3.11). The last step in the loop indicates that because J-traders provide less liquidity, the price impact is indeed higher (2.3.5).

Complementarities may generate multiple equilibria driven by self-fulfilling beliefs regarding liquidity. Indeed, suppose that there is an equilibrium. Suppose that traders believe that the liquidity is actually lower. The I-traders will then trade less aggressively. This will make the price more informative for J-traders, who will provide less liquidity. The latter confirms lower liquidity and potentially allows traders to coordinate on another equilibrium. One can also interpret the multiplicity as being driven by self-fulfilling beliefs concerning information efficiency. The latter interpretation works as follows. Suppose that there is an equilibrium. Suppose that the J-traders believe that the price informativeness is actually higher than that in equilibrium. They will then provide less liquidity. This would lead to a higher price impact of I-traders, who will trade less aggressively, confirming the higher price informativeness and potentially justifying another equilibrium.

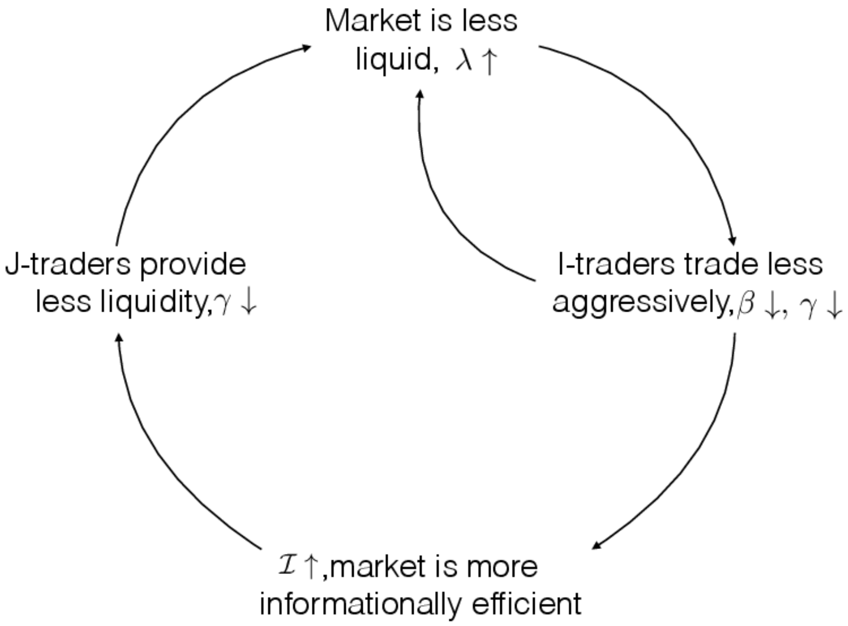

Figure 2.4.1: Equilibrium mechanism: two feedback loops. The smaller one corresponds to within-complementarities. The larger one corresponds to between-complementarities.

Theorem 3 provides sufficient conditions for uniqueness and multiplicity, which I discuss below. The complementarities between I- and J-traders are facilitated by the price inference of J-traders. The more informative the price is relative to the signal, the more J-traders rely on prices and the more the two groups of traders interact. Sufficient condition (2.3.2) ensures that the price is not too informative: if τ_I is low enough, there is enough noise in the price. This condition ensures that the between-complementarities (i.e., the larger loop in Figure 2.4.1) are not too strong to generate multiple equilibria. As is well known from the literature, the within-complementarities alone do not generate multiplicity are needed.[19], and hence no additional conditions to weaken the feedback in the small loop in Figure 2.4.1

The sufficient condition for multiplicity $\tau_I > \underline{\tau_2}$ ensures that price informativeness is high enough, such that the price inference channel, through which I- and J-traders interact, is important. The condition $\tau_I < \overline{\tau_2}$ ensures that price informativeness is not too high, and hence more/less aggressive trading by the I-traders can change the informativeness significantly. Thus, condition (2.3.4) ensures that the between-complementarities are strong enough. The condition that $w_I < \overline{w}$ ensures that the

[19]Indeed, within-complementarities are present even in the pure common values setting of, e.g., Kyle (1989), Vayanos (1999) and Rostek and Weretka (2015). However, the equilibrium in those models is unique.

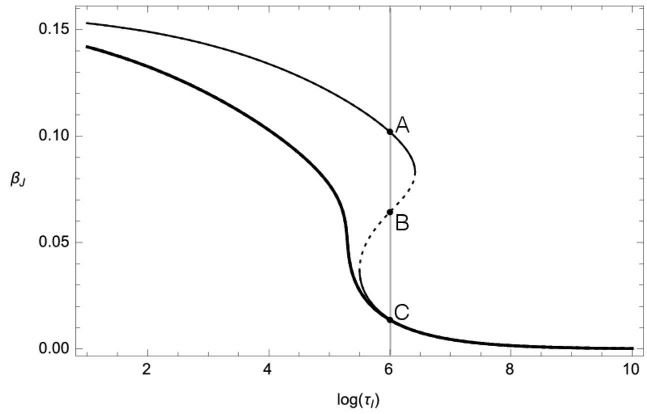

Figure 2.4.2: Multiple equilibria. The figure demonstrates dependence of β_J on $\log(\tau_I)$ for $w_I = 3$ (thick line) and $w_I = 1$ (thin line). Unstable equilibria are represented by dotted parts of the lines. The values of other parameters are $\rho = 0.9$, $\tau_J = 1$, $\tau_s = 1$, $N = 10$, $w_J = 1$.

price elasticity γ is not too small (cf. (2.3.7)). Together with the condition that N is large enough, the former condition ensures that $(N-1)\gamma$ is not too small relative to γ_J, and hence the between-complementarities are an important determinant of the price impact (cf. (2.3.5)). Thus, condition (2.3.3) ensures that the within-complementarities are strong enough.

Figure 2.4.2 illustrates the multiplicity of equilibria in the model. It plots the equilibrium sensitivities β_J against τ_I. It should be read as follows: draw a vertical line corresponding to a particular value of the parameter τ_I. Each intersection of the vertical line with the plot in Figure 2.4.2 corresponds to an equilibrium. If the line intersects with a dashed part of the plot, the equilibrium is unstable.[20] For example the equilibria A and C in Figure 2.4.2 are stable, whereas equilibrium B is unstable. Observe, consistent with Theorem 1, that there is a unique equilibrium if τ_I is small enough and that when w_I is small enough, there are three equilibria for the intermediate values of τ_I.

2.4.1 Liquidity and information efficiency

In this section, I consider the case of equilibrium multiplicity and compare the equilibria in terms of liquidity and information efficiency. Recall the definitions of

[20] The stability analysis is performed in Section 2.9.

liquidity and information efficiency

$$\mathcal{L} = \frac{1}{\lambda} \text{ and } \mathcal{I} = \frac{\mathrm{Var}(v_J)}{\mathrm{Var}(v_J|s_j, p)}.$$

Recall that the multiplicity is driven by the complementarity between illiquidity and information efficiency: lower liquidity induces higher information efficiency (through I-traders being less aggressive); higher information efficiency confirms lower liquidity (through J-traders providing less liquidity). Therefore, given a particular equilibrium, traders can coordinate on another one with lower liquidity and higher information efficiency.

The above suggests that the equilibria can be ranked in terms of $\mathcal{L}$ and $\mathcal{I}$, with the equilibria that are more liquid being less information efficient and vice versa. This is confirmed in Proposition 9. If the traders were to pick an equilibrium, they would have to choose between two evils: the equilibrium with the highest liquidity is the one with the lowest information efficiency and vice versa. To resolve this tension, I compute the welfare $\mathcal{W}$ (defined as the sum of expected utilities of all traders) and provide a sufficient condition that allows me to rank equilibria in terms of welfare. See Proposition 9 below.

Proposition 9. *Suppose that there are multiple equilibria. For any two equilibria A and B: $\mathcal{L}_A > \mathcal{L}_B$ if and only if $\mathcal{I}_A < \mathcal{I}_B$. Moreover, there exists $\underline{\tau}_J$ such that if*

$$\tau_J > \underline{\tau}_J \text{ and } \tau_I < 1 - \rho^2 \tag{2.4.1}$$

holds, then $\mathcal{W}_A > \mathcal{W}_B$ if and only if $\mathcal{L}_A > \mathcal{L}_B$.

Condition (2.4.1) should be understood as follows: prices do not provide much incremental information. Indeed, τ_J being large enough ensures that J-traders face little uncertainty regarding their value. The condition that τ_I is small enough implies that the price is not too informative from the perspective of $J-$traders. If condition (2.4.1) holds, liquidity is more important and the equilibria with higher liquidity are those with higher welfare.

2.4.2 Crashes

In this section, I explore the implications of the mechanism presented above for price crashes and the associated changes in information efficiency, liquidity, volume, volatility and welfare.

In what follows, I refer to the expected price $E[p]$ simply as price. I refer to the standard deviation of the price simply as volatility and denote it as σ_p

$$\sigma_p \equiv \sqrt{\mathrm{Var}(p)}.$$

The expected trading volume (volume hereafter) is defined as

$$\mathcal{V} \equiv \frac{1}{2} \cdot E\left[\int_0^1 |x_j(p^*)|\, dj + N \cdot |x_I(p^*)|\right].$$

I define a *crash (jump)* in an endogenous variable such as price, volatility or volume as follows.

Definition 2. Suppose that there is an endogenous variable X and a parameter of the model $\zeta \in \{\tau_I, \tau_J, \tau_s, \rho, w_I, w_J, N\}$. A crash (jump) of X is either of the two situations. (1) There are multiple equilibria. A crash (jump) is a sunspot switch from the equilibrium in which X is high (low) to the equilibrium in which X is low (high). (2) There is unique equilibrium, in which $\frac{dX}{d\zeta} = -\infty$ ($\frac{dX}{d\zeta} = +\infty$).

For example, the thick line in Figure 2.4.2 exhibits a crash of β_J when $\log(\tau_I)$ is close to 5.2, and a thin line exhibits a crash when $\log(\tau_I)$ is between 5.5 and 6.5. The proposition below characterizes the behavior of endogenous objects in the event of a price crash. I focus on the case of a price crash because in most markets, prices rarely jump up.[21] The corresponding statements for the case of jumps can be easily obtained in a way analogous to the proposition below.

Proposition 10. *Two scenarios are consistent with a price crash. (1) The price crash is associated with a liquidity crash, a jump in volatility and a jump in information efficiency, and if (2.4.1) holds, there is also a crash in the trading volume and welfare. This is the case when the I-traders are net buyers, i.e.,* $\bar{v}_I > \bar{v}_J$*. (2) The price crash is associated with a jump in liquidity, a crash in volatility and a crash in information efficiency, and if (2.4.1) holds, there is also a jump in the trading volume. This is the case when the I-traders are net sellers, i.e.,* $\bar{v}_I < \bar{v}_J$*.*

The above proposition identifies two scenarios consistent with a price crash. In the first scenario, the I-traders are net buyers. This scenario is represented

[21] The notable exception is currency markets: exchange rates do jump up.

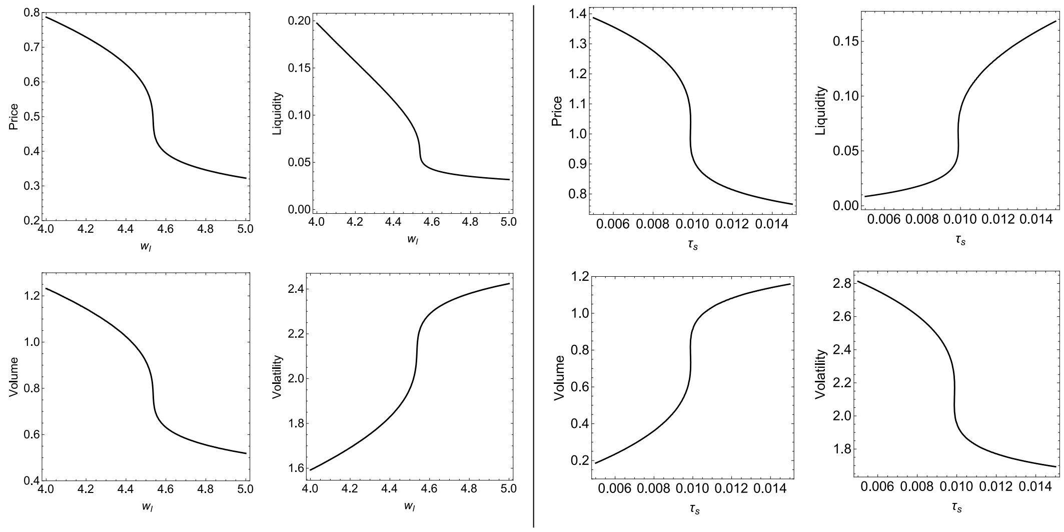

Figure 2.4.3: Two scenarios of a price crash. In the left panel, the I-traders are net buyers. The parameter values are $\bar{v}_I = 1.5$, $\bar{v}_J = 0$, and $\tau_s = 0.01$. In the right panel, the I-traders are net sellers. The parameter values are $\bar{v}_I = 0$, $\bar{v}_J = 1.5$, and $w_I = 4.5$. The values of the remaining parameters are the same for the two panels: $N = 10$, $w_J = 1$, $\rho = 0.9$, $\tau_J = 0.1$, and $\tau_I = 6.13$.

in the left panel of Figure 2.4.3. Let us interpret it in the context of securities markets (Example 1). A small change in the risk-bearing capacity of I-traders (an increase in w_I) reduces liquidity. This initial liquidity shock is amplified due to two feedback loops. Due to the liquidity shock, I-traders provide less liquidity, which feeds back into a higher price impact. As I-traders also trade less aggressively, the price becomes more informative and the J-traders provide less liquidity. This also feeds back into a higher price impact. A small liquidity shock is amplified and results is a large overall drop in liquidity. Due to the increased price impact, I-traders buy less and the prices drop. Because liquidity is low, relatively small orders can cause large price changes, and hence volatility increases. The volume drops for two reasons. First, due to the higher price impact, the I-traders trade less. Second, after the crash, information efficiency increases (because I-traders trade less aggressively); therefore, the ex post values of J-traders $E[v_J|s_j, p]$ are closer to the true value v_J and are therefore more aligned. This implies less volume generated by J-traders.[22] The first scenario is associated with a drop in liquidity but

[22]There is, however, an effect that works in the opposite direction. The variation in the ex post value $\text{Var}\,(E[v_J|s_j, p])$ may increase as a result of more information. To understand why, consider an extreme case in which the precision τ_s is zero. Without any information from price, the ex post value is $E[v_J|s_j] = \bar{v}_J$ and there is no variation in it. The more information the price provides, the closer the ex post value is to the true value v_J. Because the latter is stochastic, there will be more variation in the ex post value. The expected trading volume is an increasing function of the variance of the demand, which, in turn, depends on the ex post value $E[v_J|s_j, p]$. Therefore the above mechanism can lead to an increase in trading volume.

Condition (2.4.1) ensures that even without information from price, there is sufficient variation

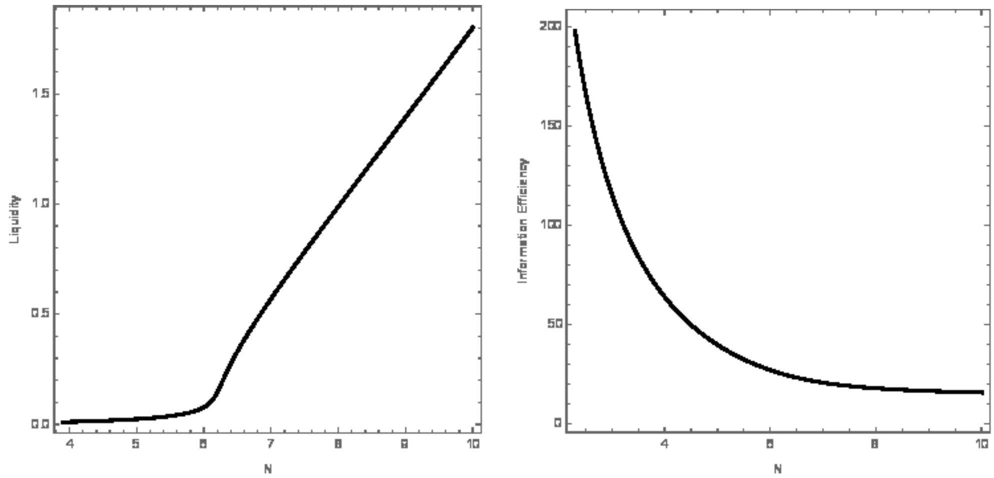

Figure 2.5.1: Tension between liquidity and information efficiency. Increasing N reduces the market power of I-traders and therefore improves liquidity, but because it induces I-traders to trade more aggressively, it reduces information efficiency.

an increase in information efficiency. If condition (2.4.1) holds (the price provides little incremental information), then such a crash is welfare-reducing and suggests a policy intervention.

In the second scenario, the I-traders are net sellers. This scenario is represented in the left panel of Figure 2.4.3. Let us interpret it in the context of commodities markets (Example 2). A small increase in the precision of information regarding the strength of the economy (an increase in τ_s) decreases the market power of producers (λ) . Due to the the mechanism discussed above, this reduction in market power is amplified and results in a substantial overall decrease in λ. Liquidity improves. Because the commodity producers have less market power, prices drop. The increase in liquidity means that volatility decreases. Volume increases because for two reasons. First, the commodity producers trade more, due to the lower price impact. Second, because there is less information, the ex post values of the firms are less aligned and there is an increase in the volume generated by them. The second scenario is associated with a drop in information efficiency but an increase in liquidity. If condition (2.4.1) holds (the informational role of price is not too important), then such a crash is welfare-improving, and no policy intervention is needed.

2.5 Comparative statics

In this section, I consider how information efficiency $\mathcal{I}$ and liquidity $\mathcal{L}$ are affected by changes in the model parameters. I focus on the following parameters: τ_s, which is related to informational frictions, N, which is related to the degree of competition, and w_I and w_J, which are related to liquidity. I consider two ways of obtaining comparative statics with respect to N.

1. No other parameters of the model change with N.
2. The convexity w_I is proportional to N, i.e., $w_I = w_1 N$, where w_1 is some constant. Other parameters are not affected by N.

The idea behind the second approach to obtain the comparative statics is as follows. Consider Example 2, in which the I-traders are producers. Decreasing (increasing) N in the second way corresponds to a merger (split) of existing producers.[23] Indeed, suppose that there are $N = n \cdot M$ producers with costs $C(x; N) = c \cdot x + \frac{w_I(N)}{2} x^2$. Suppose that every n producers have merged into 1. After the merger, there are M producers, each having n production units. To minimize the cost, producers will divide the production evenly across production units. Thus to obtain the output x, they will produce x/n units at each of the production units. Therefore, the cost function becomes $C(x; M) = nC(x/n; N) = c \cdot x + \frac{w_I(N)}{2n} x^2$. Therefore, $w_I(N/n) = w_I(N)/n$, and the coefficient w_I is indeed proportional to to the number of producers. In financial markets, the second approach to obtaining the comparative statics can regarded as a reduced-form approach to modeling the wealth effect (see Makarov and Schornick (2010)).

In the proposition below, I examine the comparative statics with respect to N.[24]

Proposition 11. *In the unique equilibrium, irrespective of whether w_I does not depend on N, or $w_I = w_1 N$,*

$$\frac{d\mathcal{I}}{dN} < 0 \text{ and } \frac{d\mathcal{L}}{dN} > 0.$$

in the ex post value $E[v_J|s_j] = \beta_J w_J s_j$ that the above effect is not too strong.

[23] The first way of obtaining the comparative statics corresponds to entry/exit.

[24] Although, by definition, N takes discrete values, the quantities $\mathcal{I}$ and $\mathcal{L}$ are continuous functions of N, and hence I provide the results for the derivatives of those functions, rather than finite differences, to simplify exposition.

The proposition above implies that there is tension between liquidity and information efficiency. Increasing the number of I-traders improves liquidity: with more I-traders, each of them has less market power, and thus the price impact is lower. However, because more liquidity induces I-traders to trade more aggressively, it reduces information efficiency. This is illustrated in Figure 2.5.1.

Next, I examine the comparative statics with respect to τ_s.

Proposition 12. *In the unique equilibrium*

$$\frac{d\mathcal{L}}{d\tau_s} > 0,$$

for $\tau_s > \frac{1-2\rho^2}{1-\rho^2}\tau_J$. *In particular, if* $\rho > \frac{1}{\sqrt{2}}$, *then* $\frac{d\mathcal{L}}{d\tau_s} > 0$ for all τ_s.

The intuition is as follows. With more precise signals, the J-traders learn more from their signals and less from prices. Their price elasticities increase, and liquidity improves.

The comparative statics for information efficiency are driven by two forces. On the one hand, increasing τ_s has a positive, direct effect on $\mathcal{I} \equiv \frac{\mathrm{Var}(v_J)}{\mathrm{Var}(v_J|s_j,p)} = \frac{\tau_J+\tau_s+\tau_\pi}{\tau_J}$. On the other hand, as the proposition above indicates, increasing τ_s improves liquidity and makes the I-traders trade more aggressively. This may have a negative effect on the precision τ_π of the price signal. If the second effect prevails, the *information aggregation paradox* obtains: aggregating an ex ante more precise information (higher τ_s) market conveys less information ex post (lower $\mathcal{I}$). Intuitively, the second force is stronger when the J-traders learn more from prices, which is the case when τ_s is low: this is illustrated in Figure 2.5.2.

Figure 2.5.2 illustrates that there is tension between liquidity and information efficiency when τ_s is small, such that there is an information aggregation paradox. When τ_s is large, there is no tension: improving the precision of information (i.e., by reducing the information acquisition costs) improves both liquidity and information efficiency.

I next examine the comparative statics with respect to w_I and w_J. An increase in w_I or w_J is interpreted as a decrease in risk-bearing capacity, which can be due to tightened of regulations or an external liquidity shock. Consider first the effect of a change in w_I and w_J on information efficiency. Intuitively, if w_J decreases, J-traders trade more aggressively on their signals and the price becomes more informative.

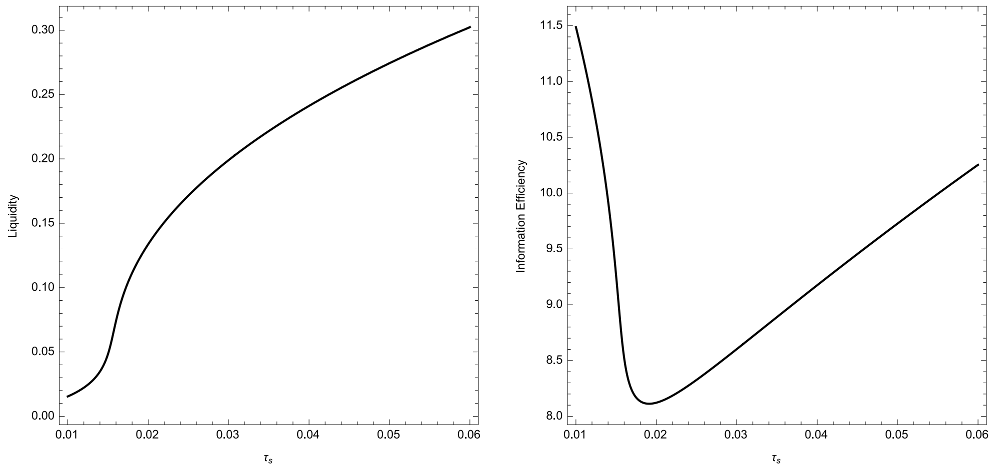

Figure 2.5.2: Left panel: liquidity is increasing in the precision of the signal τ_s. The higher the precision is, the less the J-traders learn from prices, the higher the price elasticity of their demand is and the greater the liquidity. Right panel: for small values of τ_s, there is an *information aggregation paradox*; the aggregation of more information yields less information ex post. The parameter values are $N = 10$, $w_I = 4.5$, $w_J = 1$, $\rho = 0.9$, $\tau_J = 0.1$, and $\tau_I = 7$.

An increase in w_I induces I-traders to trade less aggressively and therefore has a similar effect. This is intuition is confirmed in the proposition below.

Proposition 13. *In the unique equilibrium,* $\frac{d\mathcal{I}}{dw_J} < 0$ *and* $\frac{d\mathcal{I}}{dw_I} > 0$.

A decrease in the risk-bearing capacity of I-traders (an increase in w_I) has a direct negative effect on liquidity. It also has an indirect effect: an increase in w_I increases information efficiency, which has a negative effect on $\mathcal{L}$ because J-traders provide less liquidity. Therefore, the overall effect of the liquidity shock to I-traders on liquidity should be negative. This is confirmed in the proposition below.[25]

Proposition 14. *In the unique equilibrium,* $\frac{d\Gamma}{dw_I} < 0$*, where* $\Gamma = N\gamma + \gamma_J$ *is a slope of the aggregate demand.*

Combining the results of Propositions 13 and 14, it is clear that if the risk-bearing capacity of I-traders increases, liquidity improves but the information efficiency deteriorates. Thus, there is tension between liquidity and information efficiency.

A shock to the risk-bearing capacity of J-traders has conflicting effects on liquidity. The direct effect is negative: if w_J increases, J-traders provide less liquidity. However, an increase in w_J has a negative effect on information efficiency. This can induce J-traders to provide more liquidity. This effect can be amplified through

[25] The proposition examines the effect of w_I on Γ. The numerical result is that $\mathcal{L}$ is also decreasing in w_I.

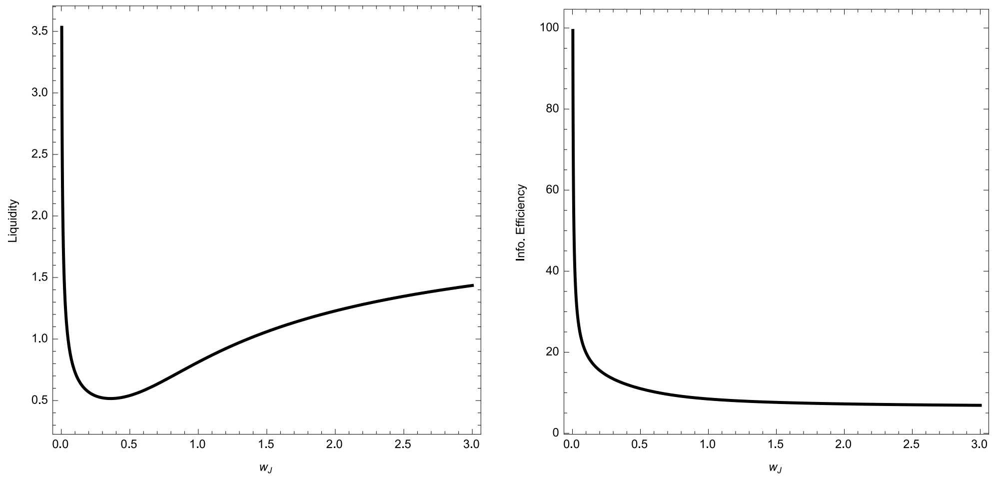

Figure 2.5.3: Left panel: liquidity paradox. An adverse shock to the risk-bearing capacity of J-traders may lead to an increase in liquidity. Right panel: information efficiency is decreasing in w_J. The parameter values are $N = 4$, $w_I = 1$, $\rho = 0.9$, $\tau_s = 0.1$, $\tau_J = 0.1$, and $\tau_I = 1$.

I-traders being more aggressive. If the second effect dominates, a *liquidity paradox* obtains: an adverse liquidity shock leads to an improvement in liquidity. This is represented in Figure 2.5.3.

I finally consider the question of how combining two markets into one (or, alternatively, breaking up an existing exchange into two) affects the market quality of the combined market.

The exercise is as follows. Consider a market with N I-traders and a unit mass of J-traders. Divide this market into two: let $M < N$ I-traders trade with a proportional measure M/N of J-traders in "market 1"; let the remaining traders trade in "market 2". Markets 1 and 2 are completely segmented. I am interested in how the information efficiency and liquidity of markets 1 and 2 are related to those of a combined market ("1+2"). Intuitively, markets 1 and 2 are less competitive than a combined market $1 + 2$, and hence the liquidity should be lower. However, because I-traders are less aggressive in smaller markets, information efficiency can increase. The proposition below confirms this intuition: breaking up a market into two is bad for liquidity but is good for information efficiency.

Proposition 15. *Liquidity and information efficiency in markets* 1, 2 *and* $1 + 2$ *described above are related as follows:* $\mathcal{L}_{1+2} > \mathcal{L}_1$ *and* $\mathcal{L}_{1+2} > \mathcal{L}_2$*; however,* $\mathcal{I}_{1+2} < \mathcal{I}_1$ *and* $\mathcal{I}_{1+2} < \mathcal{I}_2$.

Note that in the competitive economy, proportionally scaling the measures of agents in the market has no effect on equilibrium, as it cancels out through market

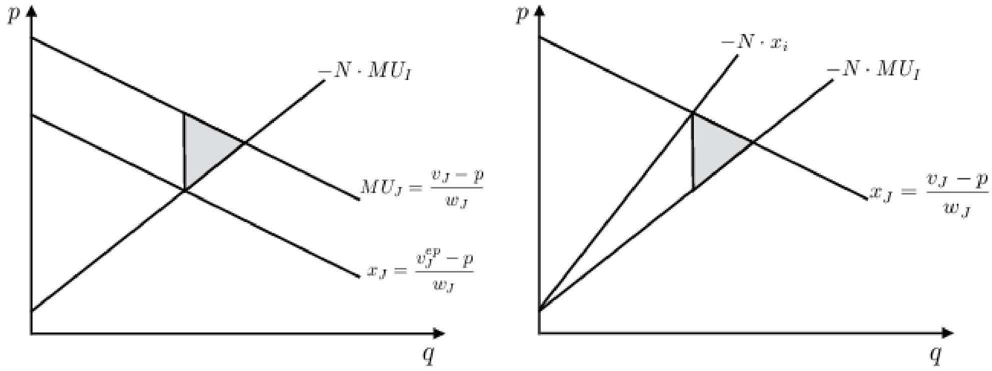

Figure 2.6.1: The role of liquidity and information efficiency in determining welfare: a gray triangle represents a welfare loss. Left panel: the role of information. In the economy with I-traders being price takers (such that liquidity plays no role), with more information the ex post value v_J^{ep} is less biased relative to v_J (state by state) which helps to reduce welfare loss. Right panel: the role of liquidity in determining welfare. In the economy in which J-traders know their value perfectly well (such that information plays no role), I-traders reduce their demands, which results in a welfare loss.

clearing. Comparing economy $1+2$ with either economy 1 or economy 2 provides another way of isolating the effect of competition on liquidity and information efficiency. Competition is good for liquidity but is bad for information efficiency.

2.6 Welfare

It is often argued that greater competition is associated with greater welfare. I examine the validity of this claim in the context of my model.

Define

$$\mathcal{U}_I \equiv E\left[(v_I - p)\,x_i(p) - \frac{w_I x_i(p)^2}{2}\right],$$

the expected utility of an I-trader, and

$$\mathcal{U}_J \equiv E\left[(v_J - p)\,x_j(p) - \frac{w_J x_j(p)^2}{2}\right],$$

the expected utility of a J-trader. The total welfare is then

$$\mathcal{W} \equiv N \cdot \mathcal{U}_I + \mathcal{U}_J.$$

With no informational frictions, the effect of competition on welfare is unambiguous (e.g., Tirole (1988)): with more competition, welfare improves. The intuition is as follows: when N increases, large traders have less market power and there is less

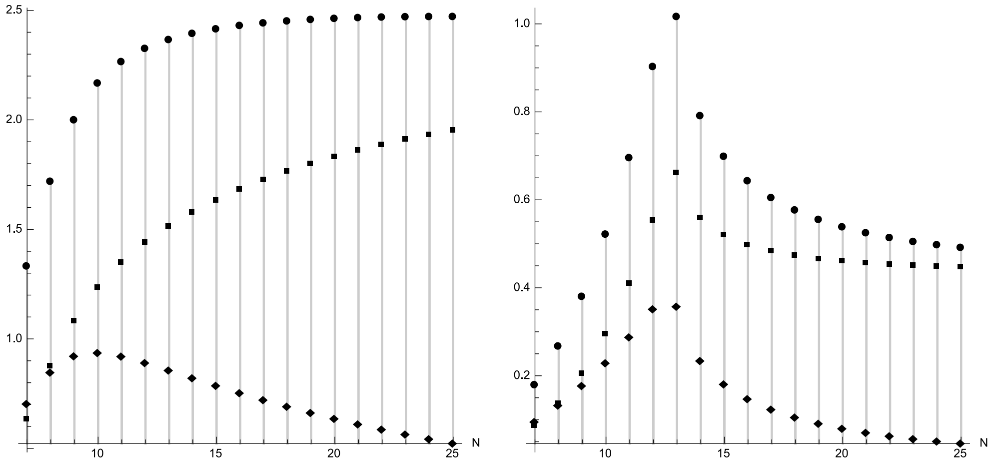

Figure 2.6.2: Welfare $\mathcal{W}$ (circles), expected utility of a $J-$trader $\mathcal{U}_J$ (squares) and the expected utility of I-traders $N \cdot \mathcal{U}_I$ (diamonds) . Left panel: $\tau_s = 0.1$. When the informational frictions are not too strong, the standard intuition applies: competition is good for welfare and for small traders. Right panel: $\tau_s = 0.01$. When informational frictions are strong, welfare becomes non-monotonic in N. Moreover, it is possible that everyone, including small traders, is worse off as a result of more competition. The remaining parameter values are $w_I = 8$, $w_J = 1$, $\rho = 0.3$, $\tau_J = 0.1$, and $\tau_I = 25$.

reduction in their demands. This helps to reduce deadweight loss, and welfare increases. Because the large traders have less market power, prices are more favorable to small traders, and hence they also become better off.

The proposition below provides a sufficient condition for the "usual" comparative statics for welfare: if prices provide little incremental information (condition (2.4.1) holds), the standard intuition applies.

Proposition 16. *Suppose that there is a unique equilibrium and (2.4.1) holds. Then,* $\frac{d\mathcal{W}}{dN} > 0$ *and* $\frac{d\mathcal{U}_I}{dN} > 0$*, irrespective of whether* w_I *does not depend on* N*, or* $w_I = w_1 N$ *.*

I show below that if condition (2.4.1) does not hold, increasing competition might actually reduce welfare. Moreover, this can be bad for everyone, including J-traders. I attribute this result to the tension between liquidity and information efficiency outlined in the section above. To proceed, it is necessary to understand the effects of liquidity and information efficiency on welfare. In the model, the two are tightly linked through the mechanism represented in Figure 2.4.1. Therefore, it is difficult to disentangle the roles of the two. To overcome this difficulty, I consider the following two thought experiments.

First, consider an economy in which the role of liquidity is "switched off". Suppose that traders behave as if the market were perfectly liquid, i.e., even I-traders take prices as given.[26] The informational frictions in this economy are the same as in the original economy. In this economy, welfare increases in information. First, with more information, the equilibrium quantities allocated to J-traders are less dispersed, which is good for welfare.[27] Second, with more information, the ex post values of J-traders ($E[v_J|s_j,p]$) are less biased,[28] which also increases welfare. Indeed, the maximum welfare in this economy is achieved when traders bid according to their marginal utilities

$$x_i = MU_I \equiv \frac{v_I - p}{w_I}, \text{ and } x_j = MU_J \equiv \frac{v_J - p}{w_J}.$$

However, the aggregate trade of J-traders is actually

$$\begin{aligned} x_J &\equiv \int x_j dj \\ &= \frac{v_J^{ep} - p}{w_J}, \text{ where } v_J^{ep} = \int E[v_J|s_j,p]dj. \end{aligned}$$

A bias between v_J^{ep} and the true value v_J results in a welfare loss. See the left panel of Figure 2.6.1.

Second, consider an economy in which the role of information is "switched off". Suppose that all traders know their values but that I-traders exercise their market power. This economy is, essentially, the textbook oligopoly model discussed above. The I-traders will reduce their demands, which will result in welfare loss. See the right panel of Figure 2.6.1.

Summarizing the above discussion, I conclude that taken in isolation, both liquidity and information efficiency are beneficial for welfare. This suggests that increasing competition can have an adverse effect on welfare through its adverse effects on information efficiency, as established in Section 2.5. Because the negative effect operates through the information channel, it should be more pronounced when informational frictions are high. This intuition is confirmed in Figure 2.6.2: when τ_s is small, welfare may be non-monotonic in N. Moreover, everyone, even small traders,

[26]This setting corresponds to REE in the model.

[27]It is easy to show that allocating the average quantity $x_J = \int x_j(p)dj$ to all traders (instead of allocating $x_j(p)$ to each of them) increases the ex post aggregate utility $\int_0^1 \left((v_J - p)\, x_j(p) - \frac{w_J x_j(p)^2}{2} \right) dj$ of J-traders. This is due to the concavity of their objective. The more dispersed $x_j(p)$ are relative to x_J, the smaller the ex post aggregate utility.

[28]Indeed, define the bias as $E\left[E[v_J|s_j,p] - v_J|v_J\right] = \frac{\tau_J}{\tau}(\bar{v}_J - v_J)$. It goes to zero as $\tau \to \infty$.

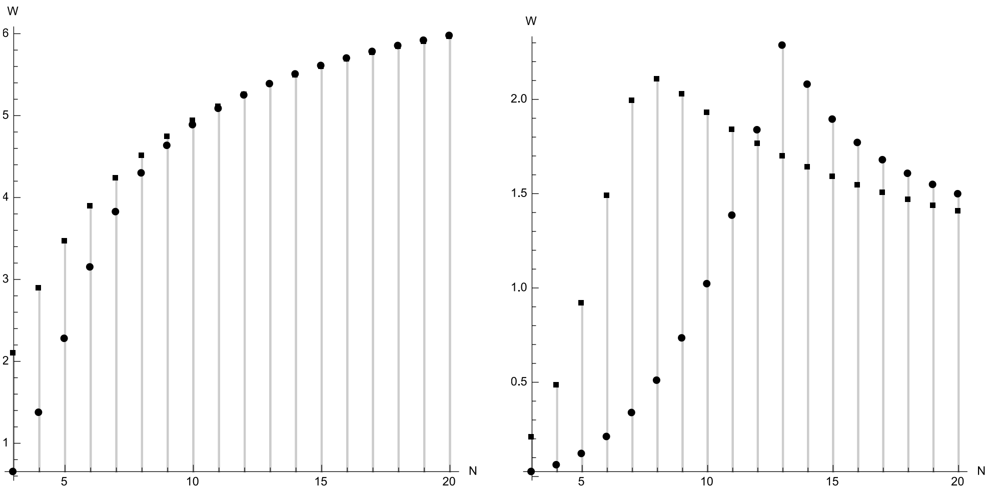

Figure 2.6.3: Welfare in the economy with 2 segmented markets (circles) and a combined market (squares). Left panel: $\tau_s = 0.3$. When the informational frictions are not too strong, the segmentation is bad for welfare. Right panel: $\tau_s = 0.01$. When informational frictions are strong, segmentation might be beneficial for welfare. The remaining parameter values are $w_I = 8$, $w_J = 1$, $\rho = 0.3$, $\tau_J = 0.1$, and $\tau_I = 25$.

can be worse off as a result of more competition.

I finally consider the question of the effects of breaking up an exchange on welfare. Common wisdom suggests that welfare should be higher in a centralized market. A centralized market should better aggregate information, and the traders should better share their risks in such a context. Figure 2.6.3 shows that this common wisdom may not be correct: it plots the welfare in an economy in which there are two identical segmented markets and the welfare when those two markets are combined into one relative to the number N of I-traders in either of the segmented markets. The right panel of the figure indicates that segmentation might be beneficial. The intuition is as follows: in two segmented markets, I-traders are less competitive, which results in lower liquidity; however, this is beneficial for information efficiency,[29] as I-traders trade less aggressively (Proposition 15). When N is large, such that the liquidity loss resulting from breaking up the market is less considerable, segmentation might be beneficial. This result is present when informational frictions are important (τ_s is low). A similar graph can be obtained for the surplus of J-traders instead of aggregate welfare.

[29]Malamud and Rostek (2015) show that breaking up an exchange can be beneficial through its effects on liquidity. Their setting does not feature asymmetric information and does not capture information efficiency. The result in this paper therefore complements that of Malamud and Rostek (2015). It would be desirable to incorporate the mechanism highlighted in this paper into the much more general market structure environment considered in Malamud and Rostek (2015).

2.7 Implications

In this section, I consider the implications of the model.

2.7.1 Commodities markets

I consider two episodes, the 2008 boom/bust in oil prices and the recent crash of oil prices, through the lens of the model.

2008 boom/bust in oil prices Oil prices reached an all-time high of \$145 per barrel in July 2008, a 40% increase from the level in January of 2008, when the US and most other developed economies were entering a recession. It is difficult to explain such a sharp increase by a shift in either demand or supply: no major disruptions in supply occurred at that time; demand from developed economies was weaker, and it is unlikely that the demand from the developing economies (which were performing well at the time) could have offset this weakness.

My explanation for this episode emphasizes the role of two forces: (1) informational frictions and (2) the market power of oil producers being endogenously amplified because of (1). In the model, market power corresponds to λ, which measures the extent to which producers can drive up the price by reducing their supply. Proposition 12 implies that the producers' market power is higher when information concerning economic fundamentals is less precise ($1/\tau_s$ is higher). This is intuitive: when informational frictions are high, firms on the demand side rely on commodity prices as a signal of the strength of the economy. When commodity producers reduce their supply, they are driving up the price. Firms partly attribute the increase in price to stronger fundamentals and demand more, which amplifies the price impact of producers. Therefore, the boom in prices can be attributed to an increase in the market power of commodity producers caused by the uncertainty regarding the strength of the economy at the time.

Singleton (2014) presents empirical evidence that supports this explanation. He finds a strong, positive correlation between the dispersion of oil price forecasts (related to $1/\tau_s$ in the model) and the oil price level. Such a relationship can be explained by an increase in the market power of producers caused by higher uncertainty. This evidence therefore supports the mechanism discussed above.

In emphasizing the role of informational frictions, my explanation is closely related to that of Sockin and Xiong (2015).[30] My explanation complements theirs by highlighting the role of commodity producers' market power and the role of informational frictions in amplifying the latter.

The bust can be explained by the price effect of the demand shock coming from CITs and hedge funds that unwounded their positions in the commodities markets during the financial crisis, as documented by Cheng et al. (2015). My model can help explain the magnitude of the price effect of this shock. Informational frictions amplified the illiquidity of the market, and a demand shock had a larger price effect.

2014 crash in oil prices Between January 2012 and October 2014, the oil price ranged from \$80 to \$110 per barrel. By the by the end of 2014, it had halved and remains in the range \$40 - \$60 to the present.

I attempt to explain this episode with assistance from my model. As illustrated in the right panel of Figure 2.4.3, a small change in the precision of information regarding the fundamentals of the economy (an increase in τ_s) can cause a sharp decrease in the market power of commodity producers (increase in $\mathcal{L}$) and a price crash in the model. This is because of the complementarity between the market power of producers and informational frictions discussed above. Therefore, the resolution of uncertainty regarding the strength of the world economy (e.g., news about "China's new normal")[31] could have sharply decreased the market power of commodity producers and caused a price crash.

The above explanation attributes the sharp decrease in price to a decrease in the market power of oil producers. Consistent with this explanation, OPEC did not cut its production following the crash.[32] If the above story is true, after the crash, information efficiency should have decreased (Proposition 10): oil prices should have become worse barometers for the global economy. This can be tested, for example, by conducting the exercise in Hu and Xiong (2013) for the periods before and after the 2014 crash.

[30] See Section V.C of Sockin and Xiong (2015) for more evidence and discussion on the importance of informational frictions.

[31] E.g., "Xi Says China Must Adapt to 'New Normal' of Slower Growth", Bloomberg, May 12, 2014

[32] See, e.g., "OPEC Pumps at Three-Year High Despite Oil Glut", WSJ Aug. 11, 2015

2.7.2 Asset markets

I showed in the Section 2.5 that changes in different aspects of the market environment (such as risk-bearing capacity, information precision, entry/exit or competition between large traders) can induce changes in liquidity and information efficiency in opposite directions. The latter implies, in particular, that when liquidity improves, the price better (worse) reflects the values of large (small) traders. In reality, one does not observe the values of the large traders. However, there are two settings in which the two can be proxied.

On-the-run treasury market and short-sellers In the treasury bonds market, the difference in the prices of the on-the-run bonds and off-the-run bonds is attributed to "specialness", the quality of the on-the-run bond of being better collateral. The difference between the price of the on-the-run and off-the-run bond can therefore be attributed to the values of the short sellers, who are buying the on-the-run bonds to use as collateral. Regard the short sellers as the I-traders; the model then implies that when the liquidity of the on-the-run market improves, the price of the on-the-run bond should increase. This is because in a more liquid market, the short sellers will trade more aggressively, driving up the price.[33] Because the off-the-run bonds are unaffected by the short sellers, the spread should increase. This is consistent with the evidence in Krishnamurthy (2002) and Banerjee and Graveline (2013)[34].

Equity markets and institutional investors Consider equity markets. Interpret I-traders as large institutional investors and J-traders as retail investors. Tension between liquidity and information efficiency implies that in a more liquid market, the price better reflects the value of institutional investors. While the latter is unobservable, arguably,[35] it should be correlated with a benchmark relative to which the institutions are evaluated. Consequently, a stock traded by institutions

[33]More formally, this is a consequence of Lemma 12 in the Appendix.

[34]Vayanos and Weill (2008) describe similar implications in a search-based model. They show that if there are more short sellers in the market, both the spread and liquidity increase. The intuition in their model is as follows: if there are more short sellers in the on-the-run bond market, the price is more reflective of their values and is thus higher. Moreover, entry of short sellers also increases liquidity by relaxing search frictions. The main difference is that in my paper, the higher price is not necessarily a consequence of the entry of short sellers but rather more aggressive trading by existing ones due to higher liquidity.

[35]See, e.g., Basak and Pavlova (2012), who derive that the marginal utility of the fund manager should be increasing in the level of the benchmark in a moral hazard framework.

should become more correlated with a benchmark when the liquidity of the stock improves. This prediction is testable.

Indirect evidence supporting the above prediction is provided by Chan et al. (2013), who show that stocks' co-movements with one another are positively related to market liquidity. The latter can be explained as follows. When there is more liquidity, the price of each stock better reflects the values of institutions and is thus more correlated with the benchmark against which the institutions are evaluated. The stocks' correlations with one another therefore also increase.

2.7.3 Policy

A recent policy debate centers on the effects of high-frequency traders (HFTs) on asset markets and commodity index traders (CITs) on commodities futures markets. It is often argued that the presence of those traders is beneficial because it improves market liquidity and, by incorporating the information those traders have into prices, information efficiency. The results in Section 2.5 confirm the beneficial effects of those groups of traders on liquidity (Proposition 11, interpret CITs and HFTs as I-traders)[36]. However, because those groups of traders may have different values from those of other traders the beneficial effect on liquidity may feed back into an adverse effect on information efficiency:[37] with more liquidity, the price will better reflect the values of HFTs and CITs as they trade more aggressively.

It is often argued that greater competition is beneficial for welfare. Moreover, this is a basis of antitrust policy around the world.[38] As shown in Propositions 11 and 15, promoting competition is beneficial for market liquidity but may be harmful for information efficiency. Consequently, the standard intuition holds when the price provides little incremental information to economic agents (Proposition 16). However, when the price is a valuable source of information, competition may

[36] Indeed, in asset markets, HFTs dominate trading at high frequencies. They also use quantitative strategies to account for their price impact. In this context, regard J-traders as retail investors.

In commodities futures markets, CITs and large hedge funds have large positions in commodities futures (Cheng et al. (2014)). They also have price impact, as shown in Cheng et al. (2014).

[37] HFTs may have a short-term horizon and care about the price in the near future. Retail investors may have a longer horizon and consequently care about the price in the more distant future.

CITs may have specific hedging needs arising from their positions in other markets. Evidence for this is provided in Cheng et al.

[38] See, e.g., US Federal Trade Commission introduction to antitrust laws https://www.ftc.gov/tips-advice/competition-guidance/guide-antitrust-laws.

harm price discovery (as the price will better reflect the values of large traders) and may have adverse effects on welfare. Moreover, all traders, large and small, could be worse off. Below, I discuss the economic environments in which the latter effects might be important.

In the context of product markets, consumers may have little information on the quality of new products. The price may be an important signal of quality. Consequently, consumers can benefit from patent protection, which restricts competition among producers (allowing them to recover the costs of designing a new product): with less competition, the price of a new product better reflects its quality.

In the context of financial markets, an important dimension of competition comes from the market structure: there are increasingly more trading venues that allow investors to execute their trades. I show that breaking up a centralized market into two separate exchanges can improve welfare because each of the two exchanges will be more informationally efficient. This result complements the results of Malamud and Rostek (2015), who show that breaking up the exchange can be beneficial because of the effects on liquidity.

2.8 Conclusion

This paper shows that traders' heterogeneity in their price impacts and values - a notable feature of many contemporaneous markets - might have unexpected consequences for liquidity, information efficiency and welfare.

I show that the heterogeneity results in a complementarity between illiquidity and information efficiency. A belief that the market is less liquid induces large traders to trade less aggressively. It makes the price relatively more sensitive to the values of small traders. From the latter's perspective, the price is more informative, and they provide less liquidity. Consequently, the market is less liquid. The complementarity has the following consequences.

First, tension might arise between liquidity and information: policy measures intended to promote liquidity may harm information efficiency and vice versa. Related to the latter, changes in the market environment (such as risk-bearing capacity, the number of large traders, information precision) can shift liquidity and information efficiency in opposite directions.

Second, I show that an increase in risk-bearing capacity may have a negative effect on liquidity (*liquidity paradox*). This is possible because despite its direct negative effect on liquidity, it has a positive effect on information efficiency, and there is a tension between the two. Similarly, an increase in the precision of information might have a negative effect on information efficiency (*information aggregation paradox*).

Third, competition is not necessarily beneficial for welfare. This is possible because competition has negative effects on information efficiency. For a similar reason, breaking up a centralized market into two separate exchanges might improve welfare.

The evidence obtained from commodities and asset markets is broadly consistent with the predictions of the model.

The model can be extended in multiple directions. Allowing for multiple assets would allow for a treatment of asset-class effects in the presence of institutional investors and interactions with market liquidity and information efficiency. Considering a dynamic extension is also interesting. Consider a large trader, who can trade in several periods. If he trades more aggressively in the first period, the price will be less informative for small traders, and they will provide more liquidity. Trading more aggressively will then be less costly for a large trader. Consequently, in the presence of the complementarity highlighted in this paper, large investors may trade faster. These extensions are left for future work.

2.9 Stability and Amplification

One of the notions of stability in game theory is associated with a stability of a fixed point of the best-response mapping determining the equilibrium.[39] This notion is commonly adopted in REE models.[40]

The idea behind this notion is the following. Suppose that agents make a small deviation from the profile of the equilibrium strategies S_0. Denote the perturbed strategy profile by S_1. Let agents play S_2, which is a best response to S_1. Let them play a best response (denote it S_3) to the strategy profile S_2, and so on. If, as a result of such tÃ¢tonnement process, the strategies will converge back to the initial equilibrium (i.e. $S_n \to S_0$ as $n \to \infty$) the equilibrium is called stable.

[39] See e.g. Fudenberg and Tirole (1994) ch. 1.2.5 in a context of Cournot duopoly.
[40] E.g. Cespa and Vives (2015).

Such analysis is tractable in symmetric models, because the best response typically maps a scalar (the equilibrium strategy of other agents) onto a scalar (the best reply of a particular agent).

In this model, however, there are two groups of traders, and the best response depends on a vector of strategies. For example the best response sensitivity β_J depends on β and γ chosen by I-traders and β_J and γ_J chosen by the other J-traders. In other words the best response maps $\mathbb{R}^4$ onto $\mathbb{R}^4$, which complicates the stability analysis of its' fixed point.

To overcome this difficulty, I represent the equilibrium as being a fixed point of a mapping characterizing the market as a whole. The characteristic I focus on is illiquidity, i.e. price impact.[41]

To understand how the price impact mapping onto itself is determined, consider the following logic. Suppose that traders believe that the price impact is l. Given this belief they optimally choose their strategies. Those strategies determine the "true" price impact, i.e. the slope of the inverse residual supply $\Lambda(l)$. In equilibrium, the assumed price impact should be equal to the "true" one. In other words, the fixed point condition $l = \Lambda(l)$ should hold.

I will call the equilibrium stable if the fixed point $l = \Lambda(l)$ is stable. The $\Lambda(l)$ maps $\mathbb{R}$ onto $\mathbb{R}$, so the stability of fixed point of this function is easier to analyze. The intuition behind such definition of stability is as follows. Suppose that the equilibrium price impact is l_0. Suppose that traders hold slightly incorrect belief l_1 about the price impact. With this belief they choose their strategies, which determines the slope of residual supply $l_2 = \Lambda(l_1)$. Now the traders realize that the price impact is l_2, not l_1. They choose their strategies, which determine the new price impact $l_3 = \Lambda(l_2)$. If the iteration of that process brings them back to the l_0, then the equilibrium is stable. The latter is equivalent to the stability of the fixed point $l = \Lambda(l)$.

I derive the mapping $\Lambda(l)$ and formally define stability below.

Suppose that the traders believe that the price impact is equal to l such that

[41] The idea of equilibrium in a model with market power being a fixed point of a mapping of price impact to itself is due to Weretka (2011). The representation of the supply function equilibrium as a fixed point of a mapping of price impact to itself is due to Rostek and Weretka (2015).

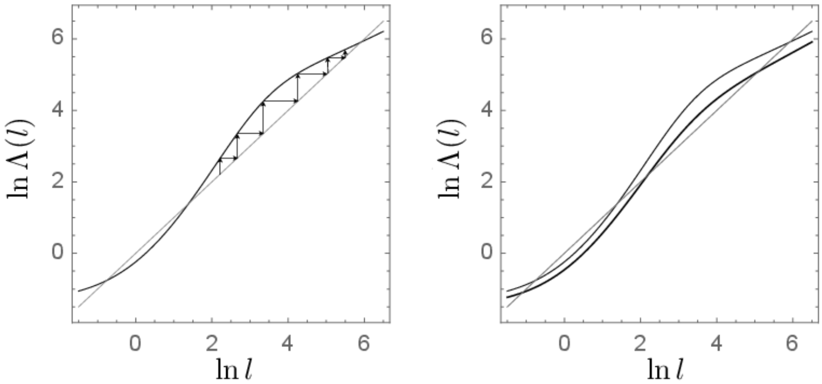

Figure 2.9.1: Stability analysis. The left panel illustrates the concept of stability. The middle equilibrium is unstable: a small deviation from it leads to further deviation, until the system converges to the top equilibrium, which is stable. The right panel shows that stable and unstable equilibria provide opposite comparative statics results: an increase in the variance of I-traders' value decreases the price impact in the stable equilibria and increases it in the unstable one. The parameter values are $\rho = 0.99$, $\tau_J = 1$, $\tau_s = 1$, $N = 10$, $w_J = 1$, $w_I = 1$, $1/\tau_I = 0.006\,(0.01)$ (thin (thick) line).

$2l + w_I > 0$.[42] The I-traders' strategy is then given by (2.3.6), implying

$$\beta = \gamma = g(l) \equiv \frac{1}{w_I + l}. \tag{2.9.1}$$

The strategy of each of J-traders given l is determined as best-response to the strategies (2.9.1) of the I-traders and to the strategies of other J-traders. Denote those by $\beta_J = b_J(l)$ and $\gamma_J = g_J(l)$.

Combining (2.3.10) and (2.3.11) one can find the implicit expression for $\beta_J = b_J(l)$

$$\frac{1}{\beta_J} = w_J \left(\frac{\tau_J + \tau_s}{\tau_s} + \frac{1}{\tau_s C^2} \left(B + \frac{\beta_J}{N \cdot g(l)} \right)^2 \right). \tag{2.9.2}$$

The left-hand side of the above is strictly decreasing while the right-hand side is strictly increasing in β_J for $\beta_J > 0$, therefore there is a unique solution $\beta_J = b_J(l)$ to (2.9.2). Moreover, differentiating the above implicitly, one can see that this function is strictly decreasing in l.

Given $\beta_J = b_J(l)$ and $\beta = g(l)$, the precision of the signal π is determined by (2.3.10), implying

$$\sqrt{\tau_\pi} = t(l) \equiv \frac{1}{C} \left(B + \frac{b_J(l)}{N \cdot g(l)} \right). \tag{2.9.3}$$

[42] The minimal price impact that can be sustained in equilibrium is $\lambda = -w_I/2$. This is because the second-order conditions would be violated otherwise. See Lemma 3 in the Appendix.

Given the above one can find the elasticity γ_J from (2.3.12)

$$\gamma_J = \frac{1}{w_J} - \frac{1}{\tau_s} t(l) \left(t(l) - \frac{B}{C} \right) (\gamma_J + Ng(l)). \tag{2.9.4}$$

In the above I used the fact that the aggregate elasticity $\Gamma = N\gamma + \gamma_J$ is equal to $\gamma_J + \frac{N}{w_I + l}$. Expressing γ_J from the equation (2.9.4) yields

$$\gamma_J = g_J(l) \equiv \frac{\tau_s - w_J N t(l) \left(t(l) - \frac{B}{C} \right) g(l)}{\tau_s w_J + w_J \cdot t(l) \left(t(l) - \frac{B}{C} \right)}. \tag{2.9.5}$$

We now know how the price elasticities $g(l)$ and $g_J(l)$ are determined. Those, in turn, determine the "true" price impact $\Lambda(l)$:

$$\Lambda(l) \equiv \frac{1}{(N-1)\, g(l) + g_J(l)}. \tag{2.9.6}$$

In equilibrium the price impact assumed by traders should be equal to the slope of the residual supply, i.e. $l = \Lambda(l)$.

The above argument is justified by the following Theorem.

Theorem 4. *The equilibrium price impact is equal to l if and only if it solves the fixed point problem $l = \Lambda(l)$ and $2l + w_I > 0$.*

We are now ready to formally define stability.

Definition 3. The equilibrium is called stable if and only if the price impact λ in that equilibrium is a stable fixed point of the function $\Lambda(\cdot)$, that is, iff the λ satisfies $|\Lambda'(\lambda)| < 1$.

The left panel of Figure 2.9.1 illustrates the above definition.[43] According to the definition the middle equilibrium on the Figure should be unstable: the function $\Lambda(l)$ crosses the 45 degree line from below, therefore its' slope is greater than one. The figure illustrates the tÃ¢tonnement process by a sequence of arrows: a small deviation from the middle equilibrium leads to further deviation, until the system converges to the top equilibrium, which is stable. The right panel illustrates that the stable and unstable equilibria yield the opposite comparative statics results. An increase in the variance of I-traders' value shifts the curve $\Lambda(l)$ down (with greater

[43] The Figure applies log-log scale, because otherwise the top equilibrium is too far apart from the two bottom ones and the plot does not read well. Since in equilibrium $l = \Lambda(l)$, one may write $\frac{d \ln \Lambda(\lambda)}{d \ln l} = \frac{l}{\Lambda(l)} \Lambda'(l) = \Lambda'(l)$. Therefore the stability can be seen graphically on the Figure 2.9.1: the equilibrium is stable iff $|\ln \Lambda'(l)| < 1$.

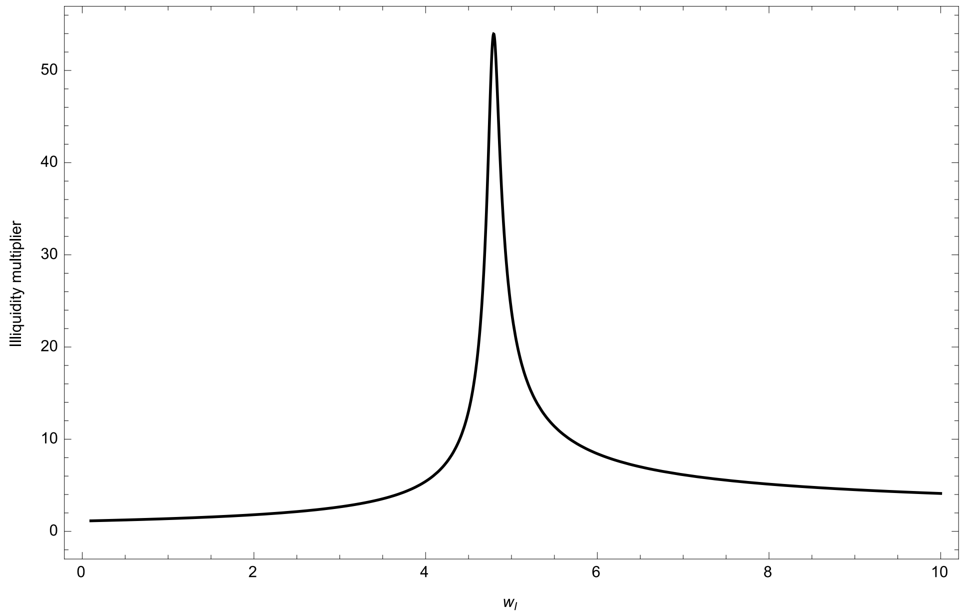

Figure 2.9.2: Illiquidity multiplier: when w_I is close to a resonant value of 4.8, the economy exhibits sharply larger amplification. The parameter values are $N = 10$, $w_J = 1$, $\rho = 0.9$, $\tau_s = 0.01$, $\tau_J = 0.1$, $\tau_I = 6$.

variance the prices are more noisy for J-traders, so they learn less from prices and their price elasticity is higher). Two equilibrium points corresponding to the crossing of $\Lambda(l)$ and the 45 degree line from above shift down, while the other equilibrium point shifts up.

2.9.1 Illiquidity multiplier

Consider an equilibrium with a price impact equal to λ. Suppose that there is an unexpected shock to a parameter of the model so that the curve $\Lambda(l)$ shifts up by $d\Lambda$ at $l = \lambda$. Since the shock is unexpected, the traders will still behave as if the price impact was λ and the change of the "true" price impact is then $d\Lambda$, the direct effect of the shock.

If the shock is expected, the traders should adjust their behavior. Consider a stable equilibrium and the tÃ¢tonnement process described above. In the first step traders realize that the "true" price impact is higher by $d\Lambda$, so they will adjust their strategies as if the price impact was $\lambda + d\Lambda$. The "true" price impact corresponding to such a belief is $\Lambda(\lambda + d\Lambda) \approx \lambda + \Lambda'(\lambda)d\Lambda$. In the second step they realize that the "true" price impact is higher by $\Lambda'(\lambda)d\Lambda$. The "true" price impact corresponding to such a belief is $\Lambda(\lambda + \Lambda'(\lambda)d\Lambda) \approx \lambda + (\Lambda'(\lambda))^2 d\Lambda$ and so on. The total change in of the equilibrium price impact is the sum of changes of price impact at each step, so we get

$$\text{change of equilibrium price impact} = d\Lambda \cdot \mathcal{M},$$

where

$$\mathcal{M} \equiv 1 + \Lambda'(\lambda) + (\Lambda'(\lambda))^2 + (\Lambda'(\lambda))^3 + ... \qquad (2.9.7)$$

is *illiquidity multiplier*.[44]

Intuitively, the direct effect $d\Lambda$ gets amplified and the total effect of the shock to a parameter is $\mathcal{M} \cdot d\Lambda$. Therefore the illiquidity multiplier $\mathcal{M}$ characterizes quantitatively the amplification mechanism of the model.

It is well know that the geometric series (2.9.7) converge if and only if $|\Lambda'(\lambda)| < 1$, i.e. iff the equilibrium is stable. Applying the formula for the sum of the geometric series, one can get that in the stable equilibrium with price impact equal to λ the multiplier is given by

$$\mathcal{M} = \frac{1}{1 - \Lambda'(\lambda)}.$$

Repeating formally the above logic for unstable equilibria, one would get that the series (2.9.7) diverge. Therefore one can think of those equilibria as the ones with extreme amplification, so that the multiplier $\mathcal{M}$ explodes.

I summarize the above discussion by formally defining the multiplier.

Definition 4. Consider a stable equilibrium with a price impact equal to λ. The illiquidity multiplier is $\mathcal{M} \equiv \frac{1}{1-\Lambda'(\lambda)}$.

Figure 2.9.2 plots the illiquidity multiplier against w_I. The figure reminds of a resonance in physics: when w_I is close to a "resonant" value of 4.8, the economy exhibits sharply larger amplification.

2.9.2 Stability analysis

The stability is easily analyzed numerically. One have to evaluate $|\Lambda'(\lambda)|$ and see whether it is smaller than 1. In the Appendix 2.10.10 I provide a closed form expression, up to a solution of a cubic equation, for the derivative $\Lambda'(\lambda)$.

The stability analysis is represented in the Figure 2.4.2. We see that all equilibria in which $\beta_J(\tau_I)$ is increasing ("upward-sloping" equilibria) are unstable. To understand why, consider an increase in τ_I and examine its' effects on the strategies of

[44]This definition is analogous to the concept of illiquidity multiplier in Cespa and Foucault (2014).

J-traders. Suppose first that I-traders' strategies are unchanged. An increase in τ_I makes the prices more informative so the J-traders learn less from their signals and more from prices. The sensitivity β_J should therefore decrease. Therefore, in order for the equilibrium β_J to increase following an increase in τ_I, the I-traders demand should become more sensitive to their value, i.e. β should increase. It implies that the equilibrium price impact λ should decrease.

From the above discussion we know that an increase in τ_I in the "upward-sloping" equilibria should lead to a decrease in λ. I will demonstrate that this is only possible if the λ corresponds to the intersection of $\Lambda(l)$ and the 45 degree line from below, which implies that such an equilibrium is unstable.

Indeed, an increase in τ_I corresponds to an upward shift of the curve $\Lambda(l)$: given the same belief l about the price impact (and therefore the same strategies of I-traders) the prices become more informative, so J-traders reduce their elasticity which increases the "true" price impact $\Lambda(l)$. But if the curve $\Lambda(l)$ shifts up, its' intersection with a 45 degree line can shift down only if $\Lambda(l)$ this intersection is from below.

2.10 Proofs

In this section I will use the following notation:

$$\theta \equiv \frac{\tau_J + \tau_s}{\tau_s} > 1,\ \xi \equiv \rho\sqrt{\frac{\tau_J}{\tau_I}},\ \varkappa \equiv \sqrt{\frac{\tau_I/\tau_s}{1-\rho^2}} > 0,$$

$$\psi \equiv \frac{w_I}{Nw_J} > 0,\ \phi \equiv \varkappa\xi = \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{\frac{\tau_J}{\tau_s}}.$$

Unless stated otherwise, the proofs are not restricted to the case $\rho \geq 0$.

2.10.1 Proof of Theorem 3

Derivation of equilibrium

The proof is split into 4 steps.

Step 1. Guess.

I conjecture that I- and J-traders have the following linear schedules in equilibrium

$$x_i = \alpha + \beta v_I - \gamma p \text{ and } x_j = \alpha_J + \beta_J s_j - \gamma_J p. \tag{2.10.1}$$

Step 2. Residual supply and the equilibrium price function.

The above guess and the market-clearing rule (2.2.4) implies that I-traders face the following inverse residual supply

$$p = \iota + \lambda \cdot x, \tag{2.10.2}$$

where the stochastic intercept ι has full support[45] and the slope is given by

$$\frac{1}{\lambda} = (N-1)\gamma + \gamma_J. \tag{2.10.3}$$

The J-traders are atomistic, so their inverse residual supply does not depend on the quantity they trade. Substituting (2.10.1) to (2.2.4) we find that it is given by

$$p = \frac{1}{\Gamma}\left(N\beta v_I + \beta_J v_J\right) + c_p, \tag{2.10.4}$$

where

$$\Gamma = N\gamma + \gamma_J, \text{ and } c_p = \frac{N\alpha + \alpha_J}{\Gamma}. \tag{2.10.5}$$

Equation (2.10.4) is also the equilibrium price function.

Step 3. Verify.

I-traders. The approach to solving I-traders problem relies on the idea of maximizing against the residual supply (Kyle (1989), Klemperer and Meyer (1989)). For a given realization of ι we find an optimal price-quantity pair $(x^*(\iota), p^*(\iota))$ on the residual supply curve (2.10.2) that maximizes the utility of a trader. Trader's optimal schedule is given parametrically by $(x^*(\iota), p^*(\iota))$ as a function of ι. The problem of finding an optimal price-quantity pair on the residual supply curve can be written as

$$\max_{(x,p)} \left(v_I - p\right) x - \frac{w_I}{2} x^2 \tag{2.10.6}$$

$$\text{s.t.: } p = \iota + \lambda \cdot x. \tag{2.10.7}$$

Taking the first order condition and eliminating ι using (2.10.7) yields the following

[45] The exact expression is $\iota = \lambda\left((N-1)\alpha + \alpha_J + (N-1)\beta v_I + \beta_J v_J\right)$.

expression for the optimal schedule of a trader i

$$x_i = \frac{1}{w_I + \lambda}(v_I - p). \tag{2.10.8}$$

The second order conditions require

$$w_I + 2\lambda > 0. \tag{2.10.9}$$

In what follows I assume that the second-order condition holds, which is justified by the Lemma 3. Comparing (2.10.8) and (2.10.1) yields

$$\beta = \gamma = \frac{1}{w_I + \lambda} > 0 \text{ and } \alpha = 0. \tag{2.10.10}$$

The guess for I-traders is therefore verified.

J-traders, being atomistic, have no price impact. For a given realisation of price they find the optimal quantity x, solving

$$\max_x \left(E[v_J|s_J, p] - p\right) x - \frac{w_J x^2}{2}.$$

The first order necessary and sufficient condition implies

$$x_j = \frac{1}{w_J}\left(E\left[v_J|s_j, p\right] - p\right). \tag{2.10.11}$$

To solve the inference problem of a J-trader I apply Lemma 4 according to which the price is informationally equivalent to a signal

$$\pi \equiv \frac{\Gamma p}{\beta_J + N\beta\xi} + c_\pi, \tag{2.10.12}$$

Where the expression for the constant c_π is given by (2.10.30). This signal can also be written as

$$\pi = v_J + \frac{1}{\sqrt{\tau_\pi}}\epsilon_\pi, \tag{2.10.13}$$

where $\epsilon_\pi \sim N(0,1)$ is independent of v_J and the noize ϵ_j in the signal s_j. Therefore, the signal π is an unbiased signal of v_J. According to Lemma 4 the precision of the signal is given by

$$\tau_\pi = \varkappa^2 \tau_s \left(\frac{\beta_J}{N\beta} + \xi \right)^2 . \tag{2.10.14}$$

Applying the Projection Theorem one gets

$$E\left[v_J | s_j, p\right] = \frac{\tau_J}{\tau} \bar{v}_J + \frac{\tau_s}{\tau} s_j + \frac{\tau_\pi}{\tau} \pi, \tag{2.10.15}$$

where

$$\tau \equiv \tau_J + \tau_s + \tau_\pi .$$

Substituting (2.10.15) and (2.10.12) to (2.10.11) and comparing to (2.10.1) one gets

$$\beta_J = \frac{1}{w_J} \frac{\tau_s}{\tau}, \tag{2.10.16}$$

$$\gamma_J = \frac{1}{w_J} \left(1 - \frac{\tau_\pi}{\tau} \frac{\Gamma}{\beta_J + N\beta\xi} \right), \tag{2.10.17}$$

$$\alpha_J = \frac{1}{w_J} \left(\frac{\tau_J}{\tau} \bar{v}_J + \frac{\tau_\pi}{\tau} c_\pi \right),$$

which verifies the guess for J-traders.

Step 4. Solve for coefficients.

To solve the model I introduce the quantity

$$\delta \equiv \varkappa \left(\frac{\beta_J}{N\beta} + \xi \right) > \phi, \tag{2.10.18}$$

and express all coefficients through it. The inequality is true because both β and β_J are positive (cf. (2.10.16) and (2.10.10)). From (2.10.14) we get

$$\frac{\tau_\pi}{\tau_s} = \delta^2,$$

which allows to rewrite (2.10.16) as

$$\beta_J = \left(w_J \left(\theta + \delta^2 \right) \right)^{-1} . \tag{2.10.19}$$

One can rewrite (2.10.18) as

$$\frac{1}{\beta} = w_I + \lambda = \frac{N}{\beta_J} \left(\frac{\delta}{\varkappa} - \xi \right)$$

Substituting (2.10.19) into which yields

$$\lambda = \frac{Nw_J}{\varkappa}\left(\delta-\phi\right)\left(\theta+\delta^2\right) - w_I. \tag{2.10.20}$$

One can express all coefficients of I-traders through δ by substituting the above expression for λ to (2.10.10).

After some algebra γ_J can be expressed as

$$\gamma_J = \frac{1}{w_J} - \delta\left(\delta-\phi\right)\Gamma. \tag{2.10.21}$$

Combining the above and $\Gamma = \gamma_J + \frac{N}{w_I+\lambda}$ one gets

$$\Gamma = \frac{\frac{1}{w_J}+\frac{N}{w_I+\lambda}}{1+\delta\left(\delta-\phi\right)}. \tag{2.10.22}$$

After some algebra, for α_J one can get the following expression:

$$\alpha_J = \frac{\beta\beta_J N(\tau_J + \delta\varkappa\tau_s(\xi\overline{v}_J - \overline{v}_I))}{\tau_s(\beta_J\delta\varkappa + \beta N)}.$$

Finally, we get the expression to pin down δ. It can be done substituting (2.10.22) to $\Gamma = \frac{1}{\lambda} + \frac{1}{w_I+\lambda}$ which yields, after some algebra

$$\lambda\left(w_I + Nw_J + \lambda\right) - w_J\left(1+\delta\left(\delta-\phi\right)\right)\left(w_I+2\lambda\right) = 0. \tag{2.10.23}$$

Substituting (2.10.20) to the above yields the following sextic equation in δ:

$$\begin{aligned} N\left(\left(\delta^2+\theta\right)\left(\delta-\phi\right)+\varkappa\right)\left(Nw_J\left(\delta^2+\theta\right)\left(\delta-\phi\right)-\varkappa w_I\right) - \\ -\kappa\left(\delta\left(\delta-\phi\right)+1\right)\left(2Nw_J\left(\delta^2+\theta\right)\left(\delta-\phi\right)-\varkappa w_I\right) &= 0. \end{aligned} \tag{2.10.24}$$

The analysis of the number of equilibria is performed in the section 2.10.1.

I summarize the results for BNE. Given δ that solves (2.10.24) and satisfies $\frac{Nw_J}{\varkappa}\left(\delta-\phi\right)\left(\theta+\delta^2\right) > \frac{w_I}{2}$, the coefficients in BNE are given by

$$\beta = \gamma = \frac{1}{w_I+\lambda} > 0,\ \alpha = 0,\ \text{where} \tag{2.10.25}$$

$$\lambda = \frac{Nw_J}{\varkappa}\left(\delta-\phi\right)\left(\theta+\delta^2\right) - w_I. \tag{2.10.26}$$

$$\beta_J = \left(w_J\left(\theta+\delta^2\right)\right)^{-1} > 0,\ \gamma_J = \frac{1}{w_J} - \delta\left(\delta-\phi\right)\Gamma, \tag{2.10.27}$$

$$\alpha_J = \frac{\beta\beta_J N(\overline{v}_J(\delta\phi\tau_s+\tau_J) - \delta\varkappa\tau_s\overline{v}_I)}{\tau_s(\beta_J\delta\varkappa+\beta N)}, \tag{2.10.28}$$

where

$$\Gamma = \frac{\frac{1}{w_J}+\frac{N}{w_I+\lambda}}{1+\delta\left(\delta-\phi\right)} > 0. \tag{2.10.29}$$

Lemma 3. *There is no equilibrium in which $2\lambda + w_I \leq 0$.*

Proof. Suppose $w_I + 2\lambda < 0$. In that case for any realisation of ι the profit maximizing quantity in the problem (2.10.6) is infinite and the market will not clear. Therefore such an equilibrium does not exist.

In the case $w_I + 2\lambda = 0$, the problem is (2.10.8) is linear and the demand of I-itraders is only finite when $p = v_I$ which is the case only when at least one of the traders submits perfectly price elastic schedule. In the latter case the price impact of other I-traders is $\lambda = 0$, not $\lambda = -\frac{w_I}{2}$, a contradiction. □

Lemma 4. *In a linear equilibrium characterised by the schedules (2.2.5) the price function is informationally equivalent to the sufficient statistic*

$$\pi \equiv \frac{\Gamma p}{\beta_J + N\beta\xi} + c_\pi,$$

where

$$c_\pi \equiv \frac{\xi\overline{v}_J - \overline{v}_I - \Gamma\frac{c_p}{N\beta}}{\frac{\beta_J}{N\beta}+\xi}. \tag{2.10.30}$$

The sufficient statistic π can be written as

$$\pi = v_J + \frac{1}{\varkappa\sqrt{\tau_s}\left(\frac{\beta_J}{N\beta}+\xi\right)}\epsilon_\pi, \tag{2.10.31}$$

where

$$\epsilon_\pi \equiv \varkappa\sqrt{\tau_s}\left(v_I - \overline{v}_I - \xi\left(v_J - \overline{v}_J\right)\right).$$

Moreover $\epsilon_\pi \sim N(0,1)$ and, for any j, ϵ_π is independent of v_J and the noize ϵ_j in *the signal s_j. The precision of π is given by*

$$\tau_\pi \equiv Var[\pi|v_J]^{-1} = \varkappa^2 \tau_s \left(\frac{\beta_J}{N\beta} + \xi\right)^2.$$

Proof. The π is a linear transformation of, and hence is informationally equivalent to, the price p.

Given the price function (2.10.4), it can be checked by a direct calculation that (2.10.31) holds.

It is clear that ϵ_π is distributed noirmally with mean zero. The variance can be computed as

$$\begin{aligned}
\mathrm{Var}[\epsilon_\pi] &= \varkappa^2 \tau_s \mathrm{Var}[v_I - \bar{v}_I - \xi\left(v_J - \bar{v}_J\right)] \\
&= \frac{\tau_I}{1-\rho^2}\left(\frac{1}{\tau_I} + \xi^2 \frac{1}{\tau_J} - 2\xi \frac{\rho}{\sqrt{\tau_I \tau_J}}\right) \\
&= \frac{\tau_I}{1-\rho^2}\left(\frac{1}{\tau_I} + \rho^2 \frac{\tau_J}{\tau_I}\frac{1}{\tau_J} - 2\rho\sqrt{\frac{\tau_J}{\tau_I}}\frac{\rho}{\sqrt{\tau_I \tau_J}}\right) \\
&= 1.
\end{aligned}$$

The ϵ_π is independent of ϵ_j, because ϵ_π is a linear combination of v_I and v_J and the two are independent of ϵ_j. To see that ϵ_π is independent of v_J compute

$$\begin{aligned}
\frac{\mathrm{cov}(\epsilon_\pi, v_J)}{\varkappa\sqrt{\tau_s}} &= \mathrm{cov}(v_I - \bar{v}_I - \xi\left(v_J - \bar{v}_J\right), v_J) \\
&= \mathrm{cov}(v_I, v_J) - \rho\sqrt{\frac{\tau_J}{\tau_I}}\frac{1}{\tau_J} \\
&= 0,
\end{aligned}$$

which, given joint normality of v_I and v_J (and hence ϵ_π and v_J) implies independence.

The formula for precision follows immediately from (2.10.31). □

Existence and sufficient conditions for uniqueness and multiplicity of equilibria

Existence

Lemma 5. *If $\phi > -\sqrt{3\theta}$ there exists at least one BNE. In particular, if $\rho \geq 0$ there always exists at least one BNE.*

Proof. According to the Theorem 3 the BNE exists if and only if there is a solution δ to the sextic equation (2.10.24) that satisfies

$$\frac{Nw_J}{\varkappa}(\delta-\phi)\left(\theta+\delta^2\right) > \frac{w_I}{2}. \tag{2.10.32}$$

Denote by $\check{\delta}$ the solution to

$$\frac{Nw_J}{\varkappa}\left(\check{\delta}-\phi\right)\left(\theta+\check{\delta}^2\right) = \frac{w_I}{2}. \tag{2.10.33}$$

Such a solution is unique and the left hand side of the above is increasing in δ for $\delta > \hat{\delta}$.[46]

After substituting $\delta = \check{\delta}$ to (2.10.24) it becomes

$$-N\left(\left(\check{\delta}^2+\theta\right)\left(\check{\delta}-\phi\right)+\varkappa\right)\frac{\varkappa w_I}{2} < 0, \tag{2.10.34}$$

i.e. the polynomial is negative at $\delta = \check{\delta}$. On the other hand, the leading coefficient of the polynomial (2.10.24) is $N^2 w_J > 0$, therefore it becomes positive for δ large enough. By the Intermediate Value Theorem there should be a solution $\delta^* > \check{\delta}$ to (2.10.24). Since for $\delta > \check{\delta}$ and the function $\frac{Nw_J}{\varkappa}(\delta-\phi)\left(\theta+\delta^2\right)$ is strictly increasing in δ, the second order condition (2.10.32) holds for $\delta = \delta^*$. □

Sufficient conditions for the uniqueness

Lemma 6. *The BNE is unique if $\xi > \underline{\xi}$, where $\underline{\xi} < 1$ is given by (2.10.44). In the case $\rho \geq 0$ the sufficient condition can be written as $\tau_I < \overline{\tau}_1$, where $\overline{\tau}_1$ is given by (2.10.45).*

Proof. The BNE corresponds to a solution of a system of equations (2.10.20) and (2.10.23) satisfying the second order condition $\lambda > -w_I/2$. After the change of variables

$$l \equiv \frac{2\lambda + w_I}{2Nw_J} > 0, \tag{2.10.35}$$

the system becomes

$$l = l(\delta) \equiv \frac{\left(\delta^2+\theta\right)(\delta-\phi)}{\varkappa} - \frac{\psi}{2} \tag{2.10.36}$$

[46] The solution exists, because LHS is less than RHS at $\delta = \phi$ and is greater than RHS for δ large enough. Since for $\delta < \phi$ the LHS is negative, there are no solutions in that region and we may consider only $\delta > \phi$. If $\phi > 0$, then the LHS is strictly increasing for $\delta > \phi$. If $-\sqrt{3\theta} < \phi < 0$ then the LHS is striclty increasing for all δ. In any case there is at most one solution and LHS is increasing in δ for $\delta > \hat{\delta}$.

$$l(Nl + N - 2\left(1 + \delta(\delta - \phi)\right)) = N\left(\left(\frac{\psi}{2}\right)^2 + \frac{\psi}{2}\right). \qquad (2.10.37)$$

We are now looking for the solutions to the above system satisfying $l > 0$.

Denote

$$y \equiv \delta(\delta - \phi).$$

From (2.10.37) one can express y through l as follows

$$y = \frac{4l^2 n + 4l(N-2) - N\psi(\psi+2)}{8l}. \qquad (2.10.38)$$

The equation (2.10.36) can be written as

$$l = \frac{\delta y + \theta(\delta - \phi)}{\varkappa} - \frac{\psi}{2},$$

which allows to get an expression of δ through y and l:

$$\delta = \frac{2\theta\phi + \varkappa\psi + 2\varkappa l}{2(\theta + y)}.$$

Substituting y from (2.10.38) the above becomes after some algebra

$$\delta = \delta(l) \equiv \frac{2\varkappa}{N} \frac{l\left(l + \theta\xi + \frac{\psi}{2}\right)}{(l - l^+)(l - l^-)}, \qquad (2.10.39)$$

where

$$l^\pm \equiv \frac{-G \pm \sqrt{G^2 + F}}{2},$$

$$G \equiv 1 + \frac{2(\theta - 2)}{N} > 1,$$

$$F \equiv 2\psi + \psi^2 > 0.$$

One can show that

$$0 < l^+ < \frac{\psi}{2},\ l^- < 0.$$

The idea for proceeding further is the following. Equation (2.10.36) gives an explicit expression for the function $l(\delta)$. Equation (2.10.39) expresses explicitly the function $\delta(l)$. The sufficient conditions for uniqueness can be obtained by analyzing how many times the two functions intersect. In what follows I consider the behavior of the two curves on the coordinate plane in which l is a vertical and δ is a horiozontal axis.

I make the following additional assumption

$$\xi > -\min\left(\frac{\psi}{2\theta}, \frac{\sqrt{3\theta}}{\varkappa}\right). \tag{2.10.40}$$

Note that when $\rho \geq 0$ the above condition always holds.

Condition (2.10.40) implies $\xi > -\frac{\psi}{2\theta}$, which implies that $\delta(l)$ is positive for $l > l^+$. Condition $\xi > -\frac{\sqrt{3\theta}}{\varkappa}$ together with $\xi > -\frac{\psi}{2\theta}$ implies that in the region $l > 0$ the curve $l(\delta)$ lies to the right of the vertical line $\delta = 0$.[47] The two curves can therefore intersect only in the region $l > l^+$ and $\delta > 0$. Since $l(\delta) < 0 < l^+$ for $\delta < \phi$ we may actually restrict our attention to the region $l > l^+$ and $\delta > \max(\phi, 0)$.

It is easy to show that for $\delta > \max(\phi, 0)$ the function $l(\delta)$ is strictly increasing.

Lemma 7 implies that for l such that $\delta(l) > \frac{2\varkappa}{N}$ the function $\delta(l)$ is strictly decreasing. Suppose that

$$\phi > \frac{2\varkappa}{N}. \tag{2.10.41}$$

The intersection of the two curves can only occur in the region $\delta > \phi > \frac{2\varkappa}{N} > 0$ and $l > l^+$. In this region $\delta(l)$ is strictly decreasing, whereas $l(\delta)$ is strictly increasing so they intersect in at most one point. Condition (2.10.41) is equivalent to

$$\xi > \frac{2}{N}. \tag{2.10.42}$$

We also know from the Lemma 7 that if $\xi > \frac{4(\theta-1)+N(2-\psi)}{2\theta N}$, then the function $\delta(l)$ is strictly decreasing in l for all $l > l^+$. Therefore

$$\xi > \frac{4(\theta-1)+N(2-\psi)}{2\theta N} \tag{2.10.43}$$

is also a sufficient condition for uniqueness. Combining (2.10.40), (2.10.42) and (2.10.43) one gets

$$\xi > \underline{\xi} = \max\left(-\min\left(\frac{\psi}{2\theta}, \frac{\sqrt{3\theta}}{\varkappa}\right), \min\left(\frac{2}{N}, \frac{4(\theta-1)+N(2-\psi)}{2\theta N}\right)\right), \tag{2.10.44}$$

[47]Note that for $l(\delta)$ can not be positive if $\delta \leq \phi$. If $\xi \geq 0$ then the condition $\delta > \phi = \kappa\xi > 0$ ensures that the statement is true. In the case $\xi < 0$ the condition $\xi > -\frac{\sqrt{3\theta}}{\varkappa}$ implies that the function $l(\delta)$ is strictly increasing. Consider $l(0)$. One can compute $l(0) = -\theta\xi - \frac{\psi}{2}$, which is negative since (2.10.40) implies that $\xi > -\frac{\psi}{2\theta}$. Therefore the function $l(\delta)$ can only be positive for strictly positive δ.

moreover, since $N > 1$ we have

$$\underline{\xi} \leq 1.$$

In the case $\rho \geq 0$ the condition (2.10.44) can be written as

$$\tau_I < \overline{\tau}_1 \equiv \left(\frac{\rho\sqrt{\tau_J}}{\min\left(\frac{2}{N}, \frac{4(\theta-1)+N(2-\psi)}{2\theta N}\right)} \right)^2. \tag{2.10.45}$$

□

Lemma 7. *Suppose that (2.10.40) holds. Then: 1) in the region $l > l^+$ the function $\delta(l)$ is strictly decreasing in l for l such that $\delta(l) > \frac{2\varkappa}{N}$; 2) the function $\delta(l)$ is strictly decreasing in l for all $l > l^+$, provided that*

$$\xi > \frac{4(\theta-1)+N(2-\psi)}{2\theta N}. \tag{2.10.46}$$

Proof. One can find that for $l > l^+$ the condition $\delta(l) > \frac{2\varkappa}{N}$ is equivalent to

$$l\left(l^+ + l^- + \theta\xi + \psi/2\right) > l^+ l^-. \tag{2.10.47}$$

Computing the derivative of $\delta(l)$ one gets

$$\begin{aligned} \delta'(l) &= \frac{\varkappa\left(2l\left(l^- l^+ - l(\theta\xi + l^- + l^+ + \psi/2)\right) + l^- l^+(2\theta\xi + \psi + 2l)\right)}{N(l-l^-)^2(l-l^+)^2} \\ &\equiv \frac{n(l)}{N(l-l^-)^2(l-l^+)^2}. \end{aligned}$$

The above is negative provided that (2.10.47) holds, which proves the first claim.

The sign of $\delta'(l)$ is the same as the sign of its' nominator $n(l)$. Below I prove that $n(l^+) < 0$ and that $n(l)$ is decreasing provided that (2.10.46) holds, which proves the second claim. Indeed,

$$n(l^+) = \varkappa l^+(l^- - l^+)(2\left(\theta\xi + \psi/2\right) + 2l^+) < 0.$$

The derivative of $n(l)$ is given, by

$$n'(l) = \kappa(4l^- l^+ - 2l(2(\theta\xi + l^- + l^+) + \psi)),$$

which is negative if

$$2(\theta\xi + l^- + l^+) + \psi > 0.$$

The above is equivalent to $\xi > \frac{4(\theta-1)+N(2-\psi)}{2\theta N}$.

□

Sufficient conditions for multiplicity of BNE, $\rho \geq 0$

Lemma 8. *Suppose $N > 4$. There are at least three equilibria if $\underline{\tau_2} < \tau_I < \overline{\tau_2}$ and $w_I < \overline{w}$, where $\overline{\tau}_1 < \underline{\tau_2} < \overline{\tau_2}$ and the expressions for the tresholds $\left(\underline{\tau_2}, \overline{\tau_2}, \overline{w}\right)$ are given by (2.10.53-2.10.55).*

Proof. Denote

$$Q \equiv -4N\xi + 8\xi + 4\psi$$
$$T \equiv 16N^2\xi\psi\left(\xi - \frac{2}{N}\right)(\psi + 2).$$

Assume that

$$Q < 0,\ \xi < \frac{1}{N},\ Q^2 + T > 0,\ \psi < 1,\ N > 4. \tag{2.10.48}$$

Consider all solutions to

$$\delta(l) = \phi. \tag{2.10.49}$$

If the conditions (2.10.48) hold then there exist two solutions to (2.10.49) given by

$$L^{\pm} = \frac{-Q \pm \sqrt{Q^2 + T}}{8N\left(\frac{2}{N} - \xi\right)}$$

Moreover, both solutions $L^{\pm} > l^{+}$[48]. The fact that there are two solutions to (2.10.49) implies that the function $\delta(l)$ attains local minimum in the region $l > l^+$ and this minimum is less than ϕ.

Consider also all solutions to

$$\delta(l) = \frac{\varkappa}{N}.$$

Given that (2.10.48) holds there are two solutions to the above. Denote the maximal

[48]It is easy to see that both solutions are positive. But $\delta(L) = \phi > 0$ is positive only if $L > l^+$.

of them by L_m. One can calculate

$$L_m = \frac{1}{2}\left(Q_m + \sqrt{Q_m^2 + T_m}\right) > L^+, \text{ where}$$

$$Q_m \equiv \frac{2(\theta-1)}{N} + 1 - \frac{2\theta\phi}{\varkappa} - \psi,$$

$$T_m \equiv -\left(\psi^2 + 2\psi\right)$$

If

$$L_m < l\left(\frac{\varkappa}{N}\right) = \frac{\left(\varkappa^2 + \theta N^2\right)(\varkappa - N\phi)}{\varkappa N^3} - \frac{\psi}{2} \equiv l_m, \tag{2.10.50}$$

then there are at least three equilibria.

The condition $Q < 0$ is equivalent to

$$\xi > \frac{\psi}{N-2}. \tag{2.10.51}$$

The condition $Q^2 + T > 0$ holds provided that[49]

$$\xi > \frac{2\psi(N(\psi+3)-2)}{N\left(N(\psi+1)^2-4\right)+4} \text{ and } N\left(N(\psi+1)^2-4\right)+4 > 0. \tag{2.10.52}$$

The second part of the above holds given (2.10.48). Note that

$$\frac{2\psi(N(\psi+3)-2)}{N\left(N(\psi+1)^2-4\right)+4} < \frac{8\psi}{N-4} > \frac{\psi}{N-2}$$

Therefore (2.10.51) and (2.10.52) hold if the weaker condition holds:

$$\xi > \underline{\xi_1} \equiv \frac{8\psi}{N-4}.$$

The above can be written as

$$\tau_I < \frac{\rho^2 \tau_J}{\underline{\xi}_1^2} \equiv \overline{\tau}_2. \tag{2.10.53}$$

Suppose that

$$l_m - Q_m > 0.$$

[49]Indeed

$$Q^2 + T = 16\xi^2\left(N\left(N(\psi+1)^2-4\right)+4\right) - 32\xi\psi(N(\psi+3)-2) + 16\psi^2.$$

Condition (2.10.52) ensures that the first two terms are positive.

Then (2.10.50) holds[50]. The above can be written as

$$\left(\frac{\varkappa^2}{N^2}-\theta\right)\left(\frac{1}{N}-\xi\right)>1-\frac{2}{N}-\frac{\psi}{2}.$$

Assume

$$\xi<\frac{1}{2N}.$$

Then the LHS of the above is greater than $\left(\frac{\varkappa^2}{N^2}-\theta\right)\frac{1}{2N}$ and the constraint holds provided that

$$\frac{\varkappa^2}{N^2}-\theta>2N-4-N\psi,$$

which is equivalent to

$$\tau_I>(1-\rho^2)\tau_s N^2\left(2N-4-N\psi+\theta\right).$$

The above holds if the stricter inequality holds:

$$\tau_I>(1-\rho^2)\tau_s N^2\left(2N-4+\theta\right).$$

The constraint that $\xi<\frac{1}{2N}$ implies that

$$\tau_I>4N^2\rho^2\tau_J.$$

The above two conditions hold provided that

$$\tau_I>\underline{\tau_2}\equiv\max\left(4N^2\rho^2\tau_J,(1-\rho^2)\tau_s N^2\left(2N-4+\theta\right)\right). \tag{2.10.54}$$

It is clear that

$$\underline{\tau_2}>4N^2\rho^2\tau_J>\overline{\tau}_1.$$

We finaly derive the conditions when $\underline{\tau_2}<\overline{\tau}_2$. One gets

$$\sqrt{\underline{\tau_2}}<\frac{\rho\sqrt{\tau_J}}{\underline{\xi}_1}=\frac{\rho\sqrt{\tau_J}}{8\psi}(N-4),$$

[50] Indeed (2.10.50) is equivalent to

$$Q_m^2+T_m-(2l_m-Q_m)^2=2l_m(2Q_m-2l_m)+T_m<0,$$

which is true.

which is equivalent to

$$w_I < \overline{w} \equiv w_J \rho \frac{N(N-4)}{8} \sqrt{\frac{\tau_J}{\underline{\tau_2}}}. \tag{2.10.55}$$

□

2.10.2 Proof of Proposition 9

According to Lemmas 9-13 the equilibrium objects considered in the Proposition can be written in terms of the model parameters and

$$\delta \equiv \varkappa \left(\frac{\beta_J}{N\beta} + \xi \right) = \sqrt{\frac{\tau_\pi}{\tau_s}}. \tag{2.10.56}$$

The liquidity is decreasing in δ, whereas the information efficiency is increasing in δ. The two have the opposite rankings. It is established in the proof of Proposition 16 that the welfare is decreasing in δ if (2.4.1) holds.

2.10.3 Proof of Proposition 10

Throughout the proof I maintain the assumption that $\rho \geq 0$.

I first start with the case of crash(jump) understood as a switch from one equilibrium to another. According to Lemmas 9-13 the equilibrium objects considered in the Proposition can be written in terms of the model parameters and

$$\delta \equiv \varkappa \left(\frac{\beta_J}{N\beta} + \xi \right) = \sqrt{\frac{\tau_\pi}{\tau_s}}. \tag{2.10.57}$$

When there is a switch from one equilibrium to another, parameters of the model obviously do not change, and the only thing that changes is δ. The change in delta leads to a change to all equilibrium quantities. Therefore, given the monotonicity of the equilibrium objects in δ (Lemmas 9-13), once we understand how the δ changes if there is a price crash we know how the equilibrium objects change.

From the Lemma 12 it is clear that if there is a price crash, there is a crash (jump) in δ if $\bar{v}_I < \bar{v}_J$ ($\bar{v}_I > \bar{v}_J$). The statements of the proposition then follow from the monotonicity of λ, σ_p, $\mathcal{I}$ and V in δ which follow from Lemmas 9-13.

If the crash(jump) is understood as a situation in which the sensitivity of endogenous object to model parameters is infinite, the proof works as follows. Suppose the model parameter of interest is τ_s. We can write

$$\lambda = \lambda(\delta, \tau_s).$$

The sensitivity can be computed as

$$\frac{\partial \lambda}{\partial \tau_s} = \lambda_\delta(\delta, \tau_s)\frac{\partial \delta}{\partial \tau_s} + \lambda_{\tau_s}(\delta, \tau_s).$$

It can be seen by direct computation that $\lambda_\delta(\delta, \tau_s)$ and $\lambda_{\tau_s}(\delta, \tau_s)$ are both finite as long as δ is finite. In general, all equilibrium quantities are smooth functions of δ and the model parameters. The δ latter is finite, because the only case when iti is not is when $\beta = 0$, which is not possible due to second-order condition $w_I + 2\lambda > 0$. Therefore $\frac{\partial \lambda}{\partial \tau_s} = -\infty$ iff $\frac{\partial \delta}{\partial \tau_s} = -\infty$. Analogous statements can be formulated for σ_p, $\mathcal{I}$, $E[p]$ and V.

Expressing the endogenous objects through δ and the parameters of the model

In this section I show that the equilibrium objects considered in the Proposition 10 can be expressed through δ given by (2.10.57) and the parameters of the model.

Price impact

Lemma 9. *One can write the price impact λ as a function of δ and the parameters of the model, moreover $\frac{\partial \lambda}{\partial \delta} > 0$.*

Proof. Equation (2.10.26) implies that

$$\lambda = \frac{N w_J}{\varkappa} (\delta - \phi) \left(\theta + \delta^2\right) - w_I,$$

which is an increasing function of for $\delta > \phi$, if $\rho \geq 0$. □

Volatility

Lemma 10. *One can write the volatiltiy σ_p as a function of δ and the parameters of the model, moreover $\frac{\partial \sigma_p}{\partial \delta} > 0$.*

Proof. The volatility can be computed as follows. Using Lemma 4 one can write

$$\begin{aligned}
\mathrm{Var}(p) &= \mathrm{Var}\left(\frac{\beta_J + N\beta\xi}{\Gamma}\pi\right) \\
&= \mathrm{Var}\left(\frac{N\beta}{\Gamma}\frac{\delta}{\varkappa}\left(v_J + \frac{1}{\sqrt{\tau_s}\delta}\epsilon_\pi\right)\right) \\
&= \left(\frac{N\beta}{\Gamma}\frac{\delta}{\varkappa}\right)^2\frac{1}{\tau_J} + \left(\frac{N\beta}{\Gamma\varkappa}\right)^2\frac{1}{\tau_s}.
\end{aligned}$$

One can compute

$$\begin{aligned}
\frac{N\beta}{\Gamma} &= \frac{N}{(w_I+\lambda)\left(\frac{1}{\lambda}+\frac{1}{w_I+\lambda}\right)} \\
&= \frac{N}{\frac{w_I}{\lambda}+2}.
\end{aligned}$$

The above is positive and is increasing in λ and, given the monotonicity of $\lambda(\delta)$, also in δ. Hence, $\sigma_p'(\delta) > 0$. □

Information efficiency

Lemma 11. *One can write the information efficiency $\mathcal{I}$ as a function of δ and the parameters of the model, moreover $\frac{\partial\mathcal{I}}{\partial\delta} > 0$.*

Proof. The information efficiency can be written as

$$\begin{aligned}
\mathcal{I} &\equiv \frac{\mathrm{Var}(v_J)}{\mathrm{Var}(v_J|s_j,p)} \\
&= \frac{\tau_J+\tau_s+\tau_\pi}{\tau_J} \\
&= \frac{\tau_J+\tau_s(1+\delta^2)}{\tau_J},
\end{aligned}$$

which is increasing in δ. □

Price

Lemma 12. *One can write the price $E[p]$ as a function of δ and the parameters of the model, moreover sign $\left(\frac{\partial E[p]}{\partial\delta}\right)$ = sign $(\bar{v}_J - \bar{v}_I)$.*

Proof. Expected price sets net expected demand to zero. Expected total demand of I traders is given by

$$X = \frac{N}{w_I+\lambda}\left(\bar{v}_I - E[p]\right). \tag{2.10.58}$$

The expected demand of J-traders is given by

$$\begin{aligned} x_J &= \frac{1}{w_J} E\left[\frac{\tau_s}{\tau} v_J + \frac{\tau_\pi}{\tau}\pi + \frac{\tau_J}{\tau}\bar{v}_J - p\right] \\ &= \frac{1}{w_J}\left(\bar{v}_J - E[p]\right). \end{aligned}$$

Equalising the expected demand and supply we find

$$E[p] = \frac{N w_J \bar{v}_I + (w_I + \lambda)\,\bar{v}_J}{w_I + \lambda + N w_J}. \tag{2.10.59}$$

Taking the derivative of the above one can find

$$\operatorname{sign}\left(\frac{\partial E[p]}{\partial \lambda}\right) = \operatorname{sign}\left(\bar{v}_J - \bar{v}_I\right),$$

which, given the monotonicity of λ in δ proves the Lemma. □

Trading volume

Lemma 13. *The trading volume is given by (2.10.67). There exists $\underline{\tau}_J$ such that if*

$$\tau_I < 1 - \rho^2, \ \ \textit{and}\ \tau_J > \underline{\tau}_J \tag{2.10.60}$$

the trading volume is a decreasing function of δ.

Proof. Denote

$$X = N x_I\left(p^*\right),$$

the aggregate trade of I-traders. Denote also

$$u_j \equiv E[v_J | s_j, p] dj = \frac{\tau_\pi}{\tau}\left(v_J + \epsilon_\pi\right) + \frac{\tau_s}{\tau}\left(v_J + \epsilon_j\right) + \frac{\tau_J}{\tau}\bar{v}_J$$

the ex-post value of a trader j. Denote

$$u_J \equiv \int_0^1 u_j dj = \frac{\tau_\pi}{\tau}\left(v_J + \epsilon_\pi\right) + \frac{\tau_s}{\tau} v_J + \frac{\tau_J}{\tau}\bar{v}_J$$

the aggregate ex-post value of J-traders. One can also write the above in a more convinient way

$$u_J = \frac{\tau_\pi + \tau_s}{\tau}\left(v_J - \bar{v}_J\right) + \frac{\tau_\pi}{\tau}\epsilon_\pi + \bar{v}_J.$$

The market clearing condition can be written as

$$N\frac{v_I - p^*}{w_I + \lambda} + \frac{u_J - p^*}{w_J} = 0.$$

From the above one can express the market-clearing price and the aggregate trade of I-traders

$$p^* = \frac{v_I N w_J + u_J(\lambda + w_I)}{\lambda + N w_J + w_I},$$

$$X = G\left(v_I - u_J\right) \text{ where } G \equiv \frac{N}{\lambda + N w_J + w_I}. \tag{2.10.61}$$

One can also compute

$$x_j\left(p^*\right) = -X + \beta_J \epsilon_j. \tag{2.10.62}$$

According to the Lemma 4 one can write

$$v_I = \overline{v}_I + \xi(v_J - \overline{v}_J) + \frac{1}{\varkappa\sqrt{\tau_s}}\epsilon_\pi,$$

which allows to rewrite

$$v_I - u_J = \overline{v}_I - \bar{v}_J + c_v(v_J - \overline{v}_J) + c_\epsilon \epsilon_\pi, \text{ where} \tag{2.10.63}$$

$$c_v \equiv \xi - \frac{\tau_\pi + \tau_s}{\tau}, \text{ and } c_\epsilon \equiv \frac{1}{\varkappa\sqrt{\tau_s}} - \frac{\tau_\pi}{\tau}. \tag{2.10.64}$$

Substituting $\frac{\tau_s}{\tau} = w_J \beta_J$ the above expressions become

$$c_v = \xi - 1 + \frac{\tau_J}{\tau_s} w_J \beta_J \text{ and } c_\epsilon \equiv \frac{1}{\varkappa\sqrt{\tau_s}} - 1 + \frac{\tau_J + \tau_s}{\tau_s} w_J \beta_J, \tag{2.10.65}$$

and it is clear that both c_v and c_ϵ are decreasing in δ since $\beta_J = \frac{1}{w_J(\theta + \delta^2)}$ is decreasing in δ.

I next use the well-known fact that for $Y \sim N(\mu, \sigma^2)$ the mean of $|Y|$ is given by

$$E[|Y|] = M\left(\mu, \sigma^2\right) \equiv \sigma\sqrt{\frac{2}{\pi}} \exp\left(\frac{-\mu^2}{2\sigma^2}\right) - \mu \cdot \operatorname{erf}\left(\frac{-\mu}{\sqrt{2}\sigma}\right) \tag{2.10.66}$$

where $\operatorname{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, \mathrm{d}t$ is the error function. The trading volume then can be written as

$$V = \frac{1}{2}\left(M\left(\mu_X, \sigma_X\right) + M\left(\mu_{x_j}, \sigma^2_{x_j}\right)\right),$$

where

$$\mu_X \equiv E[X] = G\left(\overline{v}_I - \bar{v}_J\right),$$

$$\mu_{x_j} \equiv E\left[x_j\left(p^*\right)\right] = -\mu_X,$$

$$\sigma_X^2 \equiv \mathrm{Var}[X] = G^2\left(c_v^2/\tau_J + c_\epsilon^2\right),$$

$$\sigma_{x_j}^2 \equiv \mathrm{Var}[x_j\left(p^*\right)] = \sigma_X^2 + \beta_J^2/\tau_s.$$

Applying the Lemma 14 one can write

$$V = \frac{1}{2}\left(M\left(|\mu_X|, \sigma_X^2\right) + M\left(|\mu_X|, \sigma_X^2 + \beta_J^2/\tau_s\right)\right). \tag{2.10.67}$$

If c_v and c_ϵ are positive, then $|\mu_X|$, σ_X^2 and $\sigma_X^2 + \beta_J^2/\tau_s$ are all decreasing in δ. Lemma 14 then implies that the trading volume is a decreasing function of δ.

The sufficient condition for c_v and c_ϵ to be positive can be found from (2.10.64):

$$\xi - \frac{\tau_\pi + \tau_s}{\tau_J + \tau_\pi + \tau_s} > 0 \text{ and } \frac{1}{\varkappa\sqrt{\tau_s}} - \frac{\tau_\pi}{\tau} > 0 \tag{2.10.68}$$

The inequalities hold if

$$\xi > \frac{\tau_s}{\tau_J}\left(1 + \bar{\delta}^2\right) \text{ and} \tag{2.10.69}$$

$$\frac{1}{\varkappa\sqrt{\tau_s}} > 1, \tag{2.10.70}$$

where the value of $\bar{\delta}$ is given by Lemma 15.

Consider τ_J that solves

$$\inf\left\{\tau_J \geq 0 : \xi > \frac{\tau_s}{\tau_J}\left(1 + \bar{\delta}^2\right)\right\}.$$

Denote the solution $\underline{\tau}_J$. The solution exists and is unique, since the left-hand side of $\xi > \frac{\tau_s}{\tau_J}\left(1 + \bar{\delta}^2\right)$ is increasing in τ_J and is unbouneded, whereas the right-hand side is decreasing in τ_J. The inequality (2.10.69) holds if $\tau_J > \underline{\tau}_J$.

The inequality (2.10.70) holds if

$$\tau_I < 1 - \rho^2.$$

□

Lemma 14. *The function* $M(\mu,\sigma^2)$ *defined by (**??**) is symmetric in* μ*, i.e.* $M(-\mu,\sigma^2) = M(\mu,\sigma^2) = M(|\mu|,\sigma^2)$*, is incresing in* μ *if* $\mu \geq 0$ *and is increasing in* σ*.*

Proof. The fact that $M(-\mu,\sigma^2) = M(\mu,\sigma^2)$ follows tion of $M(\cdot)$ and the symmetry of the error function $\mathrm{erf}(-z) = -\mathrm{erf}(z)$. Taking the derivatives of $M(\cdot)$ one gets

$$\frac{\partial}{\partial\mu} M(\mu,\sigma^2) = \mathrm{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right),$$

$$\frac{\partial}{\partial\sigma} M(\mu,\sigma^2) = \sqrt{\frac{2}{\pi}} e^{-\frac{\mu^2}{2\sigma^2}},$$

which proves the Lemma. □

Lemma 15. *In equilibrium* $\delta < \overline{\delta}$*, where* $\overline{\delta}$ *is given by (2.10.71);* $\overline{\delta}/\sqrt{\tau_J}$ *decreases in* τ_J*.*

Proof. From the definition of δ

$$\delta \equiv \varkappa\left(\frac{\beta_J}{N\beta} + \xi\right) = \varkappa\left(\frac{\beta_J}{N}(w_I+\lambda) + \xi\right) < \varkappa\left(\frac{1}{Nw_J\theta}(w_I+\lambda) + \xi\right)$$

$$< \varkappa\left(\frac{1}{Nw_J}(w_I+\lambda) + \xi\right).$$

We next find an upper bound for λ. One can write

$$\frac{1}{\lambda} = \Gamma - \gamma < \Gamma < \frac{1}{w_J} + \frac{N}{w_I+\lambda},$$

where the last inequality follows from (2.10.29). Since $\lambda > -\frac{w_I}{2}$ we have $\frac{N}{w_I+\lambda} < \frac{2N}{w_I}$ and

$$\lambda < \left(\frac{1}{w_J} + \frac{2N}{w_I}\right)^{-1} = \frac{w_I w_J}{w_I + 2Nw_J}.$$

Thus we get the following expression for $\overline{\delta}$

$$\overline{\delta} = \frac{\varkappa}{Nw_J}\left(w_I + \frac{w_I w_J}{w_I + 2Nw_J}\right) + \phi. \tag{2.10.71}$$

The $\overline{\delta}/\sqrt{\tau_J}$ is given by $\frac{\varkappa}{Nw_J\sqrt{\tau_J}}\left(w_I + \frac{w_I w_J}{w_I+2Nw_J}\right) + \frac{\rho}{\sqrt{1-\rho^2}}\frac{1}{\sqrt{\tau_s}}$. Clearly, it is decreasing in τ_J. □

2.10.4 Proof of Proposition 11

The equilibrium is a solution to the system (2.10.20-2.10.23), which can be written as follows

$$\lambda = L(\delta; N) \equiv \frac{N w_J}{\varkappa} (\delta - \phi) \left(\theta + \delta^2\right) - w_I, \text{ and} \tag{2.10.72}$$

$$\delta = D(\lambda; N) \equiv h\left(\frac{\lambda\,(w_I + N w_J + \lambda)}{w_J\,(w_I + 2\lambda)}\right), \tag{2.10.73}$$

where $h(x)$ is the inverse of $1 + \delta\,(\delta - \phi)$, i.e. it solves

$$x = 1 + h(x)\,(h(x) - \phi)\,.$$

Lemma 16 implies that in equilibrium $\lambda > 0$. This is not possible if $\delta < \phi$, therefore we may look for the nitersection of the two curves in the region $\delta > \phi$ and $\lambda > 0$.

Since for $\delta > \phi$ the function $1 + \delta\,(\delta - \phi)$ is strictly increasing, the function $h(x)$ is well-defined and is strictly increasing as well.

The equilibrium is therefore the intersection of the two curves, $\lambda = L(\delta; N)$ and $\delta = D(\lambda; N)$, moreover it is easy to see that $\frac{\partial L}{\partial \delta} > 0$ and $\frac{\partial D}{\partial \lambda} > 0$ for $\delta > \phi$, so both curves are strictly upward-sloping for a given N. We next compute

$$\frac{\partial L}{\partial N} = \frac{w_J\,(\delta^2 + \theta)\,(\delta - \phi)}{\varkappa} - w_I'(N),$$

which is positive both if w_I does not depend on N, and if $w_I = w_1 N$.

Analogously, we compute

$$\frac{\partial D}{\partial N} = h'(\cdot) \times \begin{cases} \frac{\lambda}{2\lambda + w_I} & , \text{ if } w_I \text{ does not depend on } N \\ \frac{\lambda^2 (w_1 + 2w_J)}{w_J (2\lambda + N w_1)^2} & , \text{ if } w_I = w_1 N. \end{cases}$$

The above is positive.

Therefore an infinitesimal increase in N shifts the curve $L(\delta; N)$ up and the curve $D(\lambda; N)$ to the right. Their new intersection will be below and to the left from the old one[51]. Thus, we have

$$\frac{d\lambda}{dN} < 0, \ \frac{d\delta}{dN} < 0.$$

Since $\mathcal{I} = \frac{\tau_J + \tau_s(1+\delta^2)}{\tau_J}$ is increasing in δ and does not directly depend on N, and $\mathcal{L}$ is inversely related to λ, we have

$$\frac{d\mathcal{I}}{dN} < 0 \text{ and } \frac{d\mathcal{L}}{dN} > 0.$$

Lemma 16. *[If $\rho \geq 0$, the equilibrium price impact is positive.*

Proof. Rewrite (2.10.23) as follows

$$\lambda = \frac{w_J\left(1+\delta\left(\delta-\phi\right)\right)\left(w_I+2\lambda\right)}{\left(w_I+Nw_J+\lambda\right)}.$$

The $\delta > \phi$, because otherwise $\lambda < -w_I$ and the second order condition $2\lambda + w_I > 0$ does not hold. Therefore $1 + \delta\left(\delta - \phi\right) > 0$. Other terms in the above are positive due to the second order condition $w_I + 2\lambda > 0$. □

2.10.5 Proof of Proposition 12

I will use the fixed point condition $l = \Lambda(l)$ to analyze the comparative statics of price impact.

Recall that

$$1/\Lambda(l;\tau_s) \equiv (N-1)\,g(l) + g_J(l;\tau_s).$$

Denote

$$\begin{aligned} z(l;\tau_s) &\equiv \frac{1}{\tau_s}t(l;\tau_s)\left(t(l;\tau_s) - \frac{B}{C}\right) \\ &= \frac{1}{\tau_s}t(l;\tau_s)\frac{b_J(l;\tau_s)}{Ng(l)C}. \end{aligned}$$

Recall that $C = \sqrt{\frac{1-\rho^2}{\tau_I}}$ does not depend on τ_s. With this notation one can write

$$1/\Lambda(l;\tau_s) = \frac{g(l)Nw_J + 1}{w_J z(l;\tau_s) + w_J} - g(l). \tag{2.10.74}$$

I am interested in $\frac{\partial\Lambda}{\partial\tau_s}$. The only term which depend on τ_s directly is $z(l;\tau_s)$.

[51] It can be shown that the curve $\lambda = L(\delta; N)$ has to intersect the curve $\delta = D(\lambda; N)$ from below, since for $\lambda = 0$ the curve $\lambda = L(\delta; N)$ is to the right of the curve $\delta = D(\lambda; N)$

Substituing (2.9.3) and differentiating implicitly (2.9.2) one can find

$$\frac{\partial}{\partial \tau_s}\left(\frac{t(l;\tau_s)b_J(l;\tau_s)}{\tau_s}\right) =$$

$$= \frac{-b_J\left(g^2N^2b_J\left(B^2+C^2(\tau_s-\tau_J)\right)+2BNgb_J^2+BC^2g^3N^3\tau_s+b_J^3\right)}{CNg\tau_s^2\left(g^2N^2\left(B^2+C^2(\tau_J+\tau_s)\right)+4BNgb_J+3b_J^2\right)}.$$

The above is negative provided that $B^2+C^2(\tau_s-\tau_J)>0$, which is equivalent to

$$\tau_s > \frac{1-2\rho^2}{1-\rho^2}\tau_J.$$

Therefore, provided that the above holds, the nominator (denominator) of (2.10.74) is increasing (decreasing) in τ_s, which implies that (the two are positive since $\Lambda(\lambda)$ is positive) $\frac{\partial \Lambda}{\partial \tau_s} < 0$. Therefore an increase in τ_s shifts the function $\Lambda(l)$ up, and its' new intersection with a 45 degree line will shift up as well ($\Lambda(l)$ intersects the 45 degree line from above, see Lemma 17).

Lemma 17. *In the unique equilibrium* $\Lambda'(\lambda) < 1$.

Proof. I derive the sign of $\Lambda'(\lambda)$ from the comparative statics with respect to τ_I.

We first prove that $\frac{\partial \lambda}{\partial \tau_I} > 0$ in the unique equilibrium. Indeed changes of τ_I has no effect on the curve (2.10.73) but shifts down the curve (2.10.72). Their intersection occurs at a point with greater λ.

Second, we prove that $\frac{\partial \Lambda(\lambda;\tau_I)}{\partial \tau_I} > 0$. Indeed, observe that in (2.10.74) only the term z depend on τ_I through $t(l;\tau_I)b_J(l;\tau_I)$. Moreover, it is easy to prove that $t(l;\tau_I)b_J(l;\tau_I)$ is decreasing in τ_I. Thereofre the denominator of (2.10.74) is decreasing in τ_I which implies that $\frac{\partial \Lambda}{\partial \tau_I} > 0$. Therefore an increase in τ_I shifts the function $\Lambda(l)$ up. Since $\frac{\partial \lambda}{\partial \tau_I} > 0$, the intersection of $\Lambda(l)$ and the 45 degree line should be from above, which proves the Lemma. □

2.10.6 Proof of Propositions 13 and 14

Make a change of variable

$$m \equiv \frac{\lambda + w_I}{Nw_J} = \frac{\left(\delta^2+\theta\right)\left(\delta-\phi\right)}{\varkappa}.$$

With this change of variable (2.10.24) becomes

$$(m+1)\frac{m-\psi}{2m-\psi}=\frac{\delta(\delta-\phi)+1}{N}.$$

The left hand side of the above is increasing in δ and is decreasing in ψ. The right hand side is increasing in δ and does not depend on ψ. An infinitisemal increase in w_I (or a decrease in w_J) shifts the LHS down and the new intersection of LHS and RHS occurs at a point with greater δ and thus greater $\mathcal{I}$.

Since m is increasing in δ we have that $\lambda + w_I$ is increasing in w_I. The $\Gamma = \frac{\frac{1}{w_J}+\frac{N}{w_I+\lambda}}{1+\delta(\delta-\phi)}$ is therefore decreasing in w_I.

2.10.7 Proof of Proposition 15

Given Lemma 18 scaling the measure of J traders by a factor $\mu = M/N$ is equivalent to scaling w_J by a factor $\frac{1}{\mu} = N/M$. Therefore λ and δ in the economy 1 solve (cf. (2.10.72-2.10.73))

$$\lambda = L(\delta) \equiv \frac{Nw_J}{\varkappa}(\delta-\phi)\left(\theta+\delta^2\right) - w_I, \text{ and} \tag{2.10.75}$$

$$\delta = D(\lambda;\mu) \equiv h\left(\mu\frac{\lambda\left(w_I+Nw_J+\lambda\right)}{w_J\left(w_I+2\lambda\right)}\right). \tag{2.10.76}$$

The first curve is unaffected by changes in μ. Decreasing μ shifts the curve to the left and the new point of intersection has lower δ and greater λ. Therefore, the liquidity decreases, whereas the information efficiency increases.

Lemma 18. *Consider an economy with N I-traders and a measure μ of J-traders with utility $u_J = (v_J - p)\,x - \frac{\hat{w}_J x^2}{2}$. Call this economy $\hat{\mathcal{E}}$. Any equilibrium in this economy is an equilibrium in the economy with N I-traders and a unit measure of J-traders with utility $u_J = (v_J - p)\,x - \frac{w_J x^2}{2}$,where $w_J = \frac{\hat{w}_J}{\mu}$. Call this economy $\mathcal{E}$. Conversely, any equilibrium in $\mathcal{E}$ is also an equilibrium in $\hat{\mathcal{E}}$.*

Proof. Guess that the equilibrium demands in the economy $\hat{\mathcal{E}}$ are given by

$$x_i = \alpha + \beta\cdot v_I - \gamma\cdot p \text{ and } x_j = a_J + b_J\cdot s_j - g_J\cdot p.$$

Denote $\alpha_J = \mu a_J$,$\beta_J = \mu b_J$ and $\gamma_J = \mu g_J$. Following the steps of the proof of

Theorem 1 one can see that α, β, γ,α_J, β_J and γ_J satisfy the same equations as the corresponding soefficients in the economy $\mathcal{E}$. $\square$

2.10.8 Proof of Proposition 16

Define

$$\mathcal{U}_J^{ep} \equiv E\left[(v_J - p)\, x_j(p) - \frac{w_J x_j(p)^2}{2} \middle| s_j, p \right]$$

the ex-post utility of a J-trader. Using the notation introduced in the Proof of the Lemma 13 the above can be written as

$$\mathcal{U}_J^{ep} = (u_j - p)\, x_j(p) - \frac{w_J x_j(p)^2}{2}.$$

Compute

$$\begin{aligned} x_j(p) &= \frac{u_j - p}{w_J} \\ &= \frac{u_J - p}{w_J} + \beta_J \epsilon_j \\ &= -X + \beta_J \epsilon_j, \end{aligned}$$

where the second line substitutes $u_j = u_J + w_J \beta_J \epsilon_j$ and the third line uses the market cleraing conditon $X + \int_0^1 x_j(p)dj = 0$. The expected utility of a J-trader can now be written as

$$\begin{aligned} \mathcal{U}_J &= E\left[\mathcal{U}_J^{ep}\right] \\ &= E\left[(u_J - p + w_J \beta_J \epsilon_j)(-X + \beta_J \epsilon_j) - \frac{w_J (-X + \beta_J \epsilon_j)^2}{2} \right] \\ &= E\left[-(u_J - p)\, X - \frac{w_J X^2}{2} \right] + \frac{w_J \beta_J^2}{2\tau_s}, \end{aligned}$$

where the third line computes the expectation with respect to ϵ_j. Substituting $X = G(v_I - u_J)$ and

$$\begin{aligned} u_J - p &= \frac{N w_J (u_J - v_I)}{N w_J + w_I} \\ &\equiv K_J (u_J - v_I) \end{aligned}$$

the above becomes

$$\begin{aligned}\mathcal{U}_J &= E\left[GK_J\left(u_J - v_I\right)^2 - \frac{w_J G^2\left(u_J - v_I\right)^2}{2}\right] + \frac{w_J\beta_J^2}{2\tau_s} \\ &\equiv H_J E\left[\left(u_J - v_I\right)^2\right] + \frac{w_J\beta_J^2}{2\tau_s},\end{aligned}$$

where I have denoted

$$H_J \equiv G\left(K_J - \frac{w_J G}{2}\right). \tag{2.10.77}$$

As in the Proof of the Lemma 13 one can write

$$v_I - u_J = \overline{v}_I - \overline{v}_J + c_v(v_J - \overline{v}_J) + c_\epsilon \epsilon_\pi,$$

which allows to calculate

$$\mathcal{U}_J = H_J\left(\left(\overline{v}_I - \overline{v}_J\right)^2 + c_v^2/\tau_J + c_\epsilon^2\right) + \frac{w_J\beta_J^2}{2\tau_s}.$$

Lemma 19 implies that H_J is increasing in N. From Proposition 11 we know that δ is decreasing in N, hence β_J is increasing in N. From the proof of Lemma 13 we know that if (2.10.60) holds, c_v^2 and c_ϵ^2 are both decreasing in δ and hence increasing in N. Therefore, the if (2.10.60) holds, $\mathcal{U}_J$ is increasing in N.

The expected utility of an I-trader can be calculated as

$$\mathcal{U}_I = \frac{1}{N}E\left[\left(v_I - p\right)X - \frac{w_I X^2}{2N}\right].$$

Substituting $X = G(v_I - u_J)$ and

$$v_I - p = \left(1 - K_J\right)\left(v_I - u_J\right)$$

the expected utility of an I-trader becomes

$$\mathcal{U}_I = H_I E\left[\left(u_J - v_I\right)^2\right],$$

where

$$H_I \equiv \frac{1}{N}G\left(\left(1 - K_J\right) - \frac{w_I}{2N}G\right).$$

Computing the expectation we get

$$\mathcal{U}_I = H_I \left((\overline{v}_I - \bar{v}_J)^2 + c_v^2/\tau_J + c_\epsilon^2 \right).$$

We finally compute the welfare

$$\begin{aligned}
\mathcal{W} &= N\mathcal{U}_I + \mathcal{U}_J \\
&= H \left((\overline{v}_I - \bar{v}_J)^2 + c_v^2/\tau_J + c_\epsilon^2 \right) + \frac{w_J \beta_J^2}{2\tau_s}.
\end{aligned}$$

Where I denoted

$$H \equiv NH_I + H_J = G - \frac{1}{2}\left(\frac{w_I}{N} + w_J\right) G^2. \tag{2.10.78}$$

From Proposition 11 we know that δ is decreasing in N, hence β_J is increasing in N. From Lemma 19 we know that H is increasing in N. Finally, from Lemma 13 we know that if (2.10.60) holds, c_v^2 and c_ϵ^2 are both decreasing in δ and hence increasing in N. Therefore, the if (2.10.60) holds, the welfare is increasing in N.

Lemma 19. *The H_J and H given by (2.10.77) and (2.10.78) is increasing in N.*

Proof. Write

$$H(G, N) = G - \frac{1}{2}\left(\frac{w_I}{N} + w_J\right) G^2.$$

Compute

$$\frac{dH}{dN} = \frac{\partial H}{\partial G}\frac{dG}{dN} + \frac{1}{2}\frac{w_I}{N^2} G^2.$$

Compute

$$\frac{\partial H}{\partial G} = 1 - \frac{\frac{w_I}{N} + w_J}{1/G} = 1 - \frac{\frac{w_I}{N} + w_J}{\frac{w_I}{N} + w_J + \frac{\lambda}{N}} > 0.$$

From

$$G = \frac{1}{\frac{w_I}{N} + w_J + \frac{\lambda}{N}}$$

it is clear that G is increasing in N (recall that according to Proposition 11 λ is decreasing in N).

Analogously,

$$\frac{dH_J}{dN} = \frac{\partial H_J}{\partial G}\underbrace{\frac{dG}{dN}}_{>0} + \frac{\partial H_J}{\partial K_J}\underbrace{\frac{dK_J}{dN}}_{>0}.$$

Compute

$$\frac{\partial H_J}{\partial G} = K_J - w_J G = \frac{\lambda N w_J}{(Nw_J + w_I)(\lambda + Nw_J + w_I)} > 0$$

and

$$\frac{\partial H_J}{\partial K_J} = G > 0.$$

Thus, $\frac{dH_J}{dN} > 0$. $\square$

2.10.9 Proof of Theorem 4

The values of price impact l such that $2l + w_I > 0$ does not hold are ruled out by the Lemma 3.

For the if part, we should show that if $l = \Lambda(l)$ then there exists an equilibrium such that l is equal to the slope of the inverse residual supply. A natural candidate is an equilibrium with $\beta = g(l)$, $\gamma = g(l)$ and $\beta_J = b_J(l)$, $\gamma_J = g_J(l)$, where the function $g(l)$ is given by (2.9.1), $b_J(l)$ is given by (2.9.2) and $g_J(l)$ is given by (2.9.5).

We shall show that if $l = \Lambda(l)$ then the above strategies constitute mutual best responses. By definition $\Lambda(l) = \frac{1}{(N-1)g(l)+g_J(l)}$ therefore by the Lemma 20 $\beta = g(l)$ and $\gamma = g(l)$ are the best responses to other I-traders playing $\beta = g(l)$ and $\gamma = g(l)$ and J-traders playing $\beta_J = b_J(l)$ and $\gamma_J = g_J(l)$.

By consrtruction $\beta_J = b_J(l)$ and $\gamma_J = g_J(l)$ are the best responses to I-traders playing $\beta = g(l)$ and $\gamma = g(l)$ and J-traders playing $\beta_J = b_J(l)$ and $\gamma_J = g_J(l)$, which proves the if part.

For the only if part we shall show that if there is an equilibrium with a price impact l such that $2l + w_I > 0$, then $l = \Lambda(l)$ should hold.

From Lemma 20 we know that in this equilibrium I-traders play $\beta = g(l)$ and $\gamma = g(l)$. Since it is an equilibrium, the strategies of J-traders should constitute mutual best responses and be a best response to $\beta = g(l)$ and $\gamma = g(l)$ played by the I-traders. By definition of $b_J(l)$ and $g_J(l)$ the J-traders should play $\beta_J = b_J(l)$ and $\gamma_J = g_J(l)$

Since the price impact l is given by $\frac{1}{(N-1)\gamma+\gamma_J}$ and we have $\gamma = g(l)$ and $\gamma_J = g_J(l)$, we get that $l = \Lambda(l)$, which proves the only if part.

Lemma 20. *The strategy $x_i(p)$ of a trader $i \in I$ is a best response to the profile of symmetric linear strategies of traders $j \in J$ and $k \in I$, $k \neq i$ characterised by the* *coefficients* $(\alpha_J, \beta_J, \gamma_J)$ *and* (α, β, γ) *such that* $\frac{2}{(N-1)\gamma+\gamma_J} + w_I > 0$ *if and only if it is given by*

$$x_i(p) = \frac{v_I - p}{w_I + l}, \text{ where } l = \frac{1}{(N-1)\,\gamma + \gamma_J}.$$

Proof. As it is discussed in the proof of the Theorem 3 the best response of a trader i solves

$$\max_{(x,p)} (v_I - p)\,x - \frac{w_I}{2}x^2 \tag{2.10.79}$$

$$\text{s.t.: } p = \iota + l \cdot x. \tag{2.10.80}$$

Taking the first order condition and eliminating ι using (2.10.80) yields the following expression for the best response of the trader i

$$x_i = \frac{1}{w_I + l}(v_I - p). \tag{2.10.81}$$

By assumption the second order condition $w_I + 2\lambda > 0$ holds, therefore the above first order necessary condition is also sufficient. $\square$

2.10.10 Stability. The expression for $\Lambda'(l)$.

First recall that $B = \rho\sqrt{\frac{\tau_J}{\tau_I}}$, $C = \sqrt{\frac{1-\rho^2}{\tau_I}}$ implying that

$$B = \xi,\; C = \frac{1}{\sqrt{\tau_s}\varkappa},$$

in the notation adopted in the appendix.

We first differentiate implicitly (2.9.2) to get the expression for $b_J'(l)$:

$$b_J'(l) = -\frac{2\varkappa^2 b_J(l)^2(b_J(l)(l+w_I)+N\xi)}{\varkappa^2 b_J(l)(l+w_I)(3b_J(l)(l+w_I)+4N\xi)+N^2\left(\theta+\varkappa^2\xi^2\right)} < 0. \tag{2.10.82}$$

Differentiation of (2.9.6) yields

$$\Lambda'(l) = \frac{\text{nominator}}{\text{denominator}},$$

where

$$\text{denominator} = \frac{1}{w_J}\left(N^2(l+(N-1)w_J+w_I) - \varkappa^2 w_J b_J(l)(l+w_I)(b_J(l)(l+w_I)+N\xi)\right)^2,$$

$$
\begin{aligned}
\text{nominator} \quad &= \quad \varkappa^2 N^2 b_J(l)^2 (l+w_I)^2 \left(2l + w_J \left(-\varkappa^2 \xi^2 + 3N - 2\right) + 2w_I\right) + (N-1)N^4 w_J - \\
&- \quad 2\varkappa^4 N \xi w_J b_J(l)^3 (l+w_I)^3 - \varkappa^4 w_J b_J(l)^4 (l+w_I)^4 + \\
&+ \quad \varkappa^2 N^3 \xi (l+w_I)^2 b_J'(l)(l + N w_J + w_I) + \\
&+ \quad b_J(l) \left(2\varkappa^2 N^2 (l+w_I)^3 b_J'(l)(l + N w_J + w_I) + \varkappa^2 \xi N^3 (l+w_I)(l + 2(N-1)w_J + \right.
\end{aligned}
$$

Substituting (2.10.82) to the above two expressions yields a closed form solution for $\Lambda'(l)$, up to a solution of a cubic equation determining $b_J(l)$:

$$
1 = w_J b_J \left(\theta + \left(\phi + \kappa\left(w_I + l\right)\frac{b_J}{N}\right)^2\right).
$$

Chapter 3

A Model of OTC Market with Multilateral Trading Platform

3.1 Introduction

Trade in over-the-counter (OTC) markets was traditionally bilateral: a customer calls a dealer and negotiates terms of trade. Recently, in some OTC markets customers may also trade using multilateral trading platforms (MTPs) allowing them to query several dealers and to trade with the one offering the best quote (see Figure 3.0.1 for an illustration). For example, more than 10% of trades in $8tn corporate bond market has been completed via MTP (Hendershott and Madhavan (2015)). While the theory of bilateral trade in OTC markets is well-established (e.g., Duffie, Garleanu and Pedersen (2005) and the literature that followed), to the best of our knowledge the the theory of multilateral trade in OTC market (in particular via MTPs) is absent. In this paper we aim to fill this gap.

The trade via MTP is different from the traditional bilateral trade along the following two dimensions. Matching: customers are matched to several dealers, instead of just one. Price mechanism: price is determined in a first-price auction, not through bargaining. We augment the classic OTC framework of Duffie, Garleanu and Pedersen (2005) along those two dimensions. In particular, we consider an economy with a continuum of risk-neutral infinitely-lived investors who can hold up to one unit of asset. Investors are of two types: high-type investors receive higher flow utility from holding the asset. The types evolve stochastically and the differences in types generate trade. The trade is subject to a search friction: investors wishing to

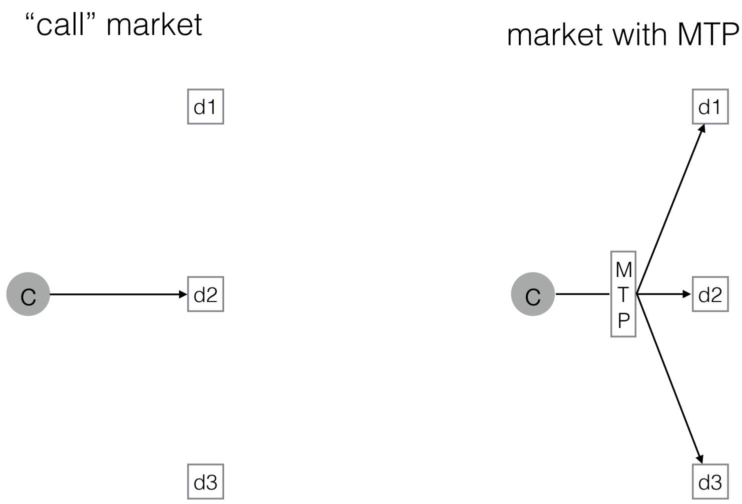

Figure 3.0.1: In a traditional, "call" market a customer (c) contacts one dealer (d) and negotiates the terms of trade. MTP allows to contact several dealers and to trade with the one offering best quote.

trade cannot do so instantaneously. We allow investors searching for quotes (playing a role of customers) to query several (n) other investors (playing a role of dealers) and to run the first-price auction among them. See Figure 3.2.1 for an illustration. We study both stationary equilibrium and transition dynamics, and the model is tractable in both.

In our model all dealers responded to customer's query have the same valuations, yet bid and ask distribution is non-degenerate. The dispersion of bid and ask prices has a strategic nature: each dealer is uncertain about how many other dealers respond and uses a mixed strategy when providing a quote. We provide testable implications linking skewness and dispersion of bid and ask prices to dealers' response rate. In particular, we show that ask prices are negatively skewed, while bid prices are positively skewed. We also show that dispersion of bid and ask prices of assets (or during periods of time) with higher response rate of dealers is smaller. The intuition is simple: higher response rate implies that the market power of each seller is smaller. The prices would be more concentrated near the competitive price and the dispersion will be smaller. The skewness of bid and ask prices is of opposite signs, because the ask distribution is shifted to the right, while the bid distribution is shifted left due to market power, producing negative (positive) skewness.

We show that increasing n may provoke a liquidity squeeze: when n is large, it is hard to find dealers as they are taken from the market quickly. As a result small changes in asset supply can cause a large swings in price. However, such a squeeze is actually efficient: for example, when asset supply is low, efficiency implies that all asset should be allocated to high-types, implying that the measure of low-types who are willing to sell the asset is zero. Moreover, we show that increasing n improves welfare both in steady state and in transition. Moreover, in the limit as $n \to \infty$ efficient allocation can be achieved in steady state even when the trade is infrequent. However, even when n is infinite, the transition to efficient allocation takes time because of search friction. We also show that allowing large n may provoke a liquidity squeeze: when n is large, it is hard to find dealers as they are taken from the market quickly. As a result small changes in asset supply can cause a large swings in price.

Our paper is related to two strands of literature: search models of OTC markets and auctions with uncertain number of bidders. Our main difference to the first strand of literature (e.g. Duffie, Garleanu and Pedersen (2005), Weill (2007), Vayanos and Weill (2008), Lagos and Rochateau (2009)) is that the trade in our model is multilateral. Price competition among several dealers and uncertainty about the number of them generates price dispersion in our model, that is strategic in nature. The latter is in contrast to price dispersion in Hugonnier et al (2015) and Shen et al (2015), where price dispersion is a consequence of dispersion in investors' valuations.

Relative to the auctions literature with uncertain number of bidders (see e.g. Klemperer (1999) for a review) and IO literature on pricing with heterogeneously informed consumers (e.g. Butters (1997), Varian (1980) and Burdett and Judd (1983)) our main difference is the endogeneity of this uncertainty. In our model uncertainty about the number of dealers who respond is a consequence of heterogeneity of traders: some of them may not respond, when there are no gains from trade. The measure of different kinds of agents is determined endogenously in our model, which is important for the liquidity squeeze result.

3.2 The model

The time is continuous and goes from zero to infinity, $t \in \mathcal{T} \equiv [0, \infty)$. There is a unit continuum of risk-neutral infinitely lived investors who discount future consumption at a rate ρ: given a consumption stream $\{c_t\}_{t\in\mathcal{T}}$ each gets utility

$$\int_0^\infty c_t e^{-\rho t} dt.$$

There is one asset and investors may choose to hold $x_t \in \{0,1\}$ units of it.[1] The supply of the asset is $s \in [0,1]$. Investors are of two types, *high* and *low*: $\theta_t \in \{1, 1-\delta\}$. Investor of type θ_t derives flow utility θ_t if holds the asset. Types evolve according to a continuous time Markov chain:

$$\begin{aligned}
\Pr\left(\theta_{t+dt} = 1 | \theta_t = 1-\delta\right) &= \lambda_u dt, \\
\Pr\left(\theta_{t+dt} = 1-\delta | \theta_t = 1\right) &= \lambda_d dt.
\end{aligned}$$

Depending on the type θ_t and asset holdings x_t we get four *kinds* of investors, $k \in \{b, s, h, o\}$: buyers, sellers, holders and outsiders. We denote the corresponding measures and values by μ_k and V_k, respectively. We summarize the notation in the table below.

kind	x_t	θ_t	Measure	Value
*b*uyer	0	1	μ_b	V_b
*s*eller	1	$1-\delta$	μ_s	V_s
*h*older	1	1	μ_h	V_h
*o*utsider	0	$1-\delta$	μ_o	V_o

Difference in types generates trade in our model. As in DGP (2005) the δ may represent, for example, hedging reasons to sell or relative tax disadvantage. The trade is subject to a friction. In order to purchase the asset buyers have to find sellers and vice versa. Investors can only search at arrival times of Poisson process with intensity α. One possible interpretation of such search friction is the following: $1/\alpha$ is a mean time to finalize investment decision, e.g. to get approval from risk-management desk.[2] When an investor can search, he simultaneously contacts n other investors from the whole population at random and asks them for quotes. An investor searching for a quote plays a role of a *customer* and investors receiving his

[1] We can extend the model similarly to DGP (2005) and add the second asset, money market account, earning the interest r. This will not change any of the results presented in the paper, so we decided to have just one asset to simplify the exposition.

[2] In such an interpretation: $\alpha dt = \Pr(\text{intended trade is approved in } dt)$.

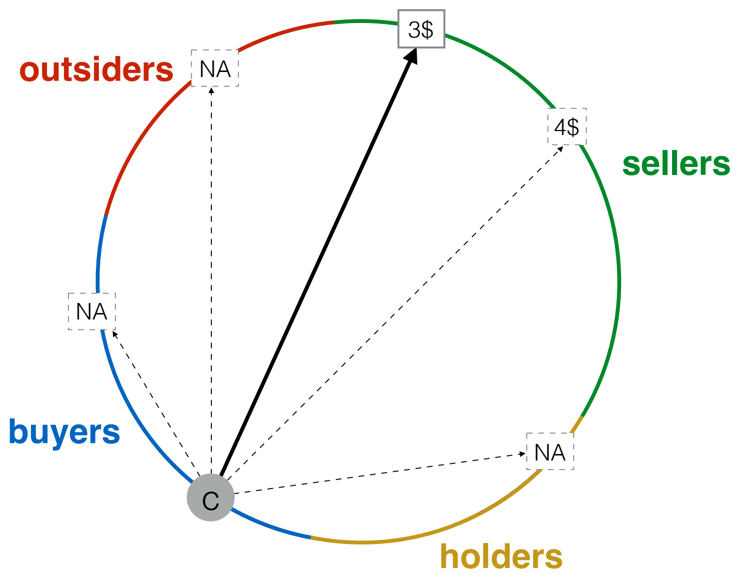

Figure 3.2.1: A buyer requests 5 other investors for a quote. A buyer plays a role of a customer (depicted as a circle), other investors (depicted as rectangles) act as dealers. Only sellers respond with a quote. The buyer trades with the seller providing best quote (3$).

query play a role of *dealers.* The quotes are take-it-or-leave-it offers. Investor trades with the one who provided best quote if there are gains from trade and does not trade otherwise.[3] We assume that only the dealers with gains from trade respond with a quote (i.e. only sellers (buyers) respond to a query for a bid (ask) quote). See Figure 3.2.1 for an illustration.

3.3 Stationary equilibrium

In this section we derive the stationary equilibrium in our model. The equilibrium objects are: 1) value functions V_k 2) measures μ_k, $k \in \{b, s, h, o\}$ and 3) price strategies. In equilibrium only sellers would be willing to sell and only buyers will be willing to buy. Therefore, without loss of generality we assume that sellers and

[3]A real life example of such a mechanism is a MarketAxess electronic trading platform for trading bonds. Here is the description from Hendershott and Madhavan (2015):

> MarketAxess is an electronic trading platform with access to many dealers in U.S. investment-grade and high-yield corporate bonds, Eurobonds, emerging markets, credit default swaps, and U.S. agency securities. MarketAxess allows an investor to query multiple dealers electronically, providing considerable time savings relative to the alternative of a sequence of bilateral negotiations with this same set of dealers. An ending time is specified for the auction. Auctions vary in length from 5 to 20 minutes, and only at the end of the auction does the investor review the dealer responses and select the best quote...MarketAxess charges dealers a fee between 0.1 and 0.5 basis points for investment-grade bonds.

only sellers provide bid prices and buyers and only buyers provide ask prices and no quotes are provided in all other cases.[4] Asking and bidding strategies of, respectively, sellers and buyers are characterized by the CDFs $A(p) \equiv Pr(\text{quoted ask price} < p)$ and $B(p) \equiv Pr(\text{quoted bid price} < p)$.

We focus on symmetric stationary Markov perfect equilibria. We derive equilibrium in three steps. First, taking values V_k and measures μ_k as given we derive the price strategies $A(p)$ and $B(p)$ in section 3.3.1. We then derive equilibrium measures μ_k and values V_k in sections 3.3.2 and 3.3.3, respectively.

3.3.1 Price strategies

For a trade to happen the price should satisfy $V_s - V_o \leq p \leq V_h - V_b$. The first inequality is to ensure that sellers are willing to sell; the second inequality ensures that the buyers are willing to buy. We denote the *reservation price of buyers (sellers)* by r_b (r_s). We denote the difference between the two, which we call *gains from trade*, by Δ.

$$r_s \equiv V_s - V_o,\ r_b \equiv V_h - V_b,\ \Delta \equiv r_b - r_s.$$

We guess, and later verify that r_s, r_b and Δ are positive. We also introduce the following notation

$$\gamma_s \equiv (1-\mu_s)^{n-1},\ \gamma_b \equiv (1-\mu_b)^{n-1}.$$

We call γ_s (γ_b) *market power* of sellers (buyers): it is equal to probability of being a monopolist conditional on being contacted.[5] The Proposition below characterizes the equilibrium price strategies, taking the measures and values as given.

Proposition 17. *There are no pure strategy equilibria. The equilibrium price strategies are given by CDFs*

$$A(p) = \frac{1}{\mu_s} - \frac{1-\mu_s}{\mu_s}\left(\frac{\Delta}{p-r_s}\right)^{\frac{1}{n-1}},\ B(p) = \frac{1-\mu_b}{\mu_b}\left(\left(\frac{\Delta}{r_b-p}\right)^{\frac{1}{n-1}} - 1\right),$$

[4] So for example if a buyer calls to holder, outsider or another buyer and asks for the price at which they sell, they provide no quotes.

[5] Indeed, consider a seller contacted by some buyer. The seller is a monopolist, i.e. the only seller among n traders contacted by the buyer, if each of the $n-1$ remaining contacts led the buyer not to a seller. The probability of not reaching a seller is $1-\mu_s$. The probability of not reaching a seller $n-1$ times is $(1-\mu_s)^{n-1}$.

with support $[r_s + \gamma_s\Delta, r_b]$ *and* $[r_s, r_b - \gamma_b\Delta]$*, respectively. The expected profits of buyers and sellers are given by, respectively*

$$\pi_s = \gamma_s\Delta,\ \pi_b = \gamma_b\Delta. \tag{3.3.1}$$

Proof. We consider sellers' strategies here. The claims for buyers are proved analogously in the appendix.

(*No pure strategy equilibrium*). Suppose all sellers charge the price p^*. Undercutting this price by a small amount is a profitable deviation, except for the case when $p^* = r_s$. However, charging $p^* = r_s$ is not an equilibrium either, as it yields a profit of zero, whereas deviating to any price $p > r_s$ yields strictly positive profit: if only the seller of interest responded (which happens with strictly positive probability) the quote p will be accepted and yield $p - r_s > 0$ for that seller.

(*Derivation of the strategy* $A(p)$). Given the above we will be looking at mixed strategies. Denote the equilibrium strategy of a seller by $A(p) = Pr(\text{quoted bid price} < p)$ and its' support by $[\overline{p_a}, \underline{p_a}]$. We show in the appendix that the price strategy has connected support and has no point masses. Consider a particular seller and suppose he quotes a price p. He gets a profit

$$\begin{aligned}
\pi_s &= Pr(\text{qoute } p \text{ is the best})(p - r_s) \\
&= (1 - \mu_s + \mu_s(1 - A(p)))^{n-1}(p - r_s) \\
&= (1 - \mu_s A(p))^{n-1}(p - r_s) \\
&= \text{const} \\
&= (1 - \mu_s)^{n-1}(\overline{p_a} - r_s).
\end{aligned}$$

The first equality is clear: a buyer will only trade with a seller at price p if the it is the best quote among the ones the buyer received. If the buyer trades with the seller of interest, the seller gets $p - r_s$. Quote p is the best if any of the other $n-1$ traders queried by the buyer either did not respond (happens with probability $1 - \mu_s$) or respond, but with a quote greater than p (happens with probability $\mu_s(1 - B(p))$). This explains the second line. In the third line we simply rearrange. For traders to play mixed strategy they have to be indifferent between any price in the support (otherwise they would charge the price yielding the greatest profit with probability 1). Therefore the profit should be constant (fourth line), in particular we can evaluate it at $p = \overline{p_a}$ (fifth line). Combining third and fifth lines we express

$$A(p) = \frac{1}{\mu_s} - \frac{1-\mu_s}{\mu_s}\left(\frac{\overline{p_a} - r_s}{p - r_s}\right)^{\frac{1}{n-1}}.$$

In the appendix we show that $\overline{p}_a = r_b$, which pins down the pricing strategy and expected profit. □

Intuitively, sellers playing a role of a dealer use mixed strategies because depending on how many other dealers respond they may either be monopolists, in which case they would like to charge the price r_b. They may also face competition with other sellers, in which case they would have to charge the price closer to r_s. As a result they mix between the prices from a segment $[r_s + c, r_b]$, where $c$is some positive number. Any price from the segment should yield the same profit π_s. In particular when charging the lowest price $r_s + c$ a seller ensures that his quote is always accepted and gets a profit $r_s + c - r_s = \pi_s$. It then follows that $c = \pi_s$, i.e. the distance c between the lowest price charged and the reservation price r_s is the profit of sellers. When charging the highest price sellers earn the maximal possible profit Δ in case their quote is accepted. The latter happens with probability γ_s. It then follows that $\pi_s = \gamma_s\Delta$, i.e. the market power γ_s is also equal to the fraction of the trade surplus that a seller gets. It is therefore analogous to bargaining power in models with Nash bargaining. Note, however, that in our model, unlike Nash bargaining models, γ_s is endogenous.

In short, the price dispersion arises due to uncertainty about the number of competitors that each dealer has. The latter uncertainty arises due to two factors which we find realistic. First, the seller does not know the current type of his potential competitors: if it is high, they will not be willing to sell and therefore will not contribute to competition. Second, he does not know their current inventories: even if their type is low they might not have the asset to sell, therefore not contributing to competition. Hendershott and Madhavan (2015) document a significant dispersion in the fraction of dealers responded, providing the empirical support to the fact that each dealer is uncertain about the number of competitors.

3.3.2 Demographics

In this section we find measures of traders μ_k, $k \in \{b, s, h, o\}$ in a stationary equilibrium. We express the measures μ_b, μ_h and μ_o through the measure of sellers. First, in a stationary equilibrium the measure of investors with $\theta = 1$ is[6]

$$\eta \equiv \frac{\lambda_u}{\lambda_u + \lambda_d}.$$

Hence,

$$\mu_b + \mu_h = \eta, \tag{3.3.2}$$

$$\mu_s + \mu_o = 1 - \eta. \tag{3.3.3}$$

From market clearing condition we also have

$$\mu_h + \mu_s = s. \tag{3.3.4}$$

Equations (3.3.2-3.3.4) allow to express

$$\mu_h = s - \mu_s,\ \mu_b = \mu_s + \eta - s,\ \mu_o = 1 - \mu_s - \eta. \tag{3.3.5}$$

It remains to pin down the measure of sellers. We consider the inflows and outflows from the population of sellers. In a short period dt a measure $\mu_s \alpha dt$ of sellers will search, of which only a fraction $1 - (1 - \mu_b)^n$ will contact at least one buyer and trade. Hence, there is an outflow $\nu_s dt$ due to seller-initiated transactions, where

$$\nu_s \equiv \left(1 - (1 - \mu_b)^n\right) \mu_s \alpha. \tag{3.3.6}$$

Since one unit of asset is traded in each transaction, ν_s is also a *seller-initiated trading volume.* Analogously, there is an outflow $\nu_b dt$ due to buyer-initiated transactions, where

$$\nu_b \equiv \left(1 - (1 - \mu_s)^n\right) \mu_b \alpha. \tag{3.3.7}$$

Again, ν_b is also a *buyer-initiated trading volume.* The outflows from the population of sellers in a short period dt include: (i) a measure$\mu_s \lambda_u dt$, who will change to holders

[6]Indeed, in a stationary equilibrium, the measure of agents who change their type from $\theta = 1$ to $\theta = 1 - \delta$ is $\eta \lambda_d$ per unit of time. It should be equal to the measure of agents who change their type from $\theta = 1 - \delta$ to $\theta = 1$, which is $(1 - \eta)\lambda_u$, per unit of time. From $\eta \lambda_d = (1 - \eta)\lambda_u$ one finds $\eta = \frac{\lambda_u}{\lambda_u + \lambda_d}$.

due to type-switching shocks, (ii) a measure $(\nu_s + \nu_b)dt$ who will become outsiders due to trade. The inflows are due to holders of a measure $\mu_h \lambda_d dt$ becoming sellers. Equating net inflow of sellers to zero one gets

$$\mu_s \lambda_u + \nu_s + \nu_b - \mu_h \lambda_d = 0. \tag{3.3.8}$$

One pins down μ substituting (3.3.5) in the last equation. The Proposition below summarizes the above discussion.

Proposition 18. *The measures μ_k, $k \in \{b, s, h, o\}$ are given by (3.3.5), where μ_s is the unique solution $\mu \in (0, s)$ to*

$$\mu\lambda_u + (1 - (1 - \mu - \eta + s)^n)\,\mu\alpha + (1 - (1 - \mu)^n)\,(\mu + \eta - s)\,\alpha - (s - \mu)\,\lambda_d = 0. \tag{3.3.9}$$

Proof. It remains to prove that there exists a unique solution to (3.3.9). Since the left hand side (LHS) of (3.3.9) is strictly increasing in μ, there is at most one solution. The LHS is continuous, strictly negative at $\mu = 0$ and strictly positive at $\mu = s$, therefore a unique solution $\mu \in (0, 1 - \eta)$ exists. □

3.3.3 Value functions

In this section we determine the value functions V_k, $k \in \{b, s, h, o\}$ in a stationary equilibrium. We start with holders. The Hamilton-Jacobi-Bellman equation writes as:

$$0 = 1 + \lambda_d(V_s - V_h) - \rho V_h. \tag{3.3.10}$$

The interpretation is as follows. Over a short period dt a holder gets a flow utility $1dt$ from holding the asset, plus, with probability $\lambda_d dt$ he can switch his type and become a seller, in which case his value changes by $V_s - V_h$, minus the "depreciation" of $\rho V_h dt$ due to discounting. For outsiders one can similarly write

$$0 = \lambda_u(V_b - V_o) - \rho V_o. \tag{3.3.11}$$

We now consider sellers. Denoting $\langle b \rangle (\langle a \rangle)$ the average bid (ask) transaction price (i.e., the average price paid in seller- (buyer-) initiated transactions), the HJB equation writes as

$$0 = 1 - \delta + \lambda_u(V_h - V_s) + \frac{\nu_s}{\mu_s}\left(\langle b\rangle - r_s\right) + \frac{\nu_b}{\mu_s}\left(\langle a\rangle - r_s\right) - \rho V_s. \tag{3.3.12}$$

In a short period dt the value of a seller changes by $(1-\delta)\, dt$ due to holding the asset. It may also change by $V_h - V_s$ with probability $\lambda_u dt$ due to a type type switching shock. A change of $\langle b\rangle - r_s$ is possible if a seller of interest initiates a transaction. Since $\nu_s dt$ sellers initiate a trade in a period dt and there are μ_s sellers in total, the probability that a seller of interest initiates a trade is $\frac{\nu_s}{\mu_s}dt$. Similarly, the value changes by $\langle a\rangle - r_s$ if a seller of interest trades in a transaction initiated by a buyer. The probability of the latter is $\frac{\nu_b dt}{\mu_s}$. Finally, the value also changes by $-\rho V_s dt$ due to discounting.

For buyers one can similarly write

$$0 = \lambda_d(V_o - V_b) + \frac{\nu_b}{\mu_b}\left(r_b - \langle a\rangle\right) + \frac{\nu_s}{\mu_b}\left(r_b - \langle b\rangle\right) - \rho V_b. \tag{3.3.13}$$

We finally determine the average bid and ask prices $\langle b\rangle$ and $\langle a\rangle$. We write the expected profit of a seller in a short period dt due to a trade initiated by a buyer in two ways. The first one was already explored when writing the HJB equation and gives $\frac{\nu_b}{\mu_s}\left(\langle a\rangle - r_s\right) dt$, . The second gives $n\alpha\mu_b\pi_s$: each seller may be contacted by a buyer with probability $n\alpha\mu_b dt$ and gets profit π_s if contacted. Hence,

$$\langle a\rangle = r_s + \frac{n\alpha\mu_b\mu_s}{\nu_b}\pi_s. \tag{3.3.14}$$

Similarly, writing the expected profit of a buyer due to a trade initiated by a seller yields

$$\langle b\rangle = r_b - \frac{n\alpha\mu_b\mu_s}{\nu_s}\pi_b. \tag{3.3.15}$$

Substituting (3.3.14-3.3.15) and (3.3.1) into (3.3.12-3.3.13) yields

$$0 = 1 - \delta + \lambda_u(V_h - V_s) + \Delta\left(\frac{\nu_s}{\mu_s} + n\alpha\mu_b\left(\gamma_s - \gamma_b\right)\right) - \rho V_s, \tag{3.3.16}$$

$$0 = \lambda_d(V_o - V_b) + \Delta\left(\frac{\nu_b}{\mu_b} + n\alpha\mu_s\left(\gamma_b - \gamma_s\right)\right) - \rho V_b. \tag{3.3.17}$$

Equations (3.3.10), (3.3.11), (3.3.16), (3.3.17) constitute a system of 4 linear equations allowing do determine all value functions. The Proposition below summarizes the derivation of equilibrium.

Proposition 19. *There exists a unique equilibrium. The price strategies and measures of agents are given by Propositions 17 and 18. The value functions are unique solutions to a linear system of equations (3.3.10), (3.3.11), (3.3.16), (3.3.17). The expressions for gains from trade and reservation prices of buyers and sellers are given by*

$$\Delta = \frac{\delta}{\rho + \lambda_u + \lambda_d + \frac{\nu_b}{\mu_b} + \frac{\nu_s}{\mu_s} + n\alpha(\mu_s - \mu_b)\left(\gamma_b - \gamma_s\right)}, \tag{3.3.18}$$

$$r_b = \frac{1}{\rho} - \frac{\Delta}{\rho}\left(\lambda_d + \frac{\nu_b}{\mu_b} + n\alpha\mu_s\left(\gamma_b - \gamma_s\right)\right), \tag{3.3.19}$$

$$r_s = \frac{1-\delta}{\rho} + \frac{\Delta}{\rho}\left(\lambda_u + \frac{\nu_s}{\mu_s} + n\alpha\mu_b\left(\gamma_s - \gamma_b\right)\right). \tag{3.3.20}$$

Proof. We derive expressions for r_b, r_s, and Δ. Subtracting (3.3.11) from (3.3.16) and (3.3.13) from (3.3.10) one gets (3.3.20) and (3.3.19). Subtracting (3.3.19) from (3.3.20) one gets (3.3.18). We uniquely express $V_h - V_s$ and $V_b - V_o$ through Δ in the appendix, from which it follows that the system (3.3.10), (3.3.11), (3.3.16), (3.3.17) has a unique solution. We also prove in the appendix that r_s, r_b and Δ are positive. □

3.3.4 Transaction price distribution

We first characterize the equilibrium transaction price distribution. The distribution of transaction ask prices (i.e., prices paid in buyer-initiated transactions) is different from the distribution of the quoted ask prices $A(p)$ (as buyers select best quotes) and is denoted by $F_a(p)$. The distribution of transaction bid prices (i.e., prices paid in seller-initiated transactions) is denoted by $F_b(p)$. The distribution of prices in all transactions is denoted by $F(p)$. The means of $F_a(p)$, $F_b(p)$ and $F(p)$ are denoted by $\langle a\rangle$, $\langle b\rangle$ and $\langle p\rangle$:

$$\langle a\rangle \equiv \int p dF_a(p),\ \langle b\rangle \equiv \int p dF_b(p),\ \langle p\rangle \equiv \int p dF(p).$$

We refer to $\langle b\rangle$ and $\langle a\rangle$ as *mean transaction bid (ask) price*. We call $\langle p\rangle$ *mean transaction price*. The standard deviations are denoted by σ_b and σ_s:

$$\sigma_a^2 \equiv \int (p - \langle p_a\rangle)^2 \, dF_a(p), \; \sigma_b^2 \equiv \int (p - \langle p_b\rangle)^2 \, dF_b(p).$$

We further normalize the above measures by the corresponding ranges of bid and ask prices: $(1-\gamma_s)\,\Delta$ in case of ask prices and $(1-\gamma_b)\,\Delta$ in case of bid prices. We call $\sigma_a/\left((1-\gamma_s)\,\Delta\right)$ *ask dispersion* and $\sigma_b/\left((1-\gamma_b)\,\Delta\right)$ *bid dispersion*. The normalization makes the measures of dispersion unitless and also scales the measures to an interval between 0 and 1. As we will also see from the proposition below the scaled dispersion depends on just two parameters: number of traders queried (n) and the response rate (μ_s in case of buyer-initiated transactions and μ_b in case of seller initiated transactions), both of which are observable in the data (Hendershott and Madhavan (2015)).

Our measure of skewness is *nonparametric skew*, which by definition is the difference between the mean and median, scaled by the standard deviation:

$$\text{skew}_{\text{ask}} = \frac{\langle p_b\rangle - F_a^{-1}(1/2)}{\sigma_a}; \; \text{skew}_{\text{bid}} = \frac{\langle p_s\rangle - F_b^{-1}(1/2)}{\sigma_b}.$$

Normalization by standard deviation ensures that the measures of skewness are unitless and is between -1 and 1. We call skew_{bid} (skew_{ask}) *bid (ask) skewness.*

We characterize the distributions and their moments in the proposition below. response

Proposition 20. *The distributions $F_a(p)$, $F_b(p)$ and $F(p)$ are given by*

$$F_a(p) = \frac{\alpha\mu_b}{\nu_b}\left(1 - \left(\frac{\gamma_s\Delta}{p - r_s}\right)^{\frac{n}{n-1}}\right) \text{ with support } [r_s + \gamma_s\Delta, r_b], \tag{3.3.21}$$

$$F_b(p) = 1 - \frac{\alpha\mu_s}{\nu_s}\left(1 - \left(\frac{\gamma_b\Delta}{r_b - p}\right)^{\frac{n}{n-1}}\right) \text{ with support } \left[r_s, \, r_b - \gamma_b\Delta\right], \tag{3.3.22}$$

$$F(p) = \frac{\nu_s F_s(p) + \nu_b F_b(p)}{\nu_s + \nu_b}. \tag{3.3.23}$$

Prices $\langle a\rangle$ and $\langle b\rangle$ are given by (3.3.14-3.3.15). The mean transaction price of $\langle p\rangle$ is given by

$$\langle p\rangle = \frac{\nu_s\,\langle b\rangle + \nu_b\,\langle a\rangle}{\nu_s + \nu_b}.$$

The bid and ask dispersions are given by

$$\sigma_b / \left((1 - \gamma_s) \Delta \right) = \phi(\mu_b, n), \; \sigma_a / \left((1 - \gamma_b) \Delta \right) = \phi(\mu_s, n), \tag{3.3.24}$$

where

$$\phi(\mu, n) \equiv \frac{n\mu(1-\mu)^{n-1}}{(1-(1-\mu)^n)(1-(1-\mu)^{n-1})} \sqrt{\frac{1-(1-\mu)^{n-2}}{(n-2)(1-\mu)^{n-2}} \frac{1-(1-\mu)^n}{n\mu^2} - 1}. \tag{3.3.25}$$

The bid and ask skewness is given by

$$\text{skew}_{\text{bid}} = -\psi(\mu_b, n), \; \text{skew}_{\text{ask}} = \psi(\mu_s, n),$$

where

$$\psi(\mu, n) = \frac{1 - \left(1 - \frac{1}{2}(1-(1-\mu)^n)\right)^{\frac{1-n}{n}} \frac{1-(1-\mu)^n}{n\mu}}{\sqrt{\frac{1-(1-\mu)^{n-2}}{(n-2)(1-\mu)^{n-2}} \frac{1-(1-\mu)^n}{n\mu^2} - 1}} > 0.$$

Proof. See Appendix. □

The proposition above yields several implications.

3.3.5 Transaction price skewness and dispersion and dealer response rate

Below we relate the properties of transaction bid and ask price distributions to the observable characteristics such as dealer response rate and number of dealers queried.

1. *The transaction ask prices are positively skewed while bid prices are negatively skewed.* The intuition is simple: due to market power sellers charge prices concentrated near monopoly price r_b. The distribution of bid prices is thus shifted to the right producing negative skewness. Similar intuition applies for bid prices that are concentrated near r_s.

2. *Price dispersion is lower for assets with higher response rate.* It follows from the fact that the function $\phi(\mu, n)$ is decreasing in μ. Hendershott and Madhavan (2015) report that traders query between 24 and 28 dealers in the corporate bond market and that this number is similar for investment-grade and high-yield bonds. They also report that the response rate is higher for investment-

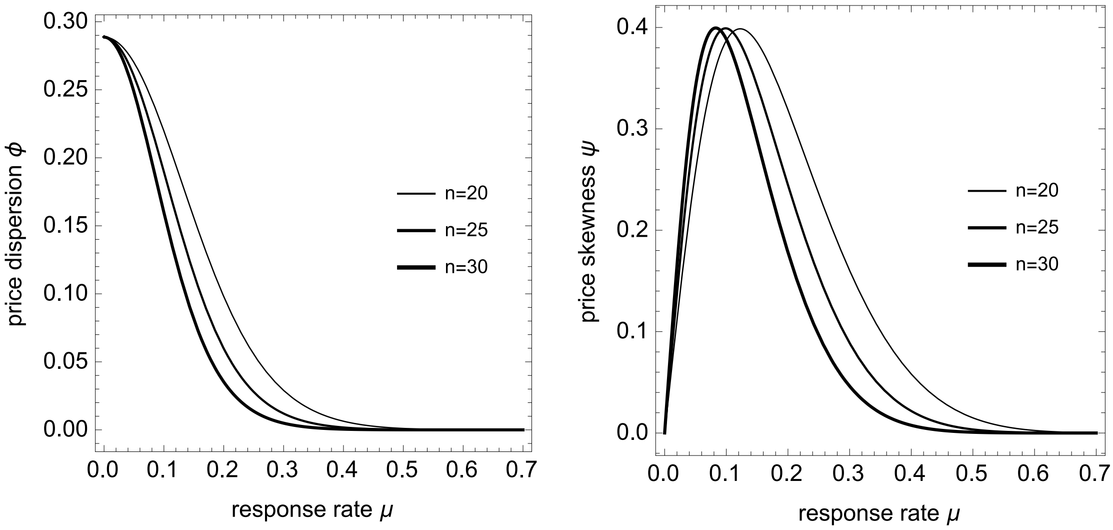

Figure 3.3.1: Price dispersion and skewness.

grade bonds. It then follows that there should be more dispersion in prices of high-yield bonds. The left panel of the Figure 3.3.1 plots the price dispersion as a function of response rate.

3. *The absolute value of skewness is the highest for assets with intermediary response rate.* As we discussed above, the source of skewness in transaction prices is a market power of dealers. High response rate implies low market power and hence low skewness. On the other hand if the response rate is small, the market power is high, and the price distribution is concentrated near the monopoly price. In the case of bid distribution the left tail will be thin, producing small negative skewness. Similar intuition applies for ask prices. Consequently, the largest value of skewness should be observed for intermediary values of response rates. The left panel of the Figure 3.3.1 plots the skewness as a function of response rate.

4. *The distribution of scaled bid (ask) prices first-order stochastically dominates that of assets with lower (higher) response rates.* We define scaled bid price as bid minus the minimal bid normalized by the range of bid prices: $\frac{p-(r_s+\gamma_s\Delta)}{(1-\gamma_s)\Delta}$. Clearly, scaled price is always between zero and one. Similarly, scaled ask price is $\frac{p-r_s}{(1-\gamma_b)\Delta}$. The intuition for stochastic dominance result is simple: higher response rate in seller-initiated transactions implies lower market power of buyers and hence higher ask prices. Similarly, higher response rate in buyer-initiated transactions implies lower market power of sellers and hence lower bid prices.

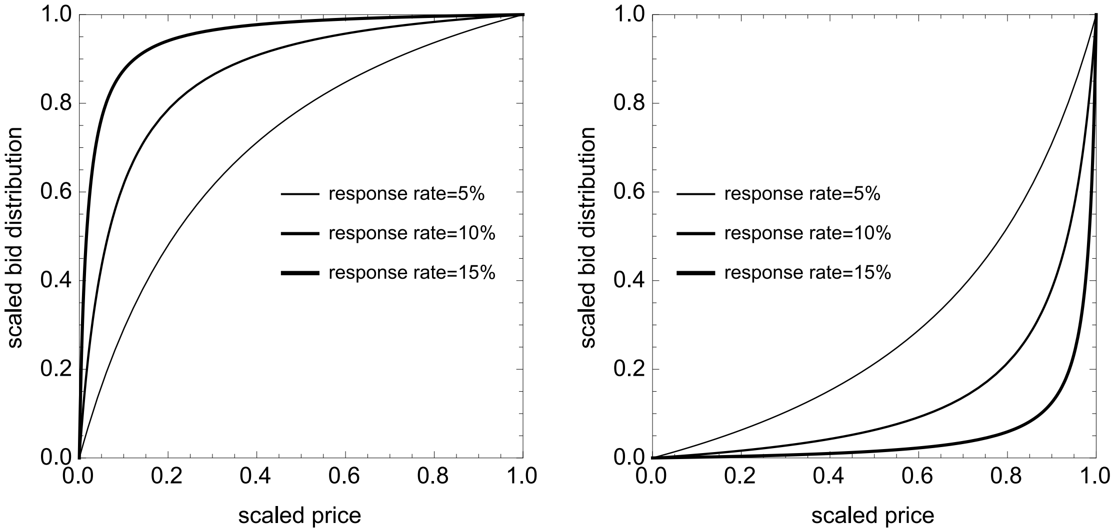

Figure 3.3.2: Distributions of scaled bid and ask prices, $n = 26$.

Given the results reported in Hendershott and Madhavan (2015) the above implies that the distribution of scaled ask (bid) prices for investment-grade (high-yield) bonds should first-order stochastically dominate that of high-yield (investment-grade) bonds. Figure 3.3.2 plots the distributions of scaled bid and ask prices for different values of response rates.

3.3.6 Price impact

In this section we consider how the market adsorbs supply shocks. Our measure of price impact is $\Lambda(s)$ which is defined as follows

$$\Lambda(s) \equiv \left| \frac{\partial \langle p \rangle (s)}{\partial s} \right|.$$

Our objective is to examine how changes in the number of dealers queried (in particular, going from a "call" market ($n = 1$) to an auction market ($n \geq 2$)) affects the above aspect of market liquidity. The Figure 3.3.3 demonstrates that as n grows large the market may become fragile: for some values of s the price impact becomes large. Moreover, Proposition 21 derives the values of s for which the price impact is infinite in the $n \to \infty$ limit and also derives the limiting function $\langle p \rangle (s)$.

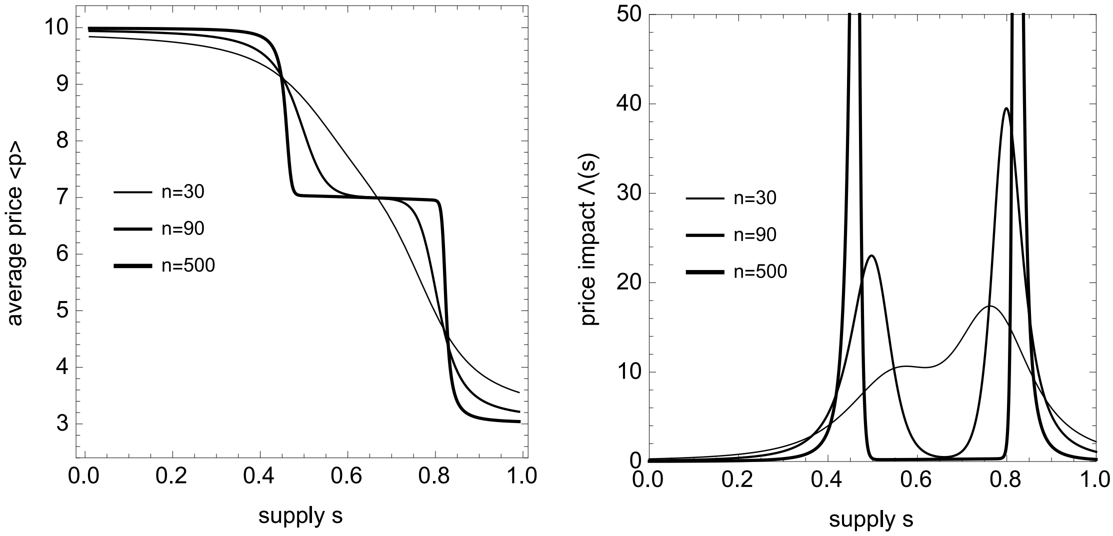

Figure 3.3.3: Average price and price impact. Parameter values: $\rho = 0.1$, $\alpha = \lambda_u = 2$, $\lambda_d = 1$, $\delta = 0.7$.

Proposition 21. *The function $\langle p \rangle (s)$ in the $n \to \infty$ limit is given by*

$$\begin{cases} \frac{1}{\rho} & \text{, } \textit{if } s \leq \underline{s}; \\ w(s)\left(\frac{1}{\rho} - \frac{\delta}{\rho}\frac{\lambda_d+\alpha}{\rho+\lambda_u+\lambda_d+2\alpha}\right) + (1-w(s))\left(\frac{1-\delta}{\rho} + \frac{\delta}{\rho}\frac{\lambda_u+\alpha}{\rho+\lambda_u+\lambda_d+2\alpha}\right) & \text{, } \textit{if } \underline{s} < s < \overline{s}; \\ \frac{1-\delta}{\rho} & \text{, } \textit{if } s \geq \overline{s}, \end{cases}$$

where

$$\underline{s} \equiv \frac{\alpha}{\alpha+\lambda_d}\eta,\ \overline{s} \equiv \left(1 + \frac{\lambda_d}{\alpha+\lambda_u}\right)\eta,$$

$$w(s) \equiv \frac{s(\alpha+\lambda_d) - \alpha\eta}{s(\lambda_d - \lambda_u) + \eta(\lambda_d+\lambda_u)} \in [0,1].$$

Consequently, $\Lambda(s)$ become infinite when $s \in \{\underline{s}, \overline{s}\}$ as $n \to \infty$.

Proof. See Appendix. □

The force behind the sharp changes in price observed in the Figure 3.3.3 is a liquidity squeeze: as n grows, the measure of sellers becomes zero for $s < \underline{s} = 0.4(4)$ (short squeeze), similarly the measure of buyers becomes zero for $s > \overline{s} = 0.8(3)$ (long squeeze). Consequently, when the supply decreases from $\underline{s} + \epsilon$ to $\overline{s} - \epsilon$ there is a short squeeze (sellers become scarce) and the price jumps. When the supply increases from $\overline{s} - \epsilon$ to $\overline{s} + \epsilon$ there is a long squeeze (buyers become scarce) and the price falls.

The intuition behind the particular value of $\underline{s}$ is as follows: $\underline{s}$ is the value of supply such that all new sellers coming to the market per unit of time (the measure of them is $\lambda_d \mu_h = \lambda_d (s - \mu_s) = \lambda_d s$) are taken from the market by searching buyers (the measure is $\alpha \mu_b = \alpha(\mu_s + \eta - s) = \alpha(\eta - s)$). Equating the two measures we find

$$\lambda_d s = \alpha(\eta - s) \Rightarrow s = \underline{s}.$$

The intuition behind the value of $\overline{s}$ is similar. Measure of new buyers coming to the market per unit of time, $\mu_o \lambda_u = (1 - s)\lambda_u$ should be equal to the measure of sellers searching $\alpha \mu_s = \alpha(s - \eta)$. Hence,

$$(1 - s)\lambda_u = \alpha(s - \eta) \Rightarrow s = \overline{s}.$$

To illustrate the mechanism further we compute the average time spent on the market by buyers and sellers. The latter depends on the price that they charge. We compute the average time for a trader charging monopoly price: r_b for sellers and r_s for buyers. The average time is given by

$$\tau_b = \left(\lambda_d + \frac{\nu_b}{\mu_b} + n\alpha\mu_s\gamma_b \right)^{-1}, \; \tau_s = \left(\lambda_u + \frac{\nu_s}{\mu_s} + n\alpha\mu_b\gamma_s \right)^{-1}.$$

The buyer may become an outsider (with intensity λ_d) or a holder due to trade that he initiates himself (intensity $\frac{\nu_b}{\mu_b}$) or due to a trade initiated by a seller (with intensity $n\alpha\mu_s\gamma_b$).[7] A similar explanation applies for a formula for τ_s. The time τ_b measure the scarcity of buyers: when τ_b is small it is hard to find a buyer, as they go from the market very quickly. Similarly, τ_s measures the scarcity of sellers. Figure 3.3.4 provides an illustration.

3.3.7 Efficiency and welfare

We start by investigating the efficient allocation. The efficient allocation is the one in which all asset is allocated to high types. Once high types are allocated the asset, the remaining is given to the low types. Thus, if $s < \eta$ efficiency implies

$$\mu_s = 0, \; \mu_h = s, \; \mu_b = \eta - s, \; \mu_o = 1 - \eta.$$

[7]In a period dt a buyer gets contacted with probability $n\alpha\mu_s dt$. Given that he charges a monopoly price a seller trades with hime with probability γ_b.

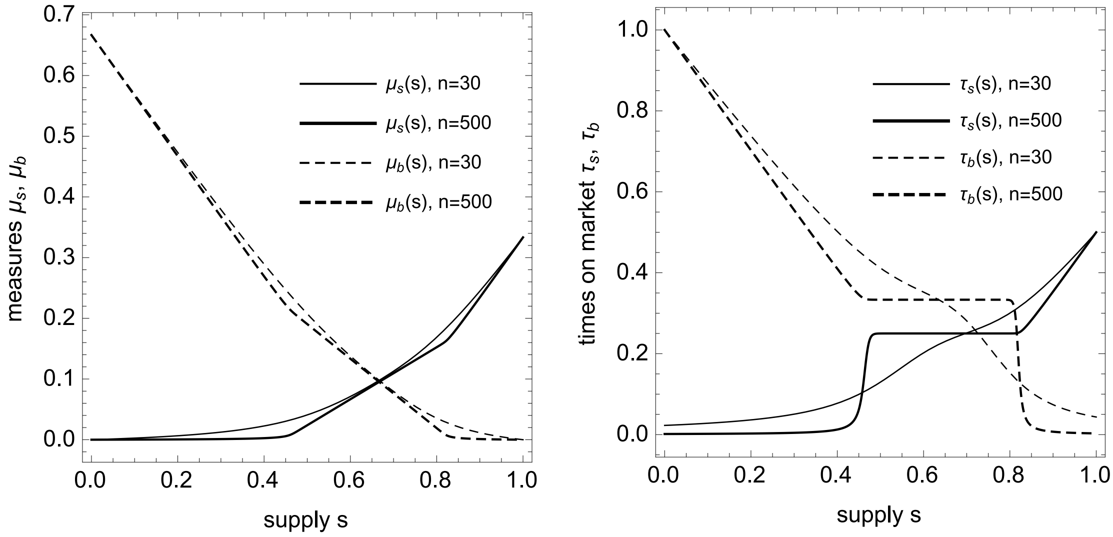

Figure 3.3.4: Measures and average times spent on the market of buyers and sellers. Parameter values: $\rho = 0.1$, $\alpha = \lambda_u = 2$, $\lambda_d = 1$, $\delta = 0.7$.

Similarly, if $s \geq \eta$ efficiency implies

$$\mu_s = s - \eta,\ \mu_h = \eta,\ \mu_b = 0,\ \mu_o = 1 - s.$$

The welfare is easy to calculate: at any time t the total utility flow is $\mu_s(1-\delta)+\mu_h$. Hence, substituting $\mu_h = s - \mu_s$ the welfare can be written as

$$\mathcal{W} = \frac{s - \mu_s \delta}{\rho}.$$

Proposition 22. *The efficient allocation is: $\mu_s = \mu_s^e \equiv [s-\eta]^+$, and the rest of the measures expressed by (3.3.5). The welfare in the efficient allocation is given by $\mathcal{W} = \frac{s-\mu_s^e\delta}{\rho}$. For finite α, as $n \to \infty$, the allocation converges to the efficient one, if $s \in [0, \underline{s}] \cup [\overline{s}, 1]$. For finite n, as $\alpha \to \infty$, the allocation converges to the efficient one for any s.*

Proof. See Appendix. □

Why the efficient allocation could not be obtained in the $n \to \infty$ limit for intermediary values of supply ($s \in [\underline{s}, \overline{s}]$)? The intuition is as follows. Because of the search friction, the agents are stuck with inefficient allocation. For example, for $s < \eta$ efficiency implies that $\mu_s = 0$. Suppose that indeed $\mu_s = 0$ and $\underline{s} < s < \eta$. Then the measure of new sellers coming to the market is $\lambda_d s$ is greater than the measure of the buyers searching (i.e. the maximal measure of sellers that transform

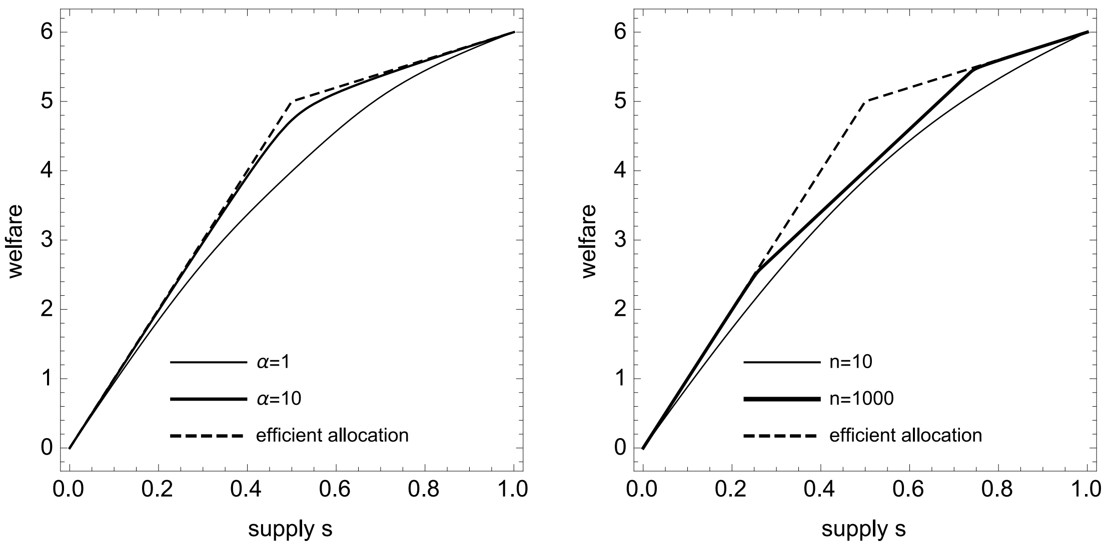

Figure 3.3.5: Approaching efficiency by changing α and n. Parameter values: $\rho = 0.1$, $\lambda_u = \lambda_d = 1$, $\delta = 0.8$, $\alpha = 1$ (right panel) and $n = 30$ (left panel).

into outsiders due to trade) $\alpha(\eta - s)$. Hence, the next instant the measure of sellers become positive. Figure 3.3.5 provides an illustration.

3.4 Transition dynamics

In this section we analyze the symmetric Markov perfect equilibrium without assuming stationarity. The equilibrium objects $V_k(t)$, $\mu_k(t)$, $A(p;t)$ and $B(p;t)$, $k \in \{b, s, h, o\}$ now depend on time. We drop the argument t where it does not cause a confusion. As before, we first characterize the price strategies taking values and measures as given. Since we did not use the stationarity when deriving the price strategies in the section 3.3.1, our derivation holds even in the non-stationary case. Therefore, it remains to derive the evolution of measures μ_k and values V_k.

As application of the techniques developed in this section we consider the dynamics of shock adsorption in this market, that the economy starts in a stationary state corresponding to the supply s, at time 0 the supply changes to s'. We then study how the economy reaches the new steady-state.

3.4.1 Demographics.

We first derive the evolution of measure $\eta(t)$ of high-type traders. The measure of agents who change their type from $\theta = 1$ to $\theta = 1 - \delta$ is $\eta\lambda_d$ per unit of time, whereas the measure of agents who change their type from $\theta = 1 - \delta$ to $\theta = 1$ is $(1-\eta)\lambda_u$, per unit of time, therefore the evolution of η is given by the following ODE:[8]

$$\dot{\eta} = -\eta\lambda_d + (1-\eta)\lambda_u.$$

Denoting the measure of high types at time 0 by η^0 and the stationary measure by η^* we get the following solution to the above ODE

$$\eta(t) = \eta^* + (\eta^0 - \eta^*)\exp(-(\lambda_u + \lambda_d)t), \tag{3.4.1}$$

$$\text{where } \eta^* \equiv \frac{\lambda_u}{\lambda_u + \lambda_d}.$$

The measures of buyers, holders and outsiders can be expressed through η and μ_s through equations (3.3.5). Equating the net inflow of sellers per unit of time to $\dot{\mu}_s$ yields the following ODE:

$$\dot{\mu}_s = \mu_h\lambda_d - \mu_s\lambda_u - \nu_s - \nu_b, \tag{3.4.2}$$

Or, substituting (3.3.5) and (3.3.6-3.3.7):

$$\dot{\mu}_s = m(\mu_s, t), \tag{3.4.3}$$

where

$$m(\mu_s, t) \equiv (s - \mu_s)\,\lambda_d - \mu_s\lambda_u -$$

$$-\left(1 - (1 - \mu_s - \eta(t) + s)^n\right)\mu_s\alpha - \left(1 - (1-\mu_s)^n\right)(\mu_s + \eta(t) - s)\,\alpha. \tag{3.4.4}$$

The initial state of the economy can be characterized by initial measure of high types (η^0) and the initial measure of sellers (μ_s^0): the remaining measures can be expressed through (3.3.5). We summarize the derivation of measures μ_k in the Proposition below.

Proposition 23. *The measures $\mu_k(t)$, $k \in \{b, s, h, o\}$ are given by (3.3.5), where $\eta(t)$ is given by (3.4.1) and $\mu_s(t)$ is the unique solution to the ODE (3.4.3) subject to the boundary condition $\mu_s(0) = \mu_s^0$.*

The ODE (3.4.3) above is a non-linear first-order ODE. Suppose we have an integral curve, i.e. a solution $\mu_s(t)$ to the ODE above, which we represent on a (μ_s, t)

[8]We denote the derivative with respect to time by dot, e. g., $\frac{d}{dt}x = \dot{x}$.

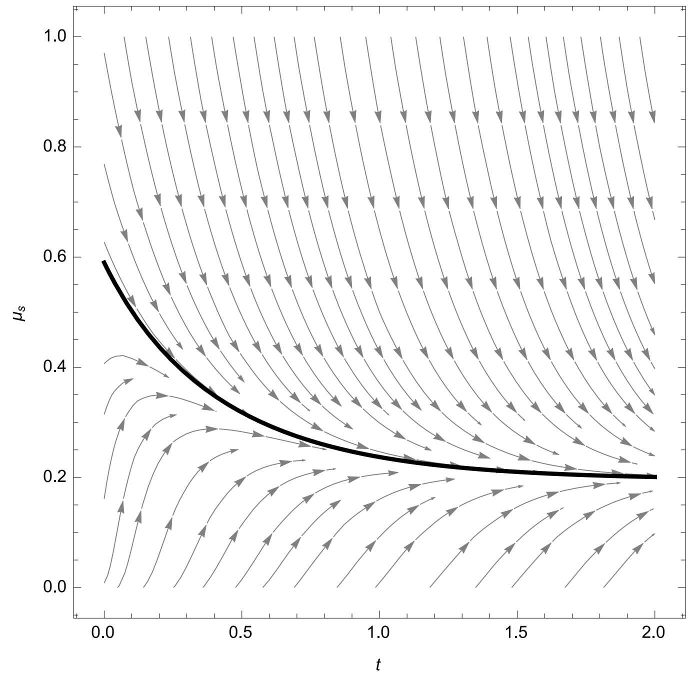

Figure 3.4.1: Slope field for the ODE (3.4.3) and the particular solution corresponding to initial condition $\mu_s^0 = 0.59$. The remaining parameter values are $\alpha = \lambda_u = \lambda_d = 1$, $n = 10$, $s = 0.6$, $\eta^0 = 0.6$.

diagram. A vector with coordinates $(1, m(\mu_s, t))$ is tangent to the integral curve at the point (μ_s, t). A collection of such vectors is called a *slope field,* corresponding to the ODE. We can characterize the slope field in close form: the function $m(\mu_s, t)$ is given by (3.4.4). Figure 3.4.1 plots the slope field for particular parameter values. It also plots the solution to (3.4.3) which is easy to obtain using standard numerical methods.

3.4.2 Value functions.

In the previous sections we derived the HJB equations by equating the change in the value of a trader, $V_t - V_{t+dt}$, to zero. In non-stationary equilibrium the latter change is equal $-\dot{V}dt$. Therefore, the HJB equations become

$$0 = \dot{V}_h + 1 + \lambda_d(V_s - V_h) - \rho V_h, \tag{3.4.5}$$

$$0 = \dot{V}_o + \lambda_u(V_b - V_o) - \rho V_o, \tag{3.4.6}$$

$$0 = \dot{V}_s + 1 - \delta + \lambda_u(V_h - V_s) + \Delta\left(\frac{\nu_s}{\mu_s} + n\alpha\mu_b\left(\gamma_s - \gamma_b\right)\right) - \rho V_s, \tag{3.4.7}$$

$$0 = \dot{V}_b + \lambda_d(V_o - V_b) + \Delta\left(\frac{\nu_b}{\mu_b} + n\alpha\mu_s\left(\gamma_b - \gamma_s\right)\right) - \rho V_b. \tag{3.4.8}$$

Subtracting (3.4.8) from (3.4.5) and (3.4.6) from (3.4.7) yields

$$0 = \dot{r}_b + 1 - \Delta\left(\lambda_d + \frac{\nu_b}{\mu_b} + n\alpha\mu_s\left(\gamma_b - \gamma_s\right)\right) - \rho r_b,$$

$$0 = \dot{r}_s + 1 - \delta + \Delta\left(\lambda_u + \frac{\nu_s}{\mu_s} + n\alpha\mu_b\left(\gamma_s - \gamma_b\right)\right) - \rho r_s.$$

Taking the difference of the above two expressions one gets the ODE for Δ

$$0 = \dot{\Delta} + \delta - \Delta\left(\rho + \lambda_d + \lambda_u + \frac{\nu_b}{\mu_b} + \frac{\nu_s}{\mu_s} + n\alpha\left(\mu_s - \mu_b\right)\left(\gamma_b - \gamma_s\right)\right). \tag{3.4.9}$$

Imposing a transversality condition

$$\lim_{t\to\infty} \exp(-\rho t)\Delta(t) = 0$$

One gets the following unique solution to (3.4.9):

$$\Delta(t) = \delta \int_t^{\infty} \exp(-R(t,\tau))d\tau, \tag{3.4.10}$$

where

$$R(t,\tau) \equiv$$
$$\int_t^{\tau}\left(\rho + \lambda_d + \lambda_u + \frac{\nu_b(z)}{\mu_b(z)} + \frac{\nu_s(z)}{\mu_s(z)} + n\alpha\left(\mu_s(z) - \mu_b(z)\right)\left(\gamma_b(z) - \gamma_s(z)\right)\right) dz. \tag{3.4.11}$$

Having the solution for $\Delta(t)$ one can compute the reservation prices

$$r_b = \frac{1}{\rho} - \int_t^{\infty} \exp(-\rho(\tau - t))\Delta(\tau)\left(\lambda_d + \frac{\nu_b(\tau)}{\mu_b(\tau)} + n\alpha\mu_s(\tau)\left(\gamma_b(\tau) - \gamma_s(\tau)\right)\right) d\tau, \tag{3.4.12}$$

$$r_s = \frac{1-\delta}{\rho} + \int_t^{\infty} \exp(-\rho(\tau - t))\Delta(\tau)\left(\lambda_u + \frac{\nu_s(\tau)}{\mu_s(\tau)} + n\alpha\mu_b(\tau)\left(\gamma_s(\tau) - \gamma_b(\tau)\right)\right) d\tau, \tag{3.4.13}$$

where we imposed the transversality condition

$$\lim_{t\to\infty} \exp(-\rho t) r_b(t) = 0,\ \lim_{t\to\infty} \exp(-\rho t) r_s(t) = 0.$$

Proposition 24. *There exists a unique symmetric Markov perfect equilibrium in the model. The price strategies and measures of agents are given by Propositions 17 and 23. The expressions for gains from trade and reservation prices of buyers and*

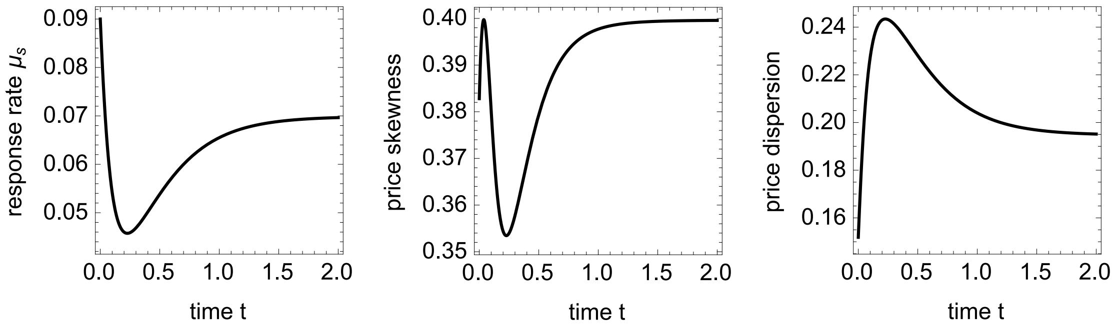

Figure 3.4.2: Dynamics of the response rate and price skewness. Parameter values are $\alpha = \lambda_u = 2$, $\lambda_d = 1$ $n = 35$, $s = 0.6$, $\eta^0 = 0.9$, $\mu_s^0 = 0.09$.

sellers are given by (3.4.10-3.3.20).

Proof. See Appendix. □

3.4.3 Transaction price skewness and dispersion and dealer response rate

The transaction prices in non-stationary equilibrium are given by the same expressions as before. In the section 3.1 we provided cross-section implications linking dealer's response rate to price skewness and dispersion. In this section we augment those by time-series implications.

1. *Price dispersion is lower in periods of time with higher response rate of dealers.*

2. *The absolute value of skewness is the highest in periods of time with intermediary response rate of dealers.*

3. *The distribution of scaled bid (ask) prices first-order stochastically dominates that in periods of time with lower (higher) response rates.*

Figure 3.4.2 illustrates points 1 and 2.

3.4.4 Price dynamics

In this section we consider the dynamics of how prices respond to a supply shock. We assume the economy starts in a stationary state corresponding to the supply s, and at time 0 the supply changes to s'. We then study how the economy reaches the new steady-state. Since the economy starts at the steady state and there is

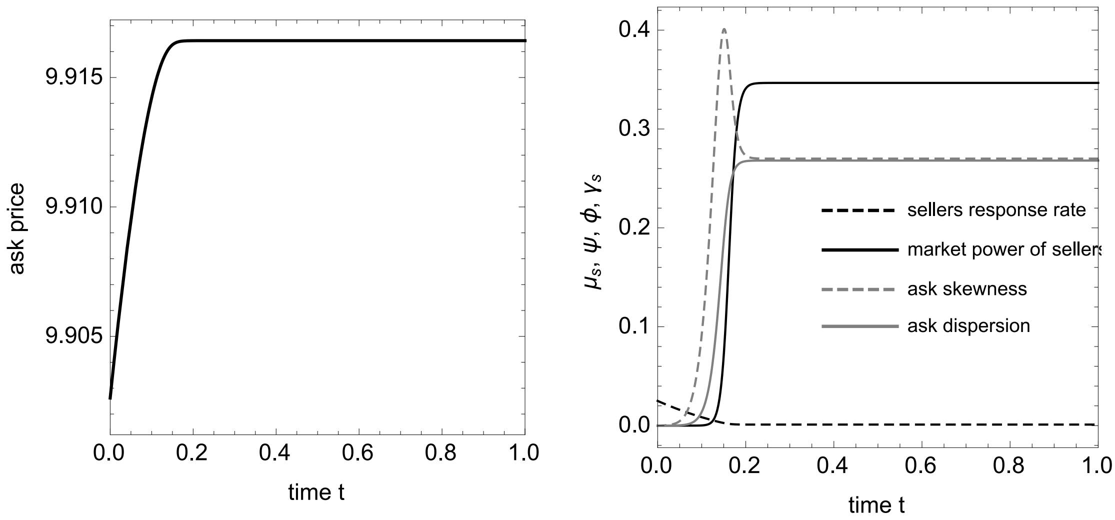

Figure 3.4.3: Dynamics of a short squeeze. Parameter values are $\alpha = \lambda_u = \lambda_d = 1$, $n = 100$, $s = 0.2$, $s' = 0.3$ $\eta^0 = 0.5$.

no change in neither λ_d nor λ_u, the initial measure of high-types is equal to its' steady-state value, $\eta^0 = \eta(t) = \eta^* \equiv \frac{\lambda_u}{\lambda_u+\lambda_d}$. The latter implies that in the ODE (3.4.3) the function $m(\mu_s, t) = m(\mu_s)$, i.e., it does not depend on t directly, which simplifies the analysis.

Figure 3.4.3 demonstrates the dynamic aspect of liquidity squeeze studied in section 2. Even though the the (ask) price changes smoothly, the (ask) skewness and dispersion jump. The intuition for the jump is skewness and dispersion is as follows: as time goes by, the measure of sellers decreases. Since n is very large, market power of sellers is almost zero at times when the measure of sellers is not zero. However, as the measure of sellers approaches zero, their market power jumps, resulting in non-trivial skewness and dispersion.

3.4.5 Efficiency and welfare

In this section we consider the dynamics of welfare. Summing the ODEs for the value functions we obtain the ODE for welfare

$$0 = \dot{\mathcal{W}} - \rho\mathcal{W} + s - \mu_s(t)\delta,$$

implying a solution

$$\mathcal{W} = \int_t^\infty \exp\left(-\rho(\tau - t)\right)\left(s - \mu_s(\tau)\delta\right) d\tau.$$

The efficient allocation implies $\mu_s = \mu_s^e$. However, because of the search friction, the efficient allocation will not be achieved instantaneously, even if $s \in [0,1] \backslash (\underline{s}, \overline{s})$.

Proposition 25. *As $n \to \infty$, the efficient allocation is achieved if $s \in [0,1] \backslash (\underline{s}, \overline{s})$, after a time*

$$\frac{\ln |\mu_s^0 - \mu_s^\infty| - \ln |\mu_s^\infty|}{2\alpha + \lambda_d + \lambda_u}, \text{ if } s < \underline{s}$$

and

$$\frac{\ln |\mu_s^0 - \mu_s^\infty| - \ln |\eta^* - s|}{2\alpha + \lambda_d + \lambda_u}, \text{ if } s > \overline{s},$$

where

$$\mu_s^\infty \equiv \frac{s(\alpha + \lambda_d) - \alpha\eta}{(2\alpha + \lambda_d + \lambda_u)}.$$

Proof. Consider the ODE

$$\dot{\mu}_s = \mu_h \lambda_d - \mu_s \lambda_u - \nu_s - \nu_b$$

in the limit as $n \to \infty$. Suppose that $0 < \mu_s < \eta^* - s$, i.e. both the measure of sellers and the measure of buyers are positive. Then, $\nu_s = \alpha\mu_s$ and $\nu_b = \alpha\mu_b = \alpha\left(\eta^* - s + \mu_s\right)$. The ODE above becomes

$$\dot{\mu}_s = s(\alpha + \lambda_d) - \alpha\eta - \mu_s(2\alpha + \lambda_d + \lambda_u),$$

which is a linear ODE. Denote $\mu_s^\infty \equiv \frac{s(\alpha+\lambda_d)-\alpha\eta}{(2\alpha+\lambda_d+\lambda_u)}$. Now note, that since by assumption $s \in [0,1] \backslash (\underline{s}, \overline{s})$, we have either $\mu_s^\infty < 0$ or $\mu_s^\infty > \eta^* - s$ (i.e. we hit one of the boundaries). Integrating the ODE we get

$$\frac{\ln |\mu_s^0 - \mu_s^\infty| - \ln |\mu_s(t) - \mu_s^\infty|}{2\alpha + \lambda_d + \lambda_u} = t,$$

from which the proposition follows. □

Figure 3.4.4 provides an illustration. We also see that increasing n improves welfare both in steady-state and in transition.

3.5 Conclusion

We present a tractable model of an OTC market with multilateral trading platform. Our model is tractable both in steady state and in transition dynamics. This makes

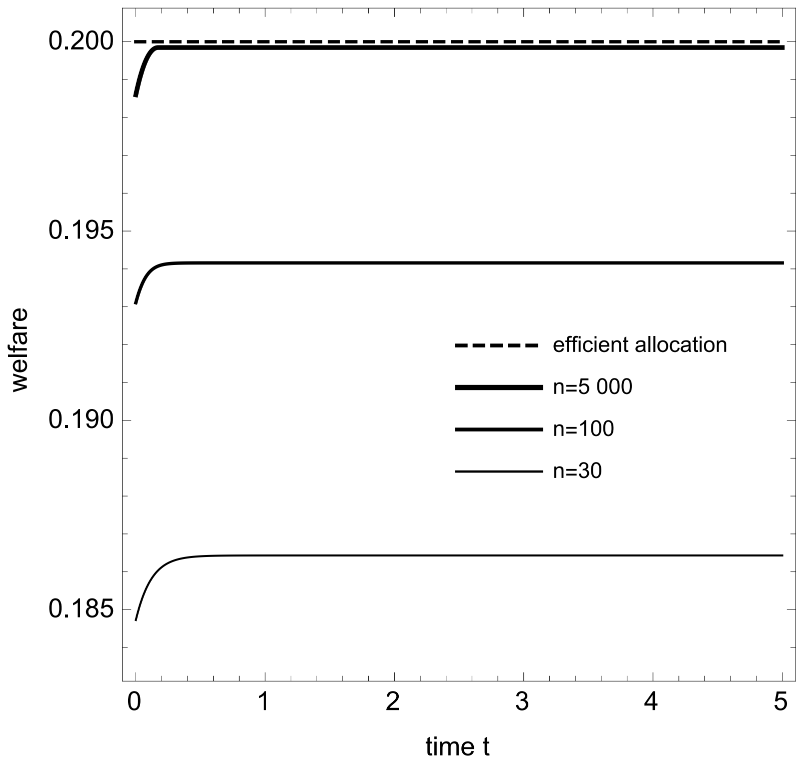

Figure 3.4.4: Dynamics of welfare. Parameter values are $\alpha = \lambda_u = \lambda_d = 1$, $n = 100$, $s = 0.3$, $s' = 0.2$ $\eta^0 = 0.5$.

us hope that the possible extensions which we outline below may be tractable as well.

First, in our model the number of dealers that a customer can query is exogenous and there is no cost of querying several dealers. It is therefore desirable to introduce such a cost and to endogenize the number of dealers queried. This extension is a subject of our ongoing research.

Second, in many markets traders use both MTP and traditional, "voice", ways of trade. It is thus desirable to give such a possibility to traders in our model and to examine what determines the choice between the two, and whether or not they can coexist in a long run.

Finally, in our model there are just two types of traders. It might be interesting to extend our results allowing a continuum of types to examine how the strategic force behind the price dispersion which our model highlights interacts with the price dispersion due to the dispersion of valuations (as highlighted by Hugonnier et al (2015) and Shen et al (2015)).

3.6 Appendix

Proof. (Proposition 17) (There are no point masses in the support of $A(p)$). Suppose that there is a point mass at some $p = p_m$. Since in a mixed strategy equilibrium the profit should be constant, the profit of following the mixed strategy is the same as the profit of charging the price p_m. However, undercutting the price $p_m > r_s$ by a small amount yields a strictly greater profit. If there is a point mass at $p = r_s$, then the profit is zero and charging any price between r_s and r_b is a deviation since it yields strictly positive profit.

(The support of $B(p)$ is connected) Suppose not and there is a gap: prices $p \in (p_1, p_2) \in [r_s + \pi_s, r_b]$ are charged with zero probability. Charging a price $p \in (p_1, p_2)$ is a deviation: price p yields greater profit than p_1 since it has the same probability of being the best one, but yields a greater profit.

($\overline{p}_a = r_b$) It is clear that prices above the reservation price r_b will not be charged in equilibrium. It remains to show that $\overline{p}_a < \overline{p}$ can not be true in equilibrium. Indeed, if $\overline{p}_a < \overline{p}$ then charging price $\hat{p} \in (\overline{p}_a; \overline{p})$ with probability one is a deviation: the profit of following the strategy $B(p)$ is $(1-\mu_s)^m (\overline{p}_a - \overline{p})$ while charging $\hat{p}$ with probability 1 yields greater profit of $(1-\mu_s)^m (\hat{p} - \overline{p})$.

(Derivation of $B(p)$) Analogous to the case of sellers one can prove that $B(p)$ has no point masses and has connected support $[\underline{p_b}, \overline{p_b}]$. The following is true for the profit of a buyer

$$
\begin{aligned}
\pi_b &= Pr(\text{qoute } p \text{ is the best})(r_b - p) \\
&= (1 - \mu_b + \mu_b B(p))^{n-1}(r_b - p) \\
&= \text{const} \\
&= (1-\mu_b)^{n-1}(r_b - \underline{p_b}).
\end{aligned}
$$

It can be proven analogous to the case of sellers that $\underline{p_b} = r_s$ from which the expressions for π_b and $B(p)$ follow. □

Proof. (Proposition 19) Subtracting (3.3.12) from (3.3.10) and (3.3.11) from (3.3.13) yields

$$
V_h - V_s = \frac{\delta}{\rho + \lambda_u + \lambda_d} - \frac{\Delta}{\rho + \lambda_u + \lambda_d}\left(\frac{\nu_s}{\mu_s} + n\alpha\mu_b\left(\gamma_s - \gamma_b\right)\right),
$$

$$V_b - V_o = \frac{\Delta}{\rho + \lambda_u + \lambda_d}\left(\frac{\nu_b}{\mu_b} + n\alpha\mu_s\left(\gamma_b - \gamma_s\right)\right).$$

To prove that $\Delta > 0$ it is sufficient to show that $n(\mu_s - \mu_b)\left(\gamma_b - \gamma_s\right) > 0$. Denote $f(x) \equiv 1 - (1-x)^n$. Note that $f(x)$ is strictly concave for $x \in (0,1)$. One can write $n(\mu_s - \mu_b)\left(\gamma_b - \gamma_s\right) = (\mu_s - \mu_b)(f'(\mu_b) - f'(\mu_s)) > 0$. The last inequality is due to concavity of $f(\cdot)$. To prove that $r_s > 0$ it is sufficient to prove that $\frac{\nu_b}{\mu_b} - n\alpha\gamma_s > 0$. The latter can be written as $\alpha\left(f(\mu_s) + f'(\mu_s)(0 - \mu_s)\right) > \alpha f(0) = 0$. Again, the inequality follows from the concavity of $f(\cdot)$. $\square$

Proof. (Proposition 20) We derive the distributions below. The computation of moments follows by straightforward integration.

$$\begin{aligned}
F_a(p) = {} & \\
& = Pr(\text{price in a buyer-init. transaction} < p) \\
& = Pr(\text{best quote recieved by a buyer} < p| \text{ a buyer contacted at least one seller}) \\
& = \frac{Pr(\text{a buyer has contacted at least one buyer } \textit{and} \text{ at least one quote } < p)}{Pr(\text{a buyer has contacted at least one seller})} \\
& = \frac{1 - Pr(\text{each contact: either not a seller } \textit{or} \text{ to a seller who quotes} > p)}{\nu_b/\left(\alpha\mu_b\right)} \\
& = \frac{1 - (1 - \mu_s + \mu_s(1 - A(p)))^n}{\nu_b/\left(\alpha\mu_b\right)} \\
& = \frac{\alpha\mu_b}{\nu_b}\left(1 - \left(\frac{\gamma_s\Delta}{p - r_s}\right)^{\frac{n}{n-1}}\right).
\end{aligned}$$

For sellers one can similarly write

$$1 - F_b(p) =$$

$$= Pr(\text{price in a seller-initiated transaction} > p)$$

$$= Pr(\text{best quote recieved by a seller} > p|\text{ a seller contacted at least one buyer})$$

$$= \frac{Pr(\text{a seller contacted at least one buyer } \textit{and} \text{ at least one quote} > p)}{Pr(\text{a seller contacted at least one buyer})}$$

$$= \frac{1 - Pr(\text{each contact is either not a buyer } \textit{or} \text{ to a buyer who quotes} < p)}{\nu_s / (\alpha\mu_s)}$$

$$= \frac{1 - (1 - \mu_b + \mu_b B(p))^n}{\nu_s / (\alpha\mu_s)}$$

$$= \frac{\alpha\mu_s}{\nu_s}\left(1 - \left(\frac{\gamma_b\Delta}{r_b - p}\right)^{\frac{n}{n-1}}\right).$$

From the above we get

$$F_b(p) = 1 - \frac{\alpha\mu_s}{\nu_s}\left(1 - \left(\frac{\gamma_b\Delta}{r_b - p}\right)^{\frac{n}{n-1}}\right).$$

□

Proof. (Proposition 21) Denote μ_s^n the equilibrium measure of sellers for a finite n, and denote by μ_s^∞ the corresponding limit. The measure μ_s^n solves (3.3.9). It is straightforward to show that μ_s^n is decreasing in n. Since the sequence μ_s^n monotonically decreases and is bounded below (by zero) the limit μ_s^∞ exists and is non-negative. Given (3.3.5), the corresponding limit of buyers exists, is given by $\mu_b^\infty = \mu_s^\infty + \eta - s$ and should also be non-negative.

We first consider that case when both μ_s^∞ and μ_b^∞ are positive.

In that case $\nu_b^\infty = \lim_{n\to\infty}(1-(1-\mu_s)^n)\,\alpha\mu_b = \alpha\mu_b$ and

$\nu_s^\infty = \lim_{n\to\infty}(1-(1-\mu_b)^n)\,\alpha\mu_s = \alpha\mu_s$. In the limit $n \to \infty$ equation (3.3.9) becomes

$$\mu_s^\infty\lambda_u + \alpha(2\mu_s^\infty + \eta - s) - (s - \mu_s^\infty)\lambda_d = 0.$$

From the above we express

$$\mu_s^\infty = \frac{s(\alpha+\lambda_d) - \alpha\eta}{2\alpha + \lambda_d + \lambda_u},\ \mu_b^\infty = \mu_s^\infty + \eta - s = \frac{\eta(\alpha+\lambda_d+\lambda_u) - s(\alpha+\lambda_u)}{2\alpha+\lambda_d+\lambda_u}.$$

It then follows that μ_s^∞ and μ_b^∞ are positive iff

$$s > \underline{s} \equiv \frac{\alpha}{\alpha+\lambda_d}\eta \text{ and } s < \overline{s} \equiv \left(1 + \frac{\lambda_d}{\alpha+\lambda_u}\right)\eta.$$

To calculate r_b, r_s and Δ one needs to calculate $\lim_{n\to\infty} n(1-\mu_s)^{n-1}$ and

$\lim_{n\to\infty} n(1-\mu_b)^{n-1}$. Since $n(1-\mu)^{n-1}$ is a decreasing function of μ_s is a decreasing sequence, $\mu_s > \mu_s^\infty$ for all n. Therefore one can write

$$0 < n(1-\mu_s)^{n-1} < \underbrace{n(1-\mu_s^\infty)^{n-1}}_{\to 0 \text{ if } \mu_s^\infty > 0}.$$

Since we conjectured that $\mu_s^\infty > 0$, by Sandwich Theorem we get $\lim_{n\to\infty} n(1-\mu_s)^{n-1} = 0$. Similarly, $\lim_{n\to\infty} n(1-\mu_b)^{n-1} = 0$.

Substituting the above into (3.3.18-3.3.20) one gets

$$\Delta = \frac{\delta}{\rho + \lambda_u + \lambda_d + 2\alpha}, \tag{3.6.1}$$

$$r_b = \frac{1}{\rho} - \frac{\Delta}{\rho}\left(\lambda_d + \alpha\right), \tag{3.6.2}$$

$$r_s = \frac{1-\delta}{\rho} + \frac{\Delta}{\rho}\left(\lambda_u + \alpha\right). \tag{3.6.3}$$

Since $\langle p_b \rangle = r_s$ and $\langle p_s \rangle = r_b$ the result follows.

We next consider the case $\mu_s^\infty = 0$. Since $\mu_b^\infty = \mu_s^\infty + \eta - s = \eta - s$, the former is possible only if $\eta - s \geq 0$. We further assume the strict equality, $\eta - s > 0$. In that case $\nu_s = 0$, but $\nu_b = \alpha\mu_b(1-(1-\mu_s)^n)$ is indeterminate. To find it we substitute $\nu_s = \mu_s = 0$ to (3.3.8) and express

$$\nu_b = \lambda_d s.$$

From $\gamma_s = (1-\mu_s)^{n-1} = \frac{1}{1-\mu_s}\left(1 - \frac{\nu_b}{\alpha\mu_b}\right)$ we also express

$$\gamma_s = 1 - \frac{\lambda_d s}{\alpha\mu_b} = 1 - \frac{\lambda_d s}{\alpha\left(\eta - s\right)},$$

which is smaller than 1 iff $s < \underline{s}$. Substituting the above into (3.3.18-3.3.20) we get

$$\Delta = 0,\ r_b = r_s = \frac{\delta}{\rho},$$

from which the result follows. The case $s > \overline{s}$ corresponding to $\mu_b^\infty = 0$ is considered analogously. □

Proof. (Proposition 22) Dividing both parts of (3.3.9) by α and taking the limits one can get that $\mu_s \to \mu_s^e$ as $\alpha \to \infty$.

It follows from the proof of Proposition *21* that $\mu_s \to \mu_s^e$ as $n \to \infty$, for $s \in [0,1]\backslash(\underline{s}, \overline{s})$. ☐

*Proof. (Proposition 24)*As in the stationary case we derive the expressions for r_b, r_s, $V_h - V_s$ and $V_b - V_o$. Having these expressions one can easily derive all value functions, The expressions for the firs two are given by (3.4.12-3.4.13). Subtracting (3.4.7) from (3.4.5) and (3.4.6) from (3.4.8) one get the following ODEs:

$$0 = \frac{d}{dt}\left(V_h - V_s\right) + \delta - \Delta\left(\frac{\nu_s}{\mu_s} + n\alpha\mu_b\left(\gamma_s - \gamma_b\right)\right) - \left(\rho + \lambda_u + \lambda_d\right)\left(V_h - V_s\right),$$

$$0 = \frac{d}{dt}\left(V_b - V_o\right) + \Delta\left(\frac{\nu_b}{\mu_b} + n\alpha\mu_s\left(\gamma_b - \gamma_s\right)\right) - \left(V_b - V_o\right)\left(\rho + \lambda_u + \lambda_d\right),$$

which one can solve in closed form as follows

$$V_h(t) - V_s(t) = \frac{\delta}{\rho + \lambda_u + \lambda_d} -$$

$$- \int_t^\infty \exp(-(\rho + \lambda_u + \lambda_d)(\tau - t))\Delta(\tau)\left(\frac{\nu_s(\tau)}{\mu_s(\tau)} + n\alpha\mu_b(\tau)\left(\gamma_s(\tau) - \gamma_b(\tau)\right)\right) d\tau, \tag{3.6.4}$$

$$V_b(t) - V_o(t) =$$

$$= \int_t^\infty \exp(-(\rho + \lambda_u + \lambda_d)(\tau - t))\Delta(\tau)\left(\frac{\nu_b(\tau)}{\mu_b(\tau)} + n\alpha\mu_s(\tau)\left(\gamma_b(\tau) - \gamma_s(\tau)\right)\right) d\tau, \tag{3.6.5}$$

where we imposed the transversality condition

$$\lim_{t\to\infty} \exp(-\rho t)\left(V_h(t) - V_s(t)\right) = 0,\ \lim_{t\to\infty} \exp(-\rho t)\left(V_b(t) - V_o(t)\right) = 0.$$

The transversality conditions ensure sufficiency. It clearly follows from (3.4.10) that our conjecture of positive gains from trade holds. ☐

Chapter 4

Bibliography

[1] Albagli E., Hellwig C., Tsyvinski A. “A theory of asset pricing based on heterogeneous information.” – National Bureau of Economic Research, 2011. – №. w17548.

[2] Almgren, Robert, Chee Thum, Emmanuel Hauptmann, and Hong Li. "Direct estimation of equity market impact." Risk 18 (2005): 5752.

[3] Almgren, Robert, and Neil Chriss. "Optimal execution of portfolio transactions." Journal of Risk 3 (2001): 5-40.

[4] Ausubel, L. M., Cramton, P., Pycia, M., Rostek, M., & Weretka, M. “Demand reduction and inefficiency in multi-unit auctions.” The Review of Economic Studies (2014): 81(4), 1366-1400.

[5] Babus, Ana, and Peter Kondor. Trading and information diffusion in over-the-counter markets. Working paper, Imperial College and Central European University, 2013.

[6] Back, Kerry, and Shmuel Baruch. "Strategic liquidity provision in limit order markets." Econometrica 81.1 (2013): 363-392.

[7] Banerjee, Snehal, and Brett Green. "Signal or noise? Uncertainty and learning about whether other traders are informed." Journal of Financial Economics (2015).

[8] Banerjee, Snehal, and Jeremy J. Graveline. "The Cost of Short-Selling Liquid Securities." The Journal of Finance 68.2 (2013): 637-664.

[9] Banerjee, Snehal, Jesse Davis, and Naveen Gondhi. "When transparency improves, must prices reflect fundamentals better?." (2015).

[10] Barlevy, Gadi, and Pietro Veronesi. "Rational panics and stock market crashes." Journal of Economic Theory 110.2 (2003): 234-263.

[11] Basak, Suleyman, and Anna Pavlova. "Asset Prices and Institutional Investors." The American Economic Review 103.5 (2013): 1728-1758.

[12] Baruch, Shmuel. "Who Benefits from an Open Limit-Order Book?." The Journal of Business 78.4 (2005): 1267-1306.

[13] Bhattacharya, Utpal, and Matthew Spiegel. "Insiders, outsiders, and market breakdowns." Review of Financial Studies 4.2 (1991): 255-282.

[14] Biais, Bruno, David Martimort, and Jean-Charles Rochet. "Competing mechanisms in a common value environment." Econometrica 68.4 (2000): 799-837.

[15] Biais, Bruno, Larry Glosten, and Chester Spatt. "Market microstructure: A survey of microfoundations, empirical results, and policy implications." Journal of Financial Markets 8.2 (2005): 217-264.

[16] Breon-Drish, Bradyn. "On Existence and Uniqueness of Equilibrium in a Class of Noisy Rational Expectations Models." The Review of Economic Studies (2015): rdv012.

[17] Brunnermeier, Markus Konrad. Asset pricing under asymmetric information: Bubbles, crashes, technical analysis, and herding. Oxford University Press, 2001.

[18] Burdett, Kenneth, and Kenneth L. Judd. "Equilibrium price dispersion." Econometrica: Journal of the Econometric Society (1983): 955-969.

[19] Butters, Gerard R. "Equilibrium distributions of sales and advertising prices." The Review of Economic Studies 44.3 (1977): 465-491.

[20] Boulatov, Alex, and Dan Bernhardt. "Robustness of equilibrium in the Kyle model of informed speculation." Annals of Finance 11.3-4 (2015): 297-318.

[21] Cespa, Giovanni, and Xavier Vives. "Dynamic trading and asset prices: Keynes vs. Hayek." The Review of Economic Studies 79.2 (2012): 539-580.

[22] Chabakauri, Georgy, Kathy Yuan, and Konstantinos E. Zachariadis. "Multi-asset noisy rational expectations equilibrium with contingent claims." Available at SSRN 2446873 (2015).

[23] Chan, Louis KC, and Josef Lakonishok. "Institutional trades and intraday stock price behavior." Journal of Financial Economics 33.2 (1993): 173-199.

[24] Chan, Louis KC, and Josef Lakonishok. "The behavior of stock prices around institutional trades." The Journal of Finance 50.4 (1995): 1147-1174.

[25] Chen, Joseph, Harrison Hong, and Jeremy C. Stein. "Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices." Journal of Financial Economics 61.3 (2001): 345-381.

[26] Cespa, Giovanni, and Thierry Foucault. "Illiquidity contagion and liquidity crashes." Review of Financial Studies (2014): hhu016.

[27] Cespa, Giovanni, and Xavier Vives. "The Beauty Contest and Short-Term Trading." The Journal of Finance (2015).

[28] Cheng, Haw, Andrei Kirilenko, and Wei Xiong. "Convective risk flows in commodity futures markets." Review of Finance (2014): rfu043.

[29] Cuoco, Domenico, and Ron Kaniel. "Equilibrium prices in the presence of delegated portfolio management." Journal of Financial Economics 101.2 (2011): 264-296.

[30] Chiyachantana, Chiraphol N., Pankaj K. Jain, Christine Jiang, and Robert A. Wood. "International evidence on institutional trading behavior and price impact." The Journal of Finance 59, no. 2 (2004): 869-898.

[31] Conrad, Jennifer, Robert F. Dittmar, and Eric Ghysels. "Ex ante skewness and expected stock returns." The Journal of Finance 68.1 (2013): 85-124.

[32] DeMarzo, Peter, and Costis Skiadas. "Aggregation, determinacy, and informational efficiency for a class of economies with asymmetric information." Journal of economic theory 80.1 (1998): 123-152.

[33] DeMarzo, Peter, and Costis Skiadas. "On the uniqueness of fully informative rational expectations equilibria." Economic Theory 13.1 (1999): 1-24.

[34] DeMarzo, Peter M., and Branko Urosevic. "Ownership dynamics and asset pricing with a large shareholder." Journal of Political Economy 114.4 (2006): 774-815.

[35] Diether, Karl B., Christopher J. Malloy, and Anna Scherbina. "Differences of opinion and the cross section of stock returns." Journal of finance (2002): 2113-2141.

[36] Du, Songzi, and Haoxiang Zhu. Welfare and optimal trading frequency in dynamic double auctions. No. w20588. National Bureau of Economic Research, 2014.

[37] Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen. "Over-the-Counter Markets." Econometrica 73.6 (2005): 1815-1847.

[38] Ganguli, Jayant Vivek, and Liyan Yang. "Complementarities, multiplicity, and supply information." Journal of the European Economic Association 7.1 (2009): 90-115.

[39] Gennotte, Gerard, and Hayne Leland. "Market liquidity, hedging, and crashes." The American Economic Review (1990): 999-1021.

[40] Goldstein, Itay, and Liyan Yang. "Commodity Financialization: Risk Sharing and Price Discovery in Commodity Futures Markets." (2015).

[41] Grossman, Sanford J. "The existence of futures markets, noisy rational expectations and informational externalities." The Review of Economic Studies (1977): 431-449.

[42] Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz. "Trading Costs of Asset Pricing Anomalies." (2014).

[43] Foster, F. Douglas, and S. Viswanathan. "Strategic trading when agents forecast the forecasts of others." Journal of Finance (1996): 1437-1478.

[44] Foster, F. Douglas, and S. Viswanathan. "Strategic trading when agents forecast the forecasts of others." The Journal of Finance 51.4 (1996): 1437-1478.

[45] Gemmill, Gordon. "Transparency and liquidity: A study of block trades on the London Stock Exchange under different publication rules." The Journal of Finance 51.5 (1996): 1765-1790.

[46] Glebkin, Sergei, Marzena J. Rostek and Ji Hee Yoon. "Uncertainty and equilibrium in supply function games", working paper (2015).

[47] Griffin, John M., Jeffrey H. Harris, and Selim Topaloglu. "The dynamics of institutional and individual trading." The Journal of Finance 58.6 (2003): 2285-2320.

[48] Gromb, Denis, and Dimitri Vayanos. "Equilibrium and welfare in markets with financially constrained arbitrageurs." Journal of financial Economics 66.2 (2002): 361-407.

[49] Hasbrouck, Joel. "Measuring the information content of stock trades." The Journal of Finance 46.1 (1991): 179-207.

[50] Hasbrouck, Joel. "High frequency quoting: Short-term volatility in bids and offers." Available at SSRN 2237499 (2015).

[51] Hausman, Jerry A., Andrew W. Lo, and A. Craig MacKinlay. "An ordered probit analysis of transaction stock prices." Journal of financial economics 31.3 (1992): 319-379.

[52] Hendershott, Terrence, Dmitry Livdan, and Norman SchÃŒrhoff. "Are institutions informed about news?." Journal of Financial Economics (2015).

[53] Hendershott, Terrence, and Ananth Madhavan. "Click or Call? Auction versus Search in the Over-the-Counter Market." The Journal of Finance 70.1 (2015): 419-447.

[54] Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill. Heterogeneity in decentralized asset markets. No. w20746. National Bureau of Economic Research, 2014.

[55] Huang, Shiyang. "Delegated Information Acquisition and Asset Pricing.", Working Paper, LSE (2015).

[56] Hu, Conghui, and Wei Xiong. "Are Commodity Futures Prices Barometers of the Global Economy?" AprÃšs le DÃ©luge: Finance and the Common Good after the Crisis, edited by E. Glen Weyl, Edward L. Glaeser, and Tano Santos (2013).

[57] Klemperer, Paul D., and Margaret A. Meyer. "Supply function equilibria in oligopoly under uncertainty." Econometrica: Journal of the Econometric Society (1989): 1243-1277.

[58] Holthausen, Robert W., Richard W. Leftwich, and David Mayers. "The effect of large block transactions on security prices: A cross-sectional analysis." Journal of Financial Economics 19.2 (1987): 237-267.

[59] Holthausen, Robert W., Richard W. Leftwich, and David Mayers. "Large-block transactions, the speed of response, and temporary and permanent stock-price effects." Journal of Financial Economics 26.1 (1990): 71-95.

[60] Jovanovic, Boyan, and Albert J. Menkveld. "Dispersion and Skewness of Bid Prices." Available at SSRN 2463066 (2015).

[61] Huberman, Gur, and Werner Stanzl. "Optimal liquidity trading." Review of Finance 9.2 (2005): 165-200.

[62] Kagel, John H., and Dan Levin. "Behavior in Multi-Unit Demand Auctions: Experiments with Uniform Price and Dynamic Vickrey Auctions." Econometrica 69.2 (2001): 413-454.

[63] Keim, Donald B., and Ananth Madhavan. "Anatomy of the trading process: Empirical evidence on the behavior of institutional traders." Journal of Financial Economics 37.3 (1995): 371-398.

[64] Keim, Donald Bruce, and Ananth Madhavan. "The upstairs market for large-block transactions: Analysis and measurement of price effects." Review of Financial Studies 9.1 (1996): 1-36.

[65] Keim, Donald B., and Ananth Madhavan. "Transactions costs and investment style: an inter-exchange analysis of institutional equity trades." Journal of Financial Economics 46.3 (1997): 265-292.

[66] Keim, Donald B., and Ananth Madhavan. "The cost of institutional equity trades." Financial Analysts Journal (1998): 50-69.

[67] Klemperer, Paul. "Auction theory: A guide to the literature." Journal of economic surveys 13.3 (1999): 227-286.

[68] Klemperer, Paul D., and Margaret A. Meyer. "Supply function equilibria in oligopoly under uncertainty." Econometrica: Journal of the Econometric Society (1989): 1243-1277.

[69] Krishnamurthy, Arvind. "The bond/old-bond spread." Journal of Financial Economics 66.2 (2002): 463-506.

[70] Korajczyk, Robert A., and Ronnie Sadka. "Are momentum profits robust to trading costs?." The Journal of Finance 59.3 (2004): 1039-1082.

[71] Kremer, Ilan, and Kjell G. Nyborg. "Underpricing and market power in uniform price auctions." Review of Financial Studies 17, no. 3 (2004): 849-877.

[72] Kyle, Albert S. "Continuous auctions and insider trading." Econometrica: Journal of the Econometric Society (1985): 1315-1335.

[73] Kyle, Albert S. "Informed speculation with imperfect competition." The Review of Economic Studies 56.3 (1989): 317-355.

[74] Kyle, Albert S. "Market structure, information, futures markets, and price formation." (1984).

[75] Lagos, Ricardo, and Guillaume Rocheteau. "Liquidity in asset markets with search frictions." Econometrica 77.2 (2009): 403-426.

[76] Lukacs, Eugene. “Characteristic functions.” Vol. 4. London: Griffin, 1970.

[77] Malamud, Semyon, and Marzena J. Rostek. Decentralized exchange. University of Geneva, 2013.

[78] Makarov, Dmitry, and Astrid V. Schornick. "A note on wealth effect under CARA utility." Finance Research Letters 7.3 (2010): 170-177.

[79] Nagel, Stefan. "Evaporating liquidity." Review of Financial Studies 25.7 (2012): 2005-2039.

[80] Nappi, Carmine. "The global aluminium industry 40 years from 1972." London: International Aluminium Institute (2013).

[81] Martin, Ian WR. "Consumption-based asset pricing with higher cumulants." The Review of Economic Studies 80.2 (2013): 745-773.

[82] Muravyev, Dmitriy. "Order flow and expected option returns." Journal of Finance, forthcoming.

[83] Nagel, Stefan. "Evaporating liquidity." Review of Financial Studies 25.7 (2012): 2005-2039.

[84] Obizhaeva, Anna A., and Jiang Wang. "Optimal trading strategy and supply/demand dynamics." Journal of Financial Markets (2012).

[85] OPEC. "Annual Statistical Bulletin 2015." Organization of Petroleum Exporting Countries, Vienna, Austria (2015).

[86] Pagano, Marco. "Trading volume and asset liquidity." The Quarterly Journal of Economics (1989): 255-274.

[87] Rostek, Marzena, and Marek Weretka. "Dynamic Thin Markets." Review of Financial Studies (2015): hhv027.

[88] Palvolgyi, Domotor, and Gyuri Venter. "Multiple equilibria in noisy rational expectations economies." Available at SSRN 2524105 (2015).

[89] Perez-Quiros, Gabriel, and Allan Timmermann. "Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities." Journal of Econometrics 103.1 (2001): 259-306.

[90] Rochet, Jean-Charles, and Jean-Luc Vila. "Insider trading without normality." The Review of Economic Studies 61.1 (1994): 131-152.

[91] Rahi, Rohit, and Jean-Pierre Zigrand. "Information Aggregation in a Competitive Economy." Available at SSRN (2014).

[92] Rosenthal, Robert W. "A model in which an increase in the number of sellers leads to a higher price." Econometrica: Journal of the Econometric Society (1980): 1575-1579.

[93] Rostek, Marzena, and Marek Weretka. "Information and strategic behavior." Journal of Economic Theory (2014).

[94] Rostek, Marzena, and Marek Weretka. "Price inference in small markets." Econometrica 80.2 (2012): 687-711.

[95] Rostek, Marzena, and Marek Weretka. "Dynamic Thin Markets." Review of Financial Studies (2015): hhv027.

[96] Roşu, Ioanid. "A dynamic model of the limit order book." Review of Financial Studies (2009): hhp011.

[97] Saar, Gideon. "Price impact asymmetry of block trades: An institutional trading explanation." Review of Financial Studies 14.4 (2001): 1153-1181.

[98] Sadka, Ronnie, and Anna Scherbina. "Analyst Disagreement, Mispricing, and Liquidity*." The Journal of Finance 62.5 (2007): 2367-2403.

[99] Shen, Ji, Bin Wei, and Hongjun Yan. Financial intermediation chains in an OTC market. Working paper, 2015.

[100] Shilony, Yuval. "Mixed pricing in oligopoly." Journal of Economic Theory 14.2 (1977): 373-388.

[101] Subrahmanyam, Avanidhar. "Risk aversion, market liquidity, and price efficiency." Review of Financial Studies 4.3 (1991): 417-441.

[102] Singleton, Kenneth J. "Investor flows and the 2008 boom/bust in oil prices." Management Science 60.2 (2013): 300-318.

[103] Stein, Jeremy C. "Informational externalities and welfare-reducing speculation." The Journal of Political Economy (1987): 1123-1145.

[104] Sockin, Michael, and Wei Xiong. "Informational frictions and commodity markets." The Journal of Finance (2015).

[105] Sun, Yeneng. "The exact law of large numbers via Fubini extension and characterization of insurable risks." Journal of Economic Theory 126.1 (2006): 31-69.

[106] Tirole, Jean. The theory of industrial organization. MIT press, 1988.

[107] Varian, Hal R. "A model of sales." The American Economic Review 70.4 (1980): 651-659.

[108] Vayanos, Dimitri. "Strategic trading and welfare in a dynamic market." The Review of Economic Studies 66.2 (1999): 219-254.

[109] Vayanos, Dimitri. "Strategic trading in a dynamic noisy market." The Journal of Finance 56.1 (2001): 131-171.

[110] Vayanos, Dimitri, and Pierre-Olivier Weill. "A Search-Based Theory of the On-the-Run Phenomenon." The Journal of Finance 63.3 (2008): 1361-1398.

[111] Vayanos, Dimitri, and Paul Woolley. "An institutional theory of momentum and reversal." Review of Financial Studies (2013): hht014.

[112] Vives, Xavier. "Strategic supply function competition with private information." Econometrica 79.6 (2011): 1919-1966.

[113] Wang, James JD, and Jaime F. Zender. "Auctioning divisible goods." Economic Theory 19.4 (2002): 673-705.

[114] Weill, Pierre-Olivier. "Leaning against the wind." The Review of Economic Studies 74.4 (2007): 1329-1354.

[115] Weretka, Marek. "Endogenous market power." Journal of Economic Theory 146, no. 6 (2011): 2281-2306.

[116] Yuan, Kathy. "Asymmetric price movements and borrowing constraints: A rational expectations equilibrium model of crises, contagion, and confusion." The Journal of Finance 60.1 (2005): 379-411.

[117] Yueshen, Bart Z. "Uncertain Market Making." Available at SSRN 2667837 (2015).