Securities lending and information transmission: a model of endogenous short-sale constraints

Andrey Pankratov*

email: andrey.pankratov@fsa.ulaval.ca

Université Laval

Abstract

I study short-sale constraints in a market with asymmetric information. I offer a novel approach endogenizing short-sale constraints by including an asset-borrowing market in my model. Short-sellers have to borrow an asset and therefore reveal information to a lender. The lender trades on her own account in addition to charging fees, which motivates the short-seller to hide the information and hinders short sales. I contribute to the literature by modeling the mechanism behind short-selling in the absence of explicit short-selling restrictions that are currently less relevant in practice. The model has new implications for profit distribution and market efficiency.

GEL codes: G11, G12, G14

Keywords: Security lenders, Short interest, Information leakages, Distribution of profits, Short-sale constraints, Market efficiency

*I wish to thank my former thesis advisor Antonio Mele for thorough discussions of this paper. I am grateful to Pete Kyle, Loriano Mancini, Laurent Frésard, Jérôme Detemple, Paul Schneider, Francesco Franzoni, Elisabeth Kempf, Peter Gruber, Tomas Fiala, Oksana Bashchenko, Alexandra Matyunina, Tina Oreški, Mirela Sandulescu, Federico Severino, Alessio Ruzza, Edoardo Slerca, Ina Bialova, Marc Jean-Paul van Uffelen, Andrea Barbon, Ekaterina and Dmitri Azarnyh, Davide Tedeschini, and Oleg Pankratov for insightful comments and suggestions. I am also extremely thankful to Thierry Foucault, Pekka Honkanen, Johan Hombert, Daniel Schmidt, and Ioanid Roşu for discussions and comments and suggestions starting from the early strages of my project.

1 Introduction

In fact he often soothed himself with the thought that in all these years he had never gambled for himself, but had always acted strictly for others instead.

—Theodore Dreiser,
The Financier

In this paper, I aim to study how short-sale constraints can arise naturally without explicit short-sale bans. I offer a parsimonious model of endogenous short-sale constraint that has implications for many other features of the financial markets – informational efficiency, volatility, and the distribution of trading profits between the agents. Unlike the existing literature about short-sale constraints, I assume that the agents are free to short-sell assets on financial markets. Nevertheless, the short-sellers have to borrow the asset from a security lender in order to open a short position. This assumption in my model is designed to mimic the real borrowing practice on financial markets, where short-sellers borrow assets from mutual funds via brokers.

I work in a setting with asymmetric information and the short-sellers reveal their signals to the security lender through their borrowing decisions. In my model, the natural short-sale constraint comes from two channels. The first channel is simple avoidance of fees. Security lenders are assumed to have some market power, and they are in a position to set fees for lending the securities. The second channel arises because of informed trading, and represents the main novelty of this paper. The short-sellers are discouraged from borrowing too much because borrowing requests signal low quality of the asset to the security lender. As a result, short sales trigger sales on the part of the security lender and create additional downward price pressure.

Compared to other papers modeling short-selling decisions, the setting in the present paper is more realistic in the context of modern financial markets because the model explicitly incorporates the mechanism behind short-selling as opposed to imposing exogenously a short-selling ban which in practice becomes less relevant.[1] The paper also answers a number of important economic questions that have not been answered in previous literature on short-

[1]For example, in the US, there is a tendency to relax the short-sale restrictions. Historically, during the period 1938-2004, US stocks were subject to so-called uptick rule which prohibited short-selling at a price lower than the previous transaction price. During the period form 2004 to 2007, the uptick rule was abolished and was only partially reinstated in 2010. The new short-sale restriction – so-called alternative uptick rule – applies only to the stocks that lost their value by at least 10% compared to the previous-day closing price.

selling and information diffusion. How exactly do security lenders profit from the lending business? Do the security lenders change the game in terms of percolation of the information, and how does the existence of third-party security lenders affect the decisions and trading profits of the short-sellers? Which factors affect the commission rates on the asset-borrowing market?

I offer an equilibrium model that allows answering all these questions in a coherent way. The model inherits its main assumptions from Kyle (1985) and augments the Kyle model with an additional agent – a security lender. I think of the security lender as of an institution that holds a big inventory of the asset on behalf of their clients who are long-term investors. The security lender is authorized by her clients rehypothecate the asset (e.g., temporarily lend the asset to third parties).

Since the authorization to lend out the assets is a privilege, the security lender has a certain bargaining power over the potential borrowers and can, therefore, charge a substantial interest rate. This interest rate works as a short-sale constraint because it discourages the informed investors from trading on negative signals.

Beyond that, there is another reason for informed agents to refrain from trading or moderate the aggressiveness – an information-driven deterrant. When an informed trader sells the asset short, she borrows this asset and by doing so reveals information to the security lender. This may create an opportunity for the security lender to also trade on the negative signal, thereby competing with the informed trader.

Moreover, even if the informed trader has a positive signal, and therefore buys the asset without a need to explicitly reveal the signal to the security lender, the very fact that the informed trader does not borrow the asset (and does not intend to sell it short), reveals that the asset value might be higher than average, which is a positive signal for the security lender and an opportunity to gain by establishing a long position. However, in the case of a positive signal, the security lender is unable to judge about its magnitude, that is, to what extent the asset fundamental value is higher than its unconditional mean. In other words, because of the lender's front-running[2] behavior, the incentive for the informed agent to buy is reduced because even in this case the informed agent share her profits with a front-running security lender.

It is crucial that for the short-selling case the information-driven deterrent is stronger. Not only the informed agent shares her profits with the front-runner, but there is also a direct incentive for her to reduce the aggressiveness of short-selling because each marginal unit of the asset borrowed triggers extra sales from the security lender. This does not happen in the

[2] I use the term "front-running" in a broad sense, to designate that the lender uses information about trading decisions of another trader for her own profits. Some other authors also follow this approach, as I point out in Section 1.1.

positive case because in such case the lender is ignorant about the size of the buy order.

Overall, I have two ways in which the security lender can profit: receiving commission and front-running. To achieve clarity, I model each of the two ways separately and then bring them together. The variants of my model that include the possibility of front-running contribute to the well-established strand of empirical literature dedicated to the signaling role of short interest, see for example Boehmer, Huszar, and Jordan (2010) and Seneca (1967). As the short interest represents a valuable piece of information, it is natural to assume that the security lender can use it.

I present three variants of the model – with commission only, with front-running only, and a hybrid of the two. In the first setting, the security lender is only allowed to charge a proportional commission for lending the asset to those who want to sell it short, but not make use of the private information. Additionally, I consider a generalization. I give the security lender the freedom to set a variable commission rate. This option can be used in order to either play a screening game with the informed traders or to create an incentive for them to short-sell more. In the second setting, the profits of the security lender come from front-running: trading on the private information contained in informed traders' decisions. In the third setting, the security lender benefits from both the commission and front-running.

The first setting is the most simple one. In this setting, the security lender chooses the interest rate and faces a trade-off between the size of the interest rate and the amount of asset borrowed from her by the market participants. If the security lender charges an interest rate that is too low, then the expected commission revenues are also low because the commission rate affects the expected revenues multiplicatively. On the other hand, if the interest rate set by the security lender is too high, then the informed agents borrow the asset extremely rarely and the expected borrowed quantity becomes very modest. As a result, the expected commission revenues become modest as well. Thus, the other extreme is also suboptimal.

The second setting is a benchmark case. I set the commission rate to zero by assumption in order to isolate the effects that arise due to the lender's ability to front-run the informed trader. In the third setting – a hybrid of commission and front-running – the security lender faces the same trade-off as in the first setting, but on top of that, another force comes into play. In this setting, the interest rate also affects the informational content of short interest. A high interest rate discourages short-selling and thus hinders the information transmission from the informed agent to the security lender. Therefore, in this hybrid case, the security lender has an additional incentive to keep the commission low. Since the security lender aims to maximize overall expected profits from the lending commission and from informed trading, this additional force introduces a distortion into the original model (the first setting). Given the same behavior of the informed trader (off-equilibrium), the security lender will have an incentive to lower the commission rate that she asks for lending the asset in order to encourage more active short-selling and be able to observe the private signal more often. In addition, the fact that the security lender herself trades on her own behalf, will in turn change the informational content of the order flow. The overall demand for the asset will be driven also by her choices.

This, in turn, must induce changes in the behavior of the market participants, and may change the market depth, consequently the optimal trading strategy of the informed trader will be adjusted again. This interplay of the incentives will lead to an equilibrium different from the one without lender's possibility to trade.

In the settings with front-running, in addition to an informed trader and a liquidity trader, I introduce an uninformed borrower. The role of the uninformed borrower is to prevent the security lender from knowing exactly what the informed agent is doing and being able to infer the asset's payoff precisely. Indeed, if the only agents borrowing the asset are the informed traders, then in case of non-zero borrowing position, the security lender can precisely infer value of the private signal. First, this implication would be rather extreme. Second, the assumption of having an uninformed borrower is necessary to have an equilibrium in the model.

Note that the uninformed borrowers in my model are distinct from liquidity traders. The liquidity traders may want to sell the asset in order to raise cash, so they sell only out of their portfolio and they do not need to borrow the asset. Instead, the uniformed borrowers sell short for different reasons. They might be misinformed and seek to profit from this trade or might want to hedge against the risk of a non-tradable endowment.

The contributions of this paper are as follows. First, I explain rationally possible sources of short-sale constraints. Second, I study the implications of the possibility that the security lender may learn about the quality of the asset from the quantity of the asset that the market participants would like to borrow. Third, I find that the commission rates for asset borrowing are likely to depend on whether the security lender has the right to trade on her own account and on the aggressiveness of the uninformed borrowers. If the security lender is unable to front-run or the uninformed borrowers are very active, this is likely to cause an increase in borrowing commission rates.

1.1 Related literature

The present paper models the disincentives to short-sell due to information-retention considerations, and thus contributes to two major strands of finance literature. The first strand is the literature about short-sale constraints. This literature includes Jones and Lamont (2002), Allen, Morris, and Postlewaite (1993), Danielsen and Sorescu (2001), Diamond and

Verrecchia (1987), and Pankratov (2018). The present paper is similar to the last two papers in this list; in the sequel of this section, I explain the distinctions.

The second strand is the literature about "predatory trading", "front-running", "back-running", "mimicking", "spoofing", etc. The common idea of this literature is that certain traders (let's call them "followers") observe the trading decisions of other possibly informed agents and trade using this information. This behavior allows the followers to earn profits either by predicting the future direction of trades (see for example, Sannikov and Skrzypacz (2016), Kervel and Menkveld (2019), and Barbon, Di Maggio, Franzoni, and Landier (2019)), by knowing contemporaneous trading needs (e.g. Danthine and Moresi (1998)), or by predicting asset fundamentals (see for example Huddart, Hughes, and Levine (2001), Yang and Zhu (2020), and Mele and Sangiorgi (2020)). Another common feature is that the players whose behavior is observed (let us call them "leaders") are completely aware of being watched and copied by the followers and take this into account. In my model, fundamentally informed traders play the role of "leaders", while the security lenders play the role of "followers". The key feature of my model is that the follower can only judge about the size of the leader's short position but is absolutely ignorant about the size of the leader's long position. Similarly to Danthine and Moresi (1998), I use the term "front-running" in a broad sense to highlight the attempt of the security lender in possession of trade-related information to use this information for her profits. This does not mean, however, that in my model, the asset lenders predict anyone's future trading needs.[3]

From the game-theoretic perspective the front-running game that I present here is similar to the model of Huddart et al. (2001) with two key distinctions. First, the model is asymmetric, and the security lender ("follower") can judge the magnitude of the private signal of the informed traders only in the negative case. Second, unlike Huddart et al. (2001), I assume that the "follower" trades in the same auction simultaneously with the "leader".

Among the papers about short-sale constraints, the present paper is most similar Diamond and Verrecchia (1987), Pankratov (2018). Moreover, the present paper can be considered as a generalization of Pankratov (2018). These two related papers analyze how short-selling increases market efficiency, while short-sale constraints diminish it. The mechanism is intuitive: in the absence of short-sale constraints, informed agents with negative signal act on their information and reveal it to the market through their trading decisions, while in the presence of short-sale constraints such a revelation is hindered.

One of the first theoretical papers studying this effect is Diamond and Verrecchia (1987). In this paper, the authors introduce an asset with an uncertain payoff. The asset payoff can take only two values ("high" and "low"). Uninformed market participants (market makers)

[3] I use the designation "front-running" throughout the paper.

infer asset payoff by observing the behavior of the informed ones. The price converges to the true payoff as the market makers repeatedly observe the behavior of informed participants and learn from them. The authors show that the short-sale constraints slow down the convergence of the asset prices to the efficient values reflecting the true payoff.

In Pankratov (2018) I study a more realistic setting and propose a model similar to the one considered in the present paper. In contrast to the present paper, in Pankratov (2018), I still consider explicit short-sale bans, but I introduce a probability distribution of the asset with many (at least three) possible outcomes. As a base case, I consider normal distribution to mimic a more realistic probability distribution of the asset's fundamental value and make the model comparable with Kyle (1985). However, in order to generate the main effects, it is enough to consider the case with three possible outcomes "bad", "medium", and "good". With this key assumption of a multinomial/normal distribution, I show that the main prediction of Diamond and Verrecchia (1987) holds, but on top of that, I can generate richer predictions. In particular, Pankratov (2018) explains rationally the asymmetric volatility effect in the stock prices (the fact that volatility in the future is inversely related to the return in the current period).[4] In a nutshell, the explanation is rather intuitive. I assume a short-sale prohibition that applies to some of the informed agents in the model. Under this assumption, a constrained agent who knows that the asset is "bad" cannot reveal this information because of the constraint. In other words, the constrained agents behave the same way regardless of whether the asset is of "bad" or "medium" quality. Thus, from the perspective of the market maker, it is harder to distinguish between "medium" and "bad" than between "medium" and "good" outcomes. This eventually makes the left half of the distribution more opaque from the market maker's point of view, i.e. delays learning in case of bad and neutral outcomes. This leads to higher volatility following negative returns. Most of the implications of the model in Pankratov (2018) hold also in the present paper as long as the asset lender sets a very high interest rate. If the rate is very high, this deterrent works exactly as short-selling ban, the borrowing demand becomes uninformative, and front-running becomes irrelevant.

2 Model structure

In this section, my goal is to merely set the stage – to explain how the agents interact among themselves, the information structure, introduce the reference names for the agents in the model, and stipulate main statistical assumptions.

[4]This phenomenon is also known as the "leverage" effect. This name comes from the first economic explanation (Black, 1976), which eventually turned out to be insufficient.

I model a market for a single asset, whose liquidation value is a continuously distributed random variable, say, normally distributed one $d \sim N(0, \sigma_d^2)$.

The model is composed of the following agents:

1. Informed trader "x" is a rational trader in a possession of private signal d – a pure signal that coincides with the asset's fundamental (liquidation) value.

2. Uninformed borrower "w" is a noise trader who trades for exogenous reasons. The uninformed borrower "w" may even choose to sell short.

3. Lender is an institution (such as a mutual fund or an ETF) that holds massive inventories of the asset and stands ready to lend them to short-sellers on short notice.

4. Liquidity trader "z" is an agent that trades for exogenous liquidity reasons. The liquidity trader "z" never goes short.[5]

5. Market maker is a deep-pocketed risk-neutral agent who observes the overall demand for the asset and sets the price to expected liquidation value given all available information. This agent can be thought of as a perfectly competitive population of arbitrageurs who trade only on public information.

The model is static i.e. there is only one round of trading. At time 0, no one has private signals about the liquidation value. At time 1, the informed trader "x" knows the liquidation value d from her private source of information. As a result of observing demand for lending (a.k.a. short interest), the lender also has a superior insight about d. At time 2 all the agents know precisely d.

In the model, the order size x is determined endogenously in equilibrium, while quantities w and z are exogenous and continuously distributed: say, $w \sim N(0, \sigma_w^2)$ and $z \sim N(0, \sigma_z^2)$. It is assumed that d, w, and z are mutually independent.

Actions in my model happen according to the following protocol:[6]

- First move ($t = 1$):

[5] Since the liquidity trader "z" buys and sells only for liquidity reasons, this trader never has an incentive to borrow the asset. If this agent needs cash, there is no reason to borrow a risky asset because it is easier to directly borrow cash.

[6] Please note that I chose a non-conventional notation. In particular by a decomposition of a real number into "negative part" and "positive part" I mean the following:

$$x = \underbrace{x^-}_{\text{negative part: } \min(x,0)} + \underbrace{x^+}_{\text{positive part: } \max(x,0)},$$

as opposed to a more traditional notation $x^- = |\min(x, 0)|$. In my notation, the "negative part" is indeed negative.

- True value of the asset d is revealed to the INFORMED TRADER "x";
- The INFORMED TRADER "x" borrows if necessary $|x^-|$ shares and submits an order to buy (sell) $x(d)$ shares;
- The UNINFORMED BORROWER "w" borrows if necessary $|w^-|$ shares and submits an order to buy (sell) w shares;

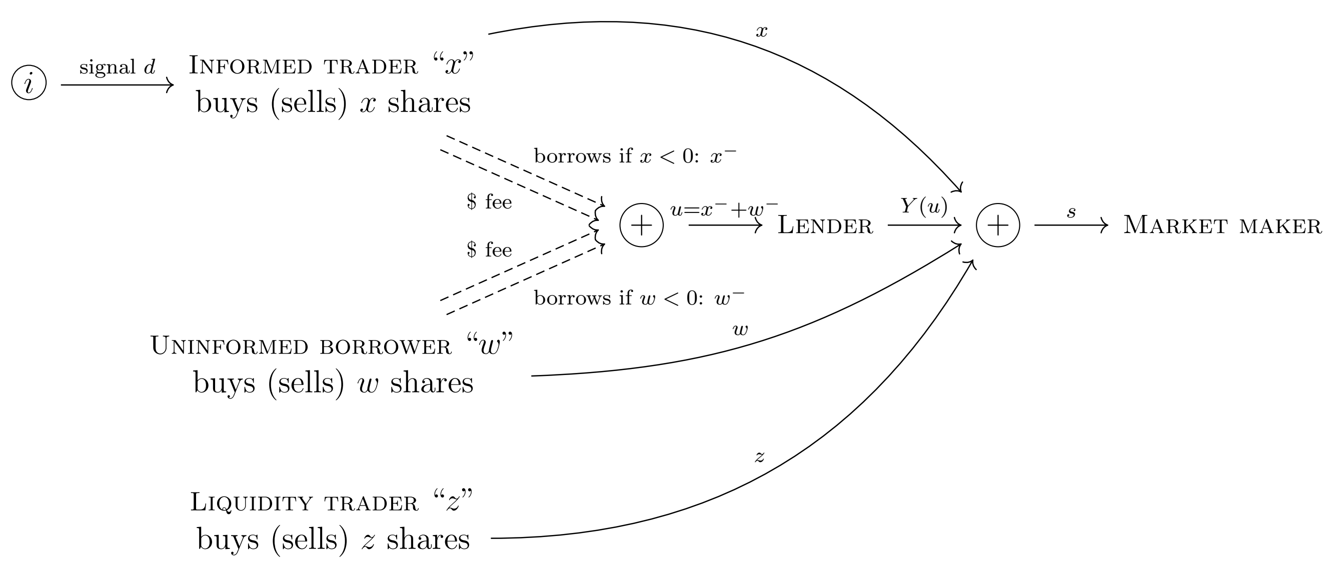
- Second move ($t = 1$):
  - The LENDER observes $x^- + w^-$;
  - The LENDER submits an front-running order to buy (sell) $y\,[x^- + w^-]$ shares;
  - The LIQUIDITY TRADER "z" submits an order to buy (sell) z shares;

- Clearing ($t = 1$):
  - The MARKET MAKER observes $x(d) + w + y\,[x^- + w^-] + z$;
  - The MARKET MAKER sets the price and executes all the orders;

- $t = 2$: The true value d is paid out.

The following diagram summarizes the model.

Figure 1: Full diagram of the model



In the following sections, I dissect my model by segregating the phenomena that arise because of commissions and because of front-running. In particular, in section 3 I isolate implications of the commissions on the financial markets by eliminating the possibility of front-running. Conversely, in section 4 I isolate implications of the front-running on the financial markets by eliminating the possibility of the LENDER charging a commission. Finally, in section 5 I analyze both effects together.

In some of the settings, I exclude the UNINFORMED BORROWER "w" from the model. In those cases, I mention it explicitly.

3 Model without front-running: lender only charges commission

If the INFORMED TRADER "x" borrows the asset, she needs to pay a commission to the LENDER. The commission rate r is set by the LENDER ex-ante. It may be either set exogenously or determined by the LENDER in an optimal way. The commission revenues are proportional to the quantity of the asset borrowed by the traders.

The INFORMED TRADER "x" submits a market order whose size depends on the realization of the private signal d and on the commission rate r. The order size is denoted as $x = X_r(d)$. The only signal that the INFORMED TRADER "x" has is d, so the order size is a function of d. This quantity is positive for buy orders and negative for sell orders. The amount borrowed by the INFORMED TRADER "x" is $|X_r(d)^-|$.

Therefore, the dollar value of the commission paid by the INFORMED TRADER "x" is as follows:

$$R_{\text{Informed trader "}x\text{"}} = r|x^-|. \tag{1}$$

The dollar value of commission paid by the UNINFORMED BORROWER "w" is given by

$$R_{\text{Uninformed borrower "}w\text{"}} = r|w^-| \tag{2}$$

The aggregate amount of commission raised by the LENDER is

$$R_{\text{Lender}} = r \cdot |u|, \tag{3}$$

where u is aggregated negative parts of the order flows:

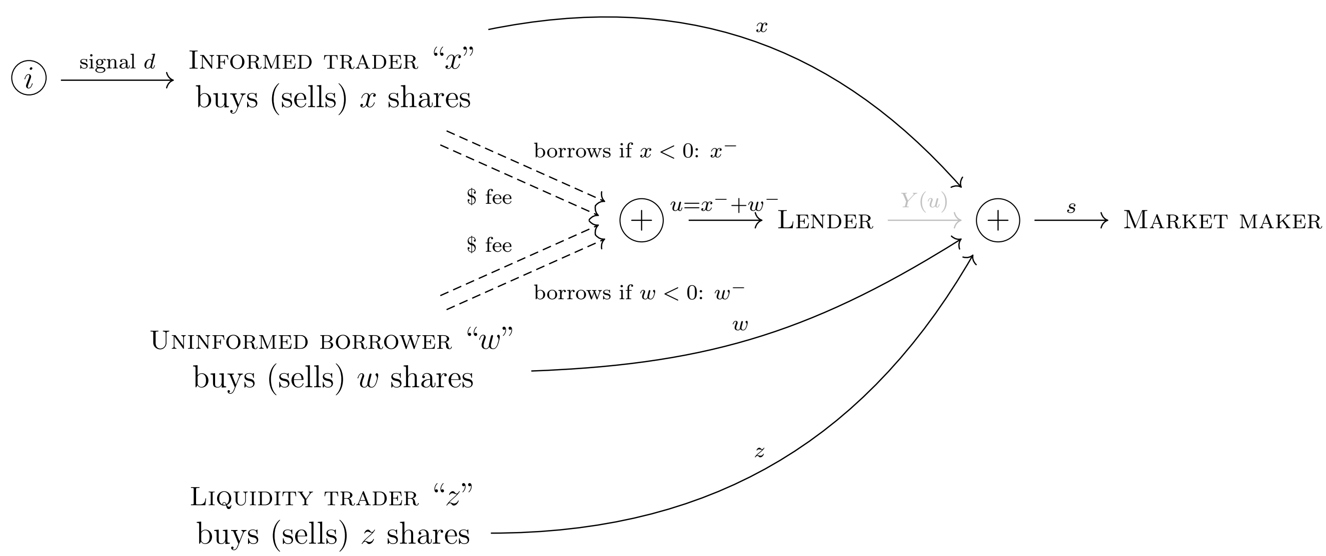
$$u = x^- + w^- \tag{4}$$

In the sequel, I explain how the trading behavior of the INFORMED TRADER "x" depends on the rate of commission rate r charged by the LENDER. In this section, I focus on a baseline model, in which the LENDER's profits only from charging a commission. On average, she earns:

$$\mathrm{E}\left(R_{\text{Lender}}\right) = \mathrm{E}\left(r \cdot |x^- + w^-|\right). \tag{5}$$

The following diagram (Figure 2) summarizes the interaction among the agents in the simplified model presented in this section. The diagram emphasizes that the front-running order "$y(u)$" of the LENDER is excluded from the model.

Figure 2: Diagram of the model of commission



3.1 Market makers

The price is set by the MARKET MAKER based on the information that she observes. At time 0, the MARKET MAKER's information set is empty and she sets the price equal to the unconditional expectation of the liquidation value:

$$p_0 = \mathrm{E}(d) = 0.$$

At time 1, the MARKET MAKER observes the aggregate order flow from the INFORMED TRADER "x", the LIQUIDITY TRADER "z", and the UNINFORMED BORROWER "w": $s = X_r(d) + w + z$, and sets the price accordingly:

$$p_1 = \mathrm{E}(d|s) =: G_r(s).$$

At date 2, the previously private information d is announced publicly and the price becomes equal to the true value:

$$p_2 = d.$$

3.2 Behavior of the informed trader

In this subsection, I describe the forces driving the behavior of the INFORMED TRADER "x". The INFORMED TRADER "x" is assumed to know the MARKET MAKER's price-setting function – $G_r(\cdot)$. The INFORMED TRADER maximizes her expected profits that she earns by holding the asset from time 1 to time 2.

$$X_r(d) = \arg\max_x \mathrm{E}\left[\underbrace{x \cdot [d - G_r(x + z + w)]}_{\text{gross profits}} \underbrace{+rx^-}_{\text{borrowing costs}} \middle| d\right] \tag{6}$$

3.3 Choice of commission rate

If the LENDER has the discretion to set the commission rate r, then I assume that this decision has to be made before the private signal d is revealed to the INFORMED TRADER "x". Irrespective of whether setting the commission rate is in LENDER's discretion, the value of r is non-random from the INFORMED TRADER's perspective.

On the other hand, strategy of the INFORMED TRADER $x_r(\cdot)$ depends on the commission rate in equilibrium, as I explain in detail. In equilibrium, the expected quantity of the asset borrowed by the INFORMED TRADER "x" is inversely related to the rate of commission r:

$$\frac{\partial\,\mathrm{E}[|X_r^-(d)|]}{\partial r} < 0.$$

There are two reasons for this inverse relationship. First, the bigger the rate of the commission r the more rarely the order size of the INFORMED TRADER "x" turns negative and the INFORMED TRADER eventually borrows the asset. Second, when the borrowing does happen, the quantity of the asset borrowed is inversely related to the size of the commission rate r, that is the commission tames the INFORMED TRADER's borrowing appetite.

In this section and in sections 4 and 5, the behavior of the UNINFORMED BORROWER "w" is entirely exogenous. In such a setting, an increase of commission rate r does not affect the UNINFORMED BORROWER's borrowing appetite: $\frac{\partial\,\mathrm{E}[|w^-|]}{\partial r} = 0$. Obviously, under this condition the LENDER has an incentive to set an extremely high commission to rip-off the UNINFORMED BORROWER "w" by charging extremely high rate r.

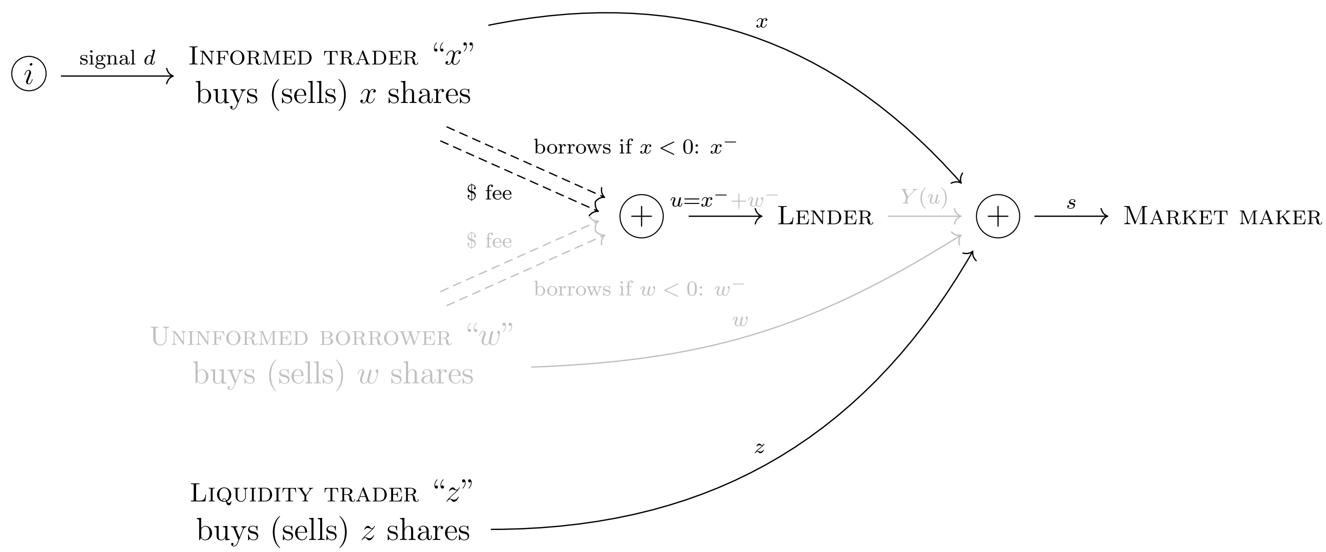
In order to address this phenomenon in this section, I simply exclude the UNINFORMED BORROWER "w" from the model. Only under this condition, I can expect to find an optimal commission rate r.[7]

The following diagram (figure 3) emphasizes that I exclude the UNINFORMED BORROWER "w" from the model in order to analyze the question of optimal commission rate r that a

[7]

monopolistic LENDER chooses to set.

Figure 3: Diagram of the model of commission without the UNINFORMED BORROWER "w"



Once I exclude the UNINFORMED BORROWER "w", the LENDER faces a sensible trade-off between the quantity of the asset borrowed by the INFORMED TRADER "x" and the size of the commission per unit of the asset. If the LENDER is a monopolist, then she solves the trade-off by maximizing

$$r \cdot \mathrm{E}[|X_r^-(d)|], \text{ where}$$

the first factor increases in r, while the second factor decreases in r.

If the LENDER sets the commission rate r too low, then the amount of borrowing remains finite while the expected commission revenues tend to zero. Instead, if the commission rate r is too high, then the expected borrowing volume (a.k.a. short interest) becomes extremely low, and, again, the LENDER underearns in terms of expected commission revenues.[8]

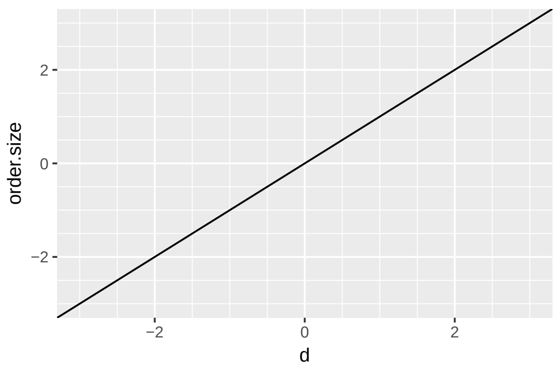
For each given value of commission rate r, I employ an iterative numerical algorithm to find the INFORMED TRADER's equilibrium trading strategy and the MARKET MAKER's equilibrium price-setting strategy. In the numerical algorithm, I use discrete approximations of random variables d and z as I describe in detail in Internet Appendix.

I show equilibrium trading strategies of the INFORMED TRADER "x" and for various exogenous commission rates in Figure 4.

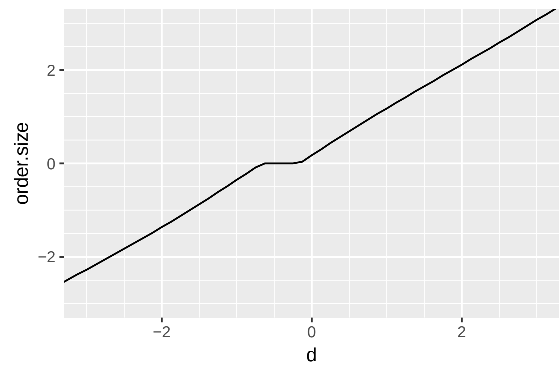
Additionally, I show the expected profits of the the LENDER as a function of commission rate in Figure 5.

[8]In this case, an excessive rate of the commission leads not only to the expected volume of trading converging to zero but also to commission revenues tending to zero. This is because as $r \to \infty$, the expected volume of borrowing $\mathrm{E}\left[-X_r(d)^-\right]$ tends to zero faster than $\frac{1}{r}$.

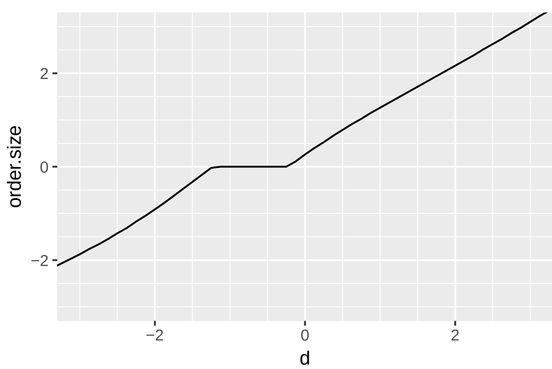
Figure 4: Equilibrium trading strategies of the INFORMED TRADER "x" for various exogenous commission rates r



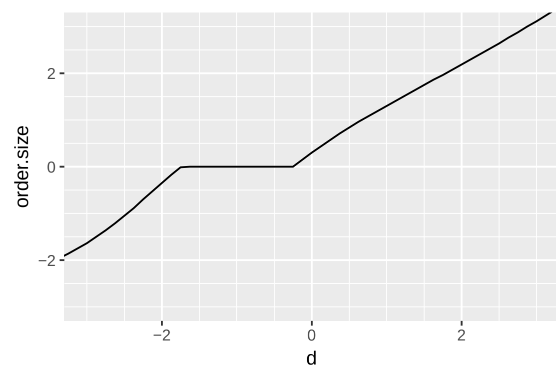
(a) Commission rate $r = 0$



(b) Commission rate $r = 0.5$



(c) Commission rate $r = 1$



(d) Commission rate $r = 1.5$

Figure 4 demonstrates that the commission discourages the INFORMED TRADER "x" from short-selling.

The mechanism is rather straightforward. If the INFORMED TRADER expects a price drop, but if its magnitude does not exceed the commission rate, the INFORMED TRADER "x" does not short-sell. The higher the commission rate is, the more likely it is that the INFORMED TRADER "x" eventually abstains from short-selling and the wider the range of values d for which this happens. In other words, the "non-trade region" widens as r increases.

Finally, I consider a monopolistic LENDER and analyze her optimal behavior in terms of choosing the commission rate r that results in the highest possible expected commission revenues. This problem parallels one of maximizing tax revenues by a government. To solve this problem graphically I plot commission revenues as a function of r: $r \mapsto r \cdot \mathrm{E}(|X_r^-(d)|)$, see Figure 5.

Figure 5: Expected profits of the LENDER as a function of commission rate r

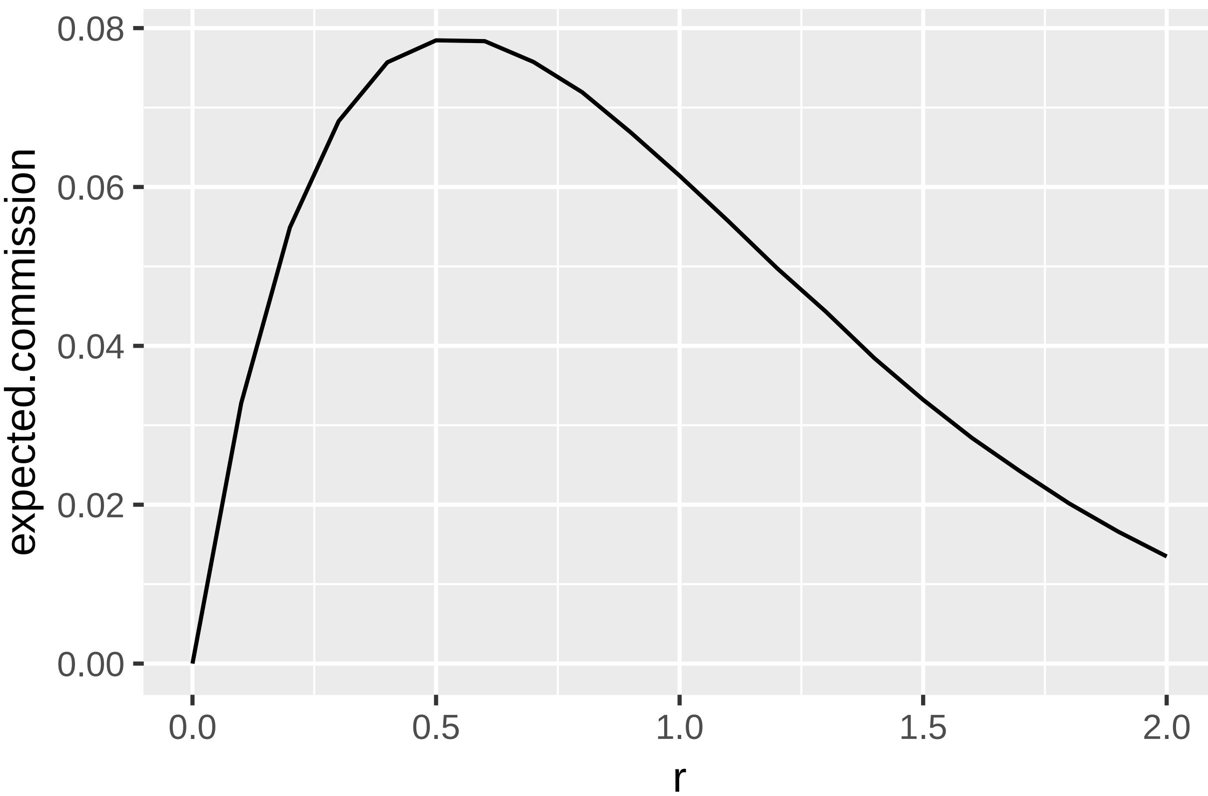


Figure 5 is a finance counterpart of Laffer curve – a tool that allows governments to analyze a trade-off between the tax rate and the magnitude of economic activity and to choose an economically efficient tax rate.

In Appendix A I additionally consider a more general setting. I analyze the possibility that the commission rate can be dependent on the quantity borrowed. By setting a variable commission rate, the LENDER can potentially discriminate between borrowers with different realizations of private signal d. Having more flexibility in setting the rate schedule, the LENDER could potentially extract higher rent from the INFORMED TRADER "x".

In economic literature, the concept of price discrimination has been shown to be an important one and beneficial for those who can exercise their market power, especially in the case of monopolistic or oligopolistic settings. Naturally, discrimination can potentially result in higher expected earning of the LENDER.

The essence and potential advantages of discriminatory commission rate can be explained as follows. The LENDER may choose to "twist" the commission rate schedule while keeping the commission rate for average borrowed quantity unchanged. Suppose, she decides to charge a higher commission rate for large quantities and lower the rate for small quantities. By doing so, the LENDER would effectively play a screening game with potential borrowers. As a result, the expected rate for a borrower with a highly negative signal would increase while the expected rate for a borrower with a moderately negative signal would decrease. Since the borrowers with a highly negative signal tend to borrow more, this would result in a higher average commission rate weighted by quantity so long as the behavior of potential borrowers does not change. So, the effect on the expected commission revenue would be *positive*. Nevertheless, the increasing schedule of commission rates creates an extra incentive to lower the borrowing quantity, and thus has a *negative* effect on expected commission revenues.

If the positive effect dominates, the "twist" would allow the LENDER to boost commission revenues by adopting a "progressive" commission rate: the higher the quantity, the higher the rate. Conversely, if negative effect dominates, then the opposite of "twist" would benefit the LENDER. In Appendix A I show that in my setting the LENDER optimally chooses to retain a flat commission schedule. In the context of the analysis provided above, the reason for such a choice is that the two effects approximately compensate each other. For each given value of commission rate r, I employ an iterative numerical algorithm to find the INFORMED TRADER's equilibrium trading strategy and the MARKET MAKER's equilibrium price-setting strategy. In the numerical algorithm, I use discrete approximations of random variables d and z as I describe in detail in Internet Appendix.

I show equilibrium trading strategies of the INFORMED TRADER "x" and for various exogenous commission rates in Figure 4.

3.4 Case of exogenous commission rate

Previously, in Figure 3 I excluded the UNINFORMED BORROWER "w" in order to case a sensible maximization problem for a monopolistic LENDER.

If the commission rate is given exogenously, dropping the UNINFORMED BORROWER "w" is no longer necessary. The model remains tractable and sensible in the presence of both the UNINFORMED BORROWER "w" and the LIQUIDITY TRADER "z". Moreover, if the rate r is exogenous, the UNINFORMED BORROWER "w" is mathematically indistinguishable from the LIQUIDITY TRADER "z": the INFORMED TRADER "x" does not directly interact with either of them, while from the perspective of the MARKET MAKER, they are indistinguishable by assumption.

As a result, the solution presented in Figure 4 can be also interpreted as a solution the model descried in Figure 2, where both INFORMED TRADER "x" and UNINFORMED BORROWER "w" are present. As UNINFORMED BORROWER "w" is indistinguishable from LIQUIDITY TRADER "z", the presence of UNINFORMED BORROWER "w" will be equivalent to just increasing $\operatorname{Var}(z)$: the implication for the equilibrium trading strategies is a proportional scale-up of INFORMED TRADER's trading strategies by the coefficient $\sqrt{\operatorname{Var}(z+w)/\operatorname{Var}(z)} = \sqrt{1+\sigma_w^2/\sigma_z^2}$.

3.5 Market efficiency

In this section, I highlight how borrowing fees influence price efficiency. In particular, to measure the price efficiency, I use the following measure: the accuracy of the forecast of d given the equilibrium price p_1. Quantitatively, I use conditional variance $\mathrm{Var}(d|p_1)$, which is depicted Figure 6.

The overall effect of the commissions is a decrease in the price efficiency: the higher the commission rate r, the higher is the $\mathrm{E}(\mathrm{Var}(d|p)) = V(d - p_1)$. The overall effect of the borrowing commissions is as expected, the commission discourages the INFORMED TRADER "x" from revealing the information. As a result, for a sizable commission (say, $r = 2$) the behavior of the INFORMED TRADER "x" for $d = -1$ and $d = -2$ is absolutely identical. The higher the commission rate, the wider is the non-trading region, which makes the market opaque.

Paradoxically, it is specifically in the opaque area the conditional variance $\mathrm{E}(d|p_1)$ is relatively small. To explain this phenomenon, I consider a limiting case: $r = +\infty$, equivalent to the case of explicit short-sale ban. I assume that $d \sim N(0, 1)$. Equilibrium trading strategy for $d > d_0$ is positive and increasing. For $d \leq d_0$, $X(d) = 0$: no-trade region is $(-\infty, d_0]$, where $d_0 < 0$.

Suppose, the MARKET MAKER observes the $s = X(d) + z = \hat{s} \ll 0$ – a highly negative order flow. Under the assumption, the z is normal, the Bayesian updating rule makes the MARKET MAKER think that $X(d) > 0$ is virtually impossible.[9] The MARKET MAKER cannot judge what exactly is the value of d within $(-\infty, d_0]$. Eventually, the conditional distribution $d|s = \hat{s}$ (or, equivalently, $d|p_1 = G(\hat{s})$) is approximately truncated normal with the support $(-\infty, d_0]$. The numerical computations show that $d_0 \approx -\frac{1}{4}$. It is possible to show that the conditional variance of the truncated standard normal between $-\infty$ and $-\frac{1}{4}$ is roughly 0.31.[10] This means that $\mathrm{Var}(d|p_1 = G(\hat{s})) \approx 0.31$. Instead, with $r = 0$, the model boils down to the standard Kyle model. In this case, the conditional distribution is as follows:

$$d|p_1 \sim N(p_1, 0.5).$$

Paradoxically, when the MARKET MAKER observes an informative negative signal, the conditional variance is higher 0.5 than in the case of uninformative negative signal 0.31. In other words, an "uninformed" MARKET MAKER faces less uncertainty than an "informed" one. In case $r = 0$ highly negative order flow values $\hat{s}$ can drive the expectation $\mathrm{E}(d|s = \hat{s})$ to highly negative values, thereby transmitting information to the MARKET MAKER; nevertheless the uncertainty about d still remains considerable: $V(d|s = \hat{s}) = \frac{1}{2}$. With $r = +\infty$, instead, even

[9] The rate of decay of the normal distribution increases in the tails because the log of density is quadratic.

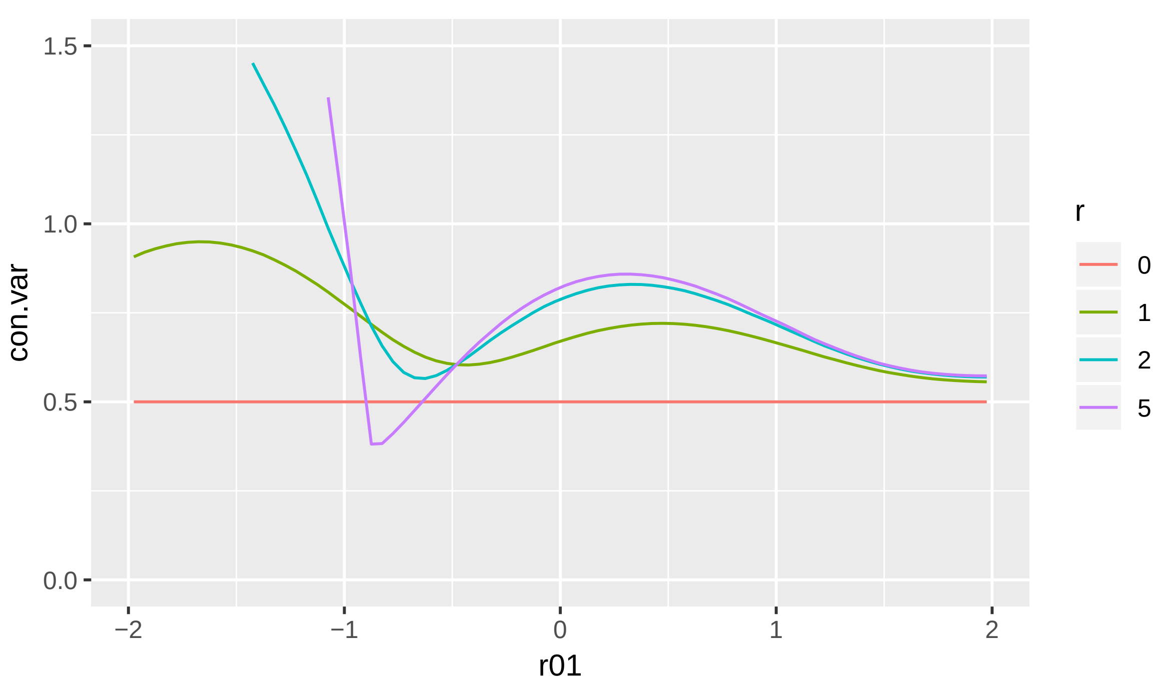
[10] A general formula is $1 - \beta\frac{\phi(\beta)}{\Phi(\beta)} - \left(\frac{\phi(\beta)}{\Phi(\beta)}\right)^2$, where β is the right truncation threshold (left one is $-\infty$).

extremely negative $\hat{s}$ only drives the expectation $\mathrm{E}(d|s = \hat{s})$ to the level $\mathrm{E}(d|d < d_0) \approx -0.96$, at most.

I already mentioned that the borrowing fees boost the average variance $\mathrm{E}(\mathrm{Var}(d|p_1))$. This effect is due to the fact that they particularly push up the conditional variance for p_1 in the region between for values of p_1 around 0.5. The reason is that in this case effect is as follows. The MARKET MAKER finds it probable that d may be within $(0, 1)$, but it can also be around zero or highly negative because negative values of d are not reflected in the order flow $s = X(d) + z$ as a result of high borrowing fees.

If the commission is finite, then $\mathrm{Var}(d|p_1)$ also exhibits high values for $p_1 < \mathrm{E}(d|d < d_0)$. This case corresponds to the possibility of short-selling. Short-selling with positive commission sends a strong signal about highly negative d; therefore, if the MARKET MAKER suspects that there might have been a short sale but is not certain about it, this creates high uncertainty about the d. Condition p_1 slightly less then $\mathrm{E}(d|d < d_0)$ precisely describes this case. This effect is particularly strong for high commission rates (say $r = 5$) in terms of the values of $V(d|p_1)$; however with high commission rates the p_1 goes below $\mathrm{E}(d|d < d_0)$ extremely rarely.

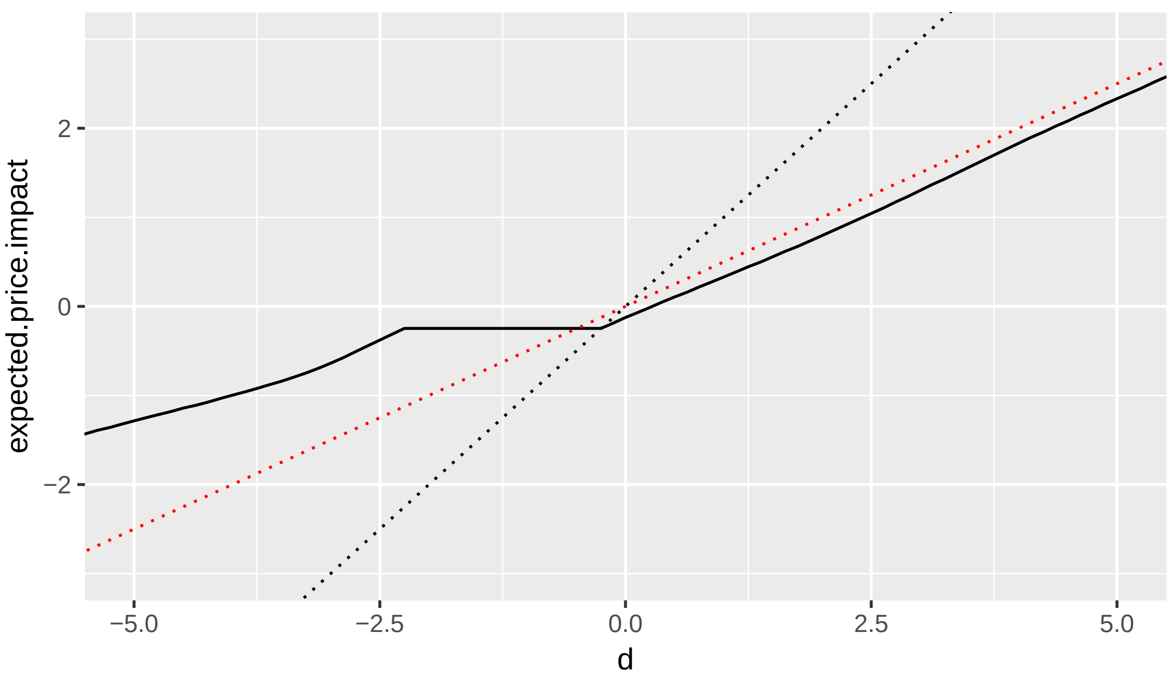
Figure 6: Price discovery – conditional variance of the fundamental value $\mathrm{Var}(d|p_1)$
In the standard Kyle model $(r = 0)$ $\mathrm{Var}(d|p_1) = \frac{1}{2}\sigma_d^2 = 0.5$.



In addition, I measure to what extent the private signal d is impounded in the market price p_1 by computing $\mathrm{E}(p_1|d)$ and visually assessing of how far it is from d. Figure 7 demonstrates that in the model with front-running and commission, the private signal d is reflected in price to a lesser extent than in the standard Kyle model: $\mathrm{E}(p_1|d)$ is farther from d than in the standard Kyle model.

It is also noteworthy that $\mathrm{E}(p_1|d = 0)$ is below zero. This effect is due to the non-trade region. Because of the commission, the INFORMED TRADER "x" abstains from trading or reduces the magnitude of sell orders. This makes the case $d = 0$ more similar to $d = -\frac{1}{2}$ than to $d = \frac{1}{2}$ from the perspective of the MARKET MAKER. This effect is described in detail in Pankratov (2018).

Figure 7: Expected price from the INFORMED TRADER's point of view $\mathrm{E}(p_1|d)$ for $r = 2$.
In the standard Kyle model (red dotted line) $\mathrm{E}(p_1|d) = \frac{1}{2}d$.
Black dotted line represents depicts $\mathrm{E}(p_1|d) = d$ – 45°-line.



4 Model of front-running[11]

In this section, I introduce another way for the LENDER to take advantage of the INFORMED TRADER "x". In particular, instead of charging a commission for lending the asset to the INFORMED TRADER "x", the LENDER can learn about the fundamental value of the asset from

[11] As I mentioned in the introduction, I use the term "front-running" in a broad sense to highlight the attempt of the security lender in possession of trade-related information to use this information for her profits. This does not mean, however, that in my model, the asset lenders predict anyone's future trading needs.

the quantity that the market participants borrow from the LENDER and use this information for trading.

Note that in this model, borrowing happens before trading. For that reason, once the LENDER learns some information about the fundamental value of the asset, she can take advantage of this information by trading on her own account.

The signal, that the LENDER observes is given by:

$$U(d,w) = X^{-}(d) + w^{-} \equiv \min(0, X(d)) + \min(0, w). \tag{7}$$

After the LENDER observes the size of the short position, she decides on the size and direction of her proprietary trade:

$$Y(\hat{u}) = \arg\max_{y} \left\{ \mathrm{E}\left[\Pi^{\text{LENDER}}_{X,G}(y,d,w,z) \middle| u = \hat{u} \right] \right\} \text{ where} \tag{8}$$

$$\Pi^{\text{LENDER}}_{X,G}(y,d,w,z) = y \cdot [d - G(X(d) + y + z + w)] \tag{9}$$

In the model of front-running on short-selling information, the MARKET MAKER's inference is done in the following way. Once the MARKET MAKER observes the aggregate order flow from the INFORMED TRADER "x", the LENDER, the LIQUIDITY TRADER "z" and, the UNINFORMED BORROWER "w", she updates her beliefs about the liquidation value of the asset d and sets the price accordingly:

$$p_1 = G(\hat{s}) \equiv \mathrm{E}\left(d \middle| X(d) + w + Y(X^{-}(d) + w^{-}) + z = \hat{s} \right). \tag{10}$$

Definition 4.1. Equilibrium with front-running is a triple of functions: $X(\cdot)$, $Y(\cdot)$, and $G(\cdot)$ such that the following conditions are met:

1. $X(\cdot)$ is an optimal strategy of the INFORMED TRADER "x":

$$X\left(\hat{d}\right) = \arg\max_{x} \left\{ x \cdot \underbrace{\left(\hat{d} - \mathrm{E}\left[G(x + w + Y(x^{-} + w^{-}) + z) \right] \right)}_{\text{expected profit per unit of asset}} \right\}, \text{ where} \tag{11}$$

   the INFORMED TRADER "x" takes strategies $Y(\cdot)$ and $G(\cdot)$ as given.

2. $Y(\cdot)$ is an optimal strategy to take advantage of the information available to the LENDER (signal u). In particular, this means that Y has to satisfy Eq. 8. The LENDER takes strategies $X(\cdot)$ and $G(\cdot)$ as given.

3. The price setting rule that the MARKET MAKER follows ($G(\cdot)$) is consistent with the

trading strategies of the INFORMED TRADER "x" and the LENDER. In particular, this means that G satisfies the Eq. 10. The MARKET MAKER takes strategies $X(\cdot)$ and $Y(\cdot)$ as given.

4.1 Case without uninformed borrowers (no equilibrium)

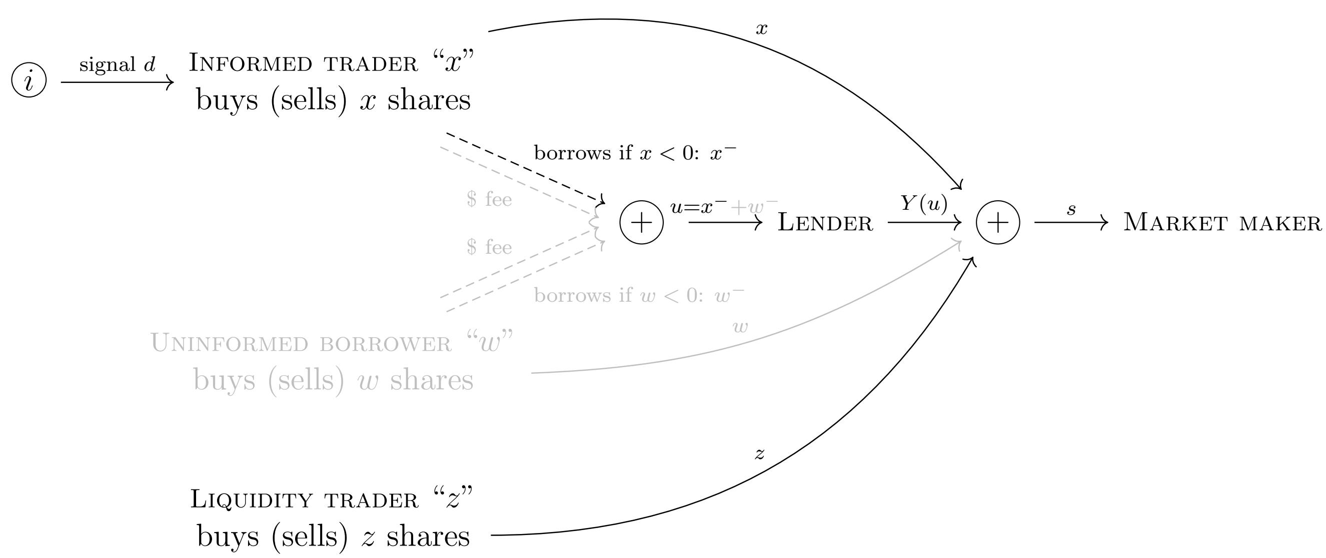
For a start, I consider a simplified case – without the UNINFORMED BORROWER "w". Excluding the UNINFORMED BORROWER "w" from the model considerably simplifies the inference process conducted by the LENDER. In particular, the signal u, observed by the LENDER is given by,

$$u = X^{-}(d) \equiv \min(0, X(d)) \leq 0, \text{ and therefore} \tag{12}$$

it becomes very easy for the LENDER to make inferences about d.

The following diagram (Figure 8) summarizes the variant of the model discussed in this subsection, and highlights the fact that the UNINFORMED BORROWER "w" is excluded from the model, and therefore the LENDER's signal depends solely on x.

Figure 8: Diagram of the model front-running without the UNINFORMED BORROWER "w"



I conduct numerical simulation of agents' behavior in an attempt to find an equilibrium. The technical details of the simulation are presented in Internet Appendix. The outcome of the simulation is that the trading strategies of the agents diverge. The divergence suggests that in the model where the LENDER can perfectly observe the size of negative orders of the INFORMED TRADER "x", may not have an equilibrium.

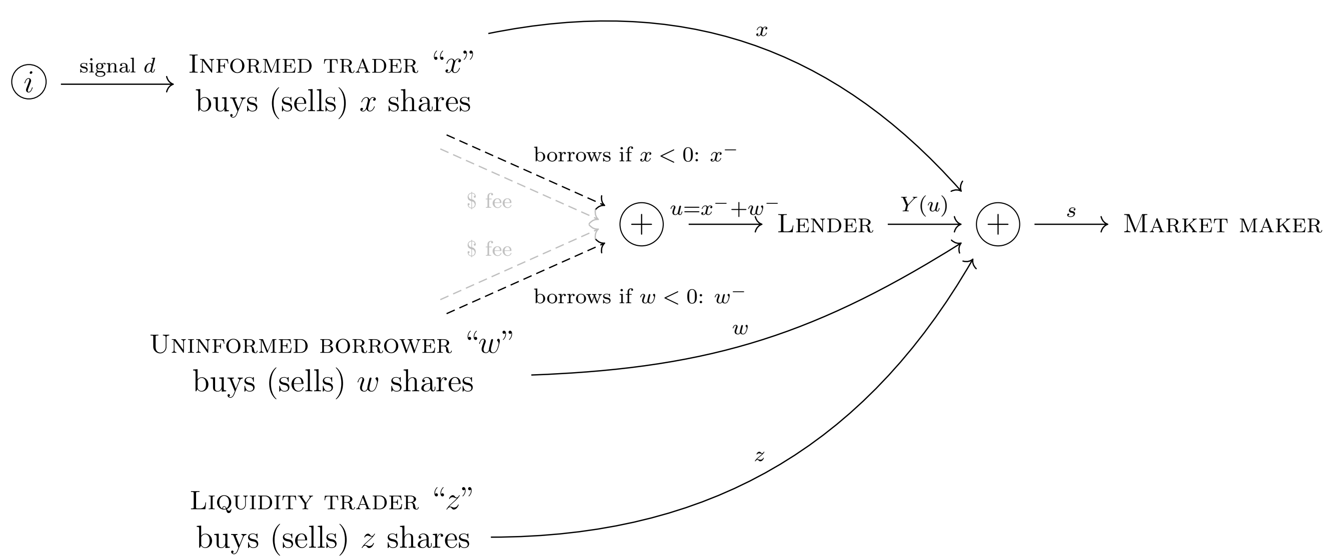
Additionally, in Appendix B, I show that in a similar model but with perfect LENDER's observability for both buy and sell orders, equilibrium does not exist. In the appendix, I assume that instead of $u = x^- + w^-$, the LENDER observes $u = x + w$. This allows me to retain the essence of the "leader-follower" essence of the model and at the same time eliminate the asymmetry between $d > 0$ and $d < 0$. I show in the appendix, that the symmetric model has a unique linear equilibrium as long as $\sigma_w > 0$. The equilibrium is known in closed form and it is possible to show that it collapses as $\sigma_w \to 0$.[12] See the appendix for the details. The absence of equilibrium in the linear version considered in the appendix suggests that the same outcome in the non-linear model is absolutely expectable.

In the model with front-running without the UNINFORMED BORROWER "w" equilibrium does not seem to exist. In order for the model to have a finite equilibrium, the existence of the noise in the volume borrowed is essential. For that reason, I provide a numeric solution of a model with $\text{Var}(w) > 0$ in the next subsection.

4.2 Case with uninformed borrowers

As I have shown, equilibrium does not seem to exist in the setting without the UNINFORMED BORROWER "w". For this reason, I consider the case $\text{Var}(w) > 0$. The following diagram (Figure 9) summarizes the variant of the model discussed in this subsection, and highlights the fact that the LENDER is not allowed to charge a commission for lending.

Figure 9: Diagram of the model front-running



This case poses a lot of challenges with respect to solving the model. The main reasons

[12] Aggressiveness of the INFORMED TRADER "x" tends to zero, while aggressiveness of the LENDER tends to infinity.

are as follows:

1. The model is asymmetric and does not have linear equilibria.

2. The decisions made by the LENDER are made with the knowledge

$$U(d, w) \equiv X^{-}(d) + w^{-} = \hat{u},$$

   therefore the objective function of the LENDER is a conditional expectation whose evaluation requires tedious computations.

3. Bayesian updating performed by the MARKET MAKER requires numeric integration, and I have to employ certain computational tricks to make this integration feasible and fast.

I summarize my numerical methodology in Appendix E with references to other details provided in other appendices. In the body of the paper instead, I focus on the economic implications and analysis of the model. Regarding technical details, in this section, I limit myself to merely revealing that I look for approximate equilibria. In particular, it means that I constrain each of the agents (the INFORMED TRADER, the LENDER, and the MARKET MAKER) to follow strategies from certain families of functions: $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, and $G \in \mathcal{G}$.

I normalize parameters σ_d and σ_z to 1, and I vary σ_w. The equilibrium strategies of each one of the players constitute univariate functions. I present them as plots in Figures 10, 11, 12, and 13.

Figure 10 demonstrates that the possibility of front-running indeed discourages the INFORMED TRADER "x" from short-selling. The aggressiveness of short-selling is always lower relative to the aggressiveness of buying. The disincentive is more severe when the INFORMED TRADER's choices are more exposed to the LENDER, that is when the UNINFORMED BORROWER "w" is less active (low σ_w). In particular, the disincentive to short-sell comes from the fact that bigger short sales trigger bigger front-running sales that in turn create extra downward price pressure. The INFORMED TRADER "x" is aware of the additional price impact, as a result, she internalizes this effect and rations the amount of short-selling.

The smaller is the activity of UNINFORMED BORROWER "w", the heavier is the burden of front-running, and the higher is the disincentive to short-sell. As I send σ_w to 0, the informed short-selling will also shrink to 0. The INFORMED TRADER "x" is no longer able to borrow without being exposed. Without the UNINFORMED BORROWER "w" borrowing market becomes a market for "lemons" (Akerlof, 1970).

For the case of purchases, the direction of the effect is the same: reduction in the random activity of the UNINFORMED BORROWER "w" drives down the purchasing activity in response to good signal realizations. Nevertheless, in the positive case, the front-running the effect is

Figure 10: INFORMED TRADER's order size as a function of private signal

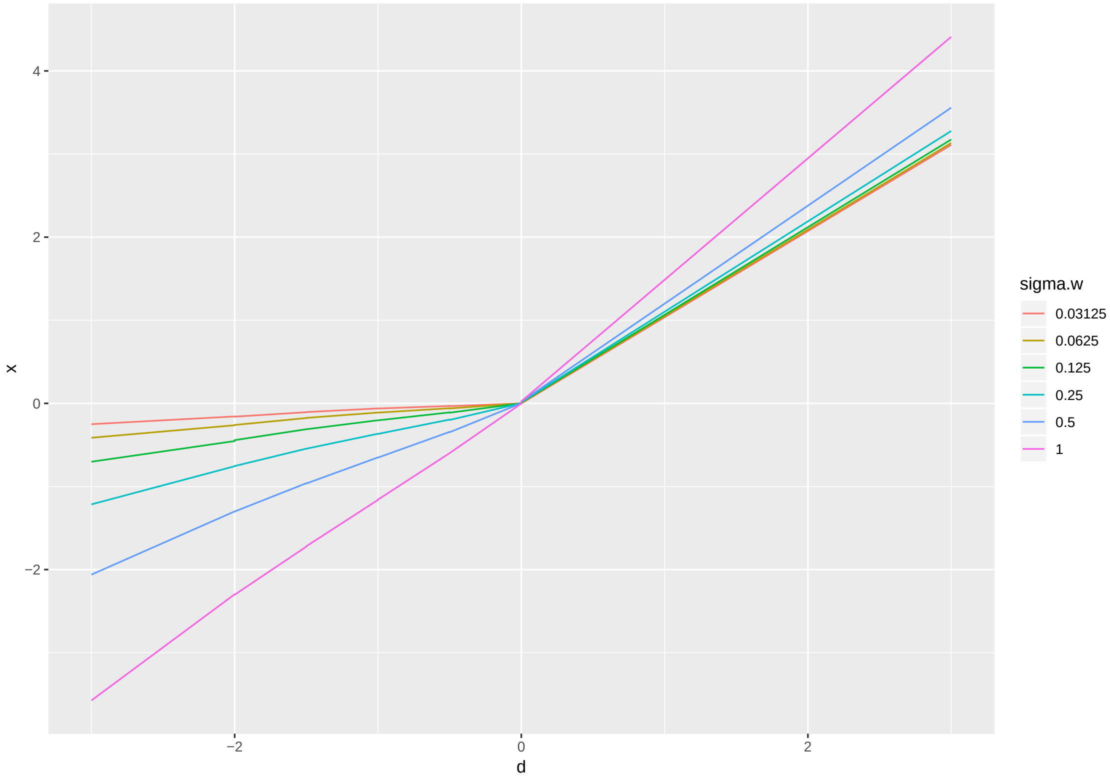


Figure 11: Front-running: LENDER's proprietary order size as a function of short interest

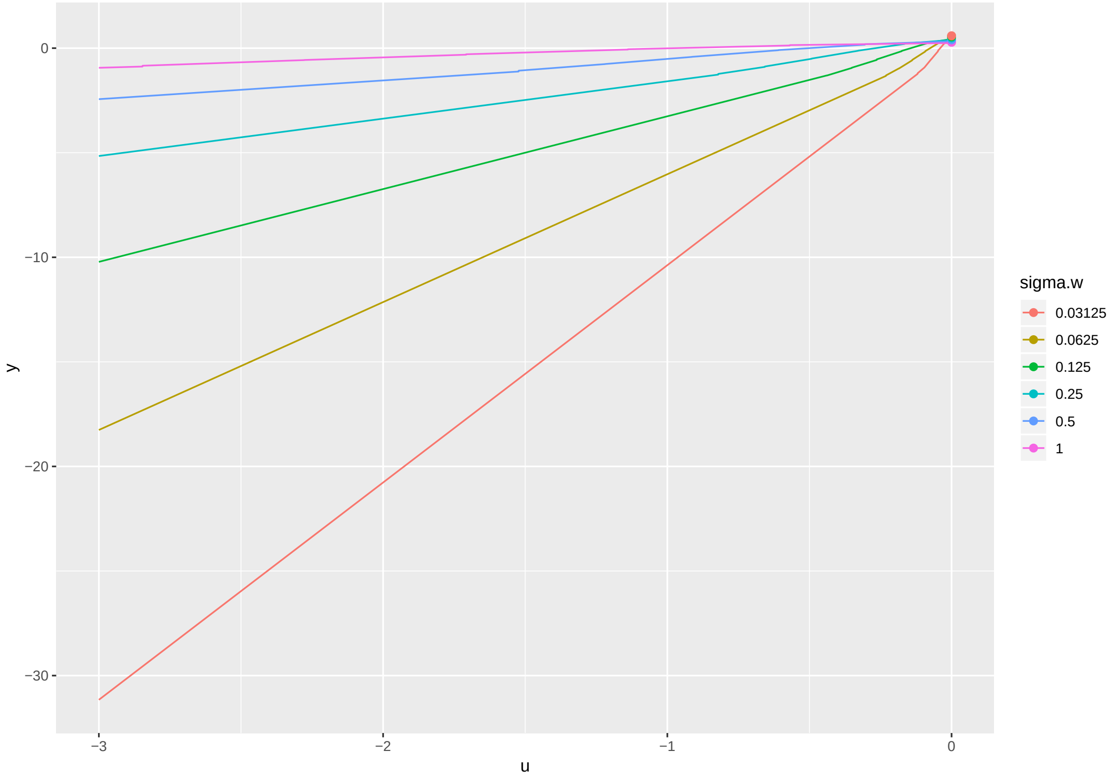


Figure 12: Front-running: LENDER's proprietary order size as a function of short interest: zoomed on the area around $u = 0$

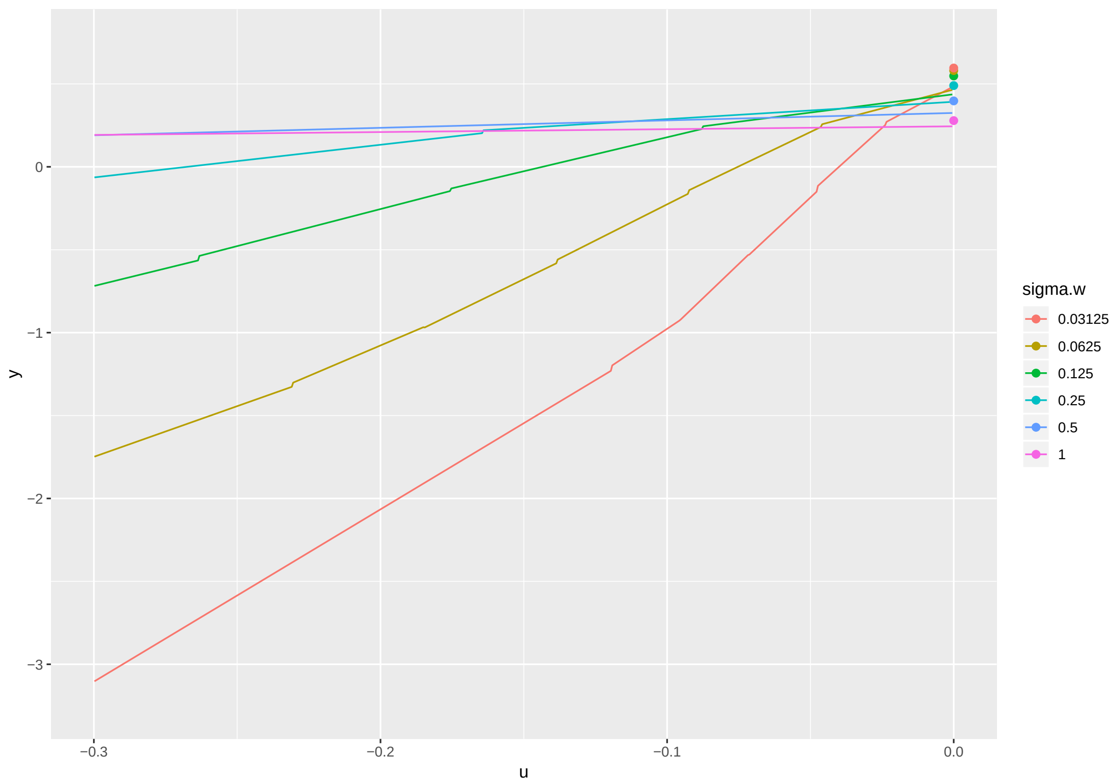
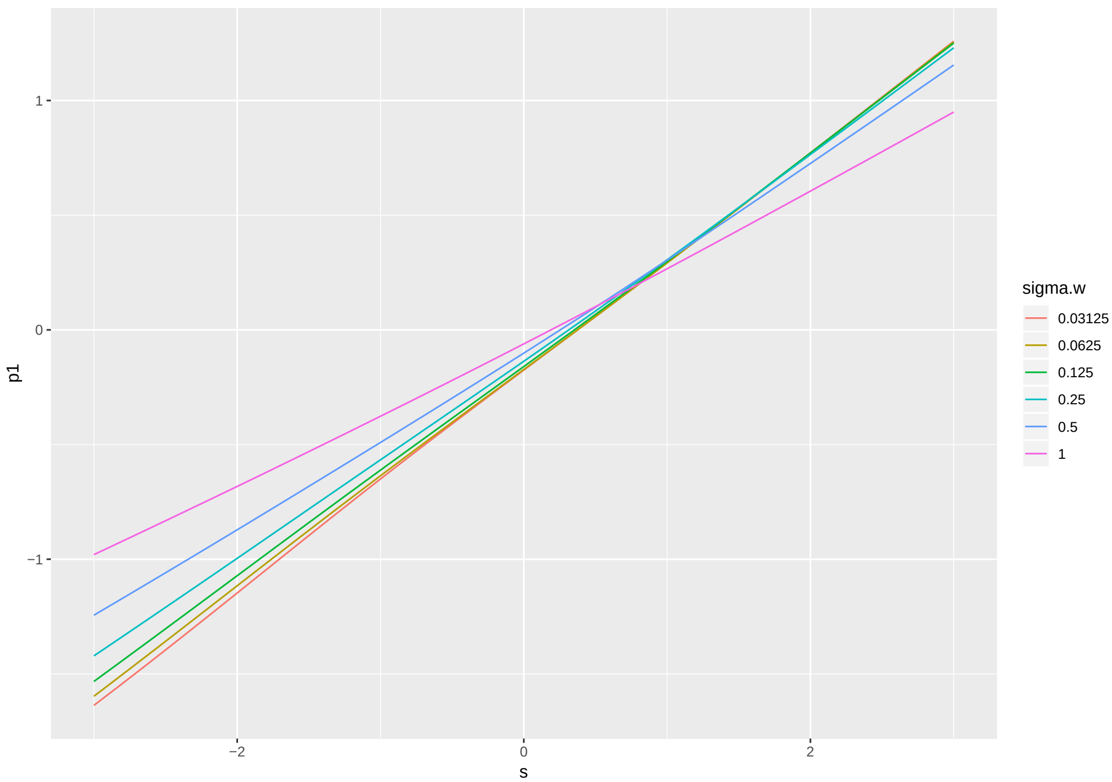


Figure 13: MARKET MAKER's price setting rule in equilibrium $p_1 = G(s), s = x + w + y + z$



much weaker. There is no explicit disincentive to hide the purchasing order from the LENDER: even though, absence of short-selling signals positive payoff to the LENDER, the decision of the LENDER cannot possibly depend on the size of the INFORMED TRADER's buy order since the order size is not observable by the LENDER, as implied by the structure of the LENDER's signal $U(d, w) = X^-(d) + w^-$.

Lowering σ_w causes the moderation of the INFORMED TRADER's buying aggressiveness simply because the UNINFORMED BORROWER "w" contributes to overall order flow

$$s = x + w + Y(x^- + w^-) + z.$$

When σ_w shrinks, the signals sent by the INFORMED TRADER "x" become less protected from the MARKET MAKER, and therefore the INFORMED TRADER "x" decides to be more cautious. It is remarkable that indeed in the positive region of d the effect is weaker: as opposed to the negative case, the activity of the INFORMED TRADER "x" does *not* tend to 0 as σ_w fades out.

Regarding the front-running, the activity itself, Figure 11 brings out the following stylized phenomena. First, the response of the LENDER to the "negative short interest" $u = x^- + w^-$ is always a monotonically increasing function: the more agents short-sell, the more the LENDER sells (or buys less). Note that the absence of short sales ($u = 0$) sends a positive signal to the LENDER.

The overall aggressiveness of front-running increases as $\mathrm{Var}(w)$ shrinks. In particular, this means that in the case $u \ll 0$, the LENDER sells more units of the asset in equilibrium. The reasons for this are as follows. First, as the UNINFORMED BORROWER "w" becomes less active, the LENDER is better able to infer d from the negative short interest u in case the INFORMED TRADER "x" does borrow. Consequently, the LENDER is better able to mimic the INFORMED TRADER "x" without incurring losses due to inevitable copying of w^-. Second, as the INFORMED TRADER "x" becomes increasingly cautious, the LENDER's interpretation of $u = x^- + w^-$ becomes more responsive to compensate for INFORMED TRADER's cautiousness. In other words, even a small move made by the INFORMED TRADER "x" means a lot from the perspective of the LENDER.

Additionally, it is remarkable that small negative values of u (small short interest) trigger purchases on the part of the LENDER. Even though the LENDER observes someone borrowing the asset, this actually sends a positive signal to the LENDER. This happens because in this case $\mathrm{E}(d|u) > 0$.

The mechanism behind is the following. A small negative value $u = \hat{u}$ reveals one of the 3 possibilities:

1. $x^- = \hat{u} < 0$, $w^- = 0$,
2. $x^- = 0$, $w^- = \hat{u} < 0$,
3. $x^- < 0$, $w^- < 0$, $x + w = \hat{u}$, where

the first case eventually means that with a finite probability d is a small negative number; the second case means that with finite probability d ranges in a set of values where $x(d) \geq 0$ with a density proportional to the unconditional density of $d : N(0, \sigma_d^2)$; and the third case is relatively unlikely and negligible. Overall, the first case tilts the expectation of d upwards, while the second case tilts the expectation downwards but the effect coming from the second case is stronger so long as $|u|$ is small enough. The details of bayesian updating $d|u$ are portrayed in Appendix F and depicted in Figure 31 within the appendix. This feature is consistent with the empirical finding of Boehmer et al. (2010) that small short interest predicts price increase.

A less obvious feature is that the LENDER's reaction to zero and small negative realizations of u is inversely related to $\mathrm{Var}(w)$, i.e. the less active the UNINFORMED BORROWER "w" is, the more the LENDER buys in response to observing $u = 0$ or $u = \hat{u} \to 0^-$. The mechanisms mentioned above do not explain such a behavior. This behavior has more intricate mechanics which is explained in Appendix D in detail.

4.2.1 Profits/losses breakdown

In this section, I analyze the competition between the INFORMED TRADER "x" and the LENDER. The split of the profits between the two agents is affected by the aggressiveness of the UNINFORMED BORROWER "w".

First, as the aggressiveness of the UNINFORMED BORROWER "w" shrinks, she submits smaller order sizes, naturally, her losses also shrink. Even though LIQUIDITY TRADER's trading decisions do not change, her expected losses increase. The main reason is that she becomes the only looser in the economy. Academically speaking, she faces more adverse selection when the relative weight of the UNINFORMED BORROWER "w" shrinks.

Overall, as the aggressiveness of the UNINFORMED BORROWER "w" shrinks, the aggregate losers' activity moderates. Naturally, the net effect is that the aggregate losses in the economy moderate as well. We are dealing with a zero-sum game. Thus, the two winners – the INFORMED TRADER "x" and the LENDER – share between them the prey from the two losers – UNINFORMED BORROWER "w" and LIQUIDITY TRADER "z".

As the aggressiveness of the UNINFORMED BORROWER "w" shrinks, the INFORMED TRADER "x" also becomes more cautious. Conversely, the LENDER becomes more informed and more aggressive. As a result, a higher proportion of the expected profits is allocated to the LENDER.

Figure 14 illustrates the overall expected profits/losses[13] in the economy, how the profits are shared between the INFORMED TRADER "x" and the LENDER (the upper part of the plot), and how the losses are shared between the UNINFORMED BORROWER "w" and the LIQUIDITY TRADER "z" (the lower part of the plot).

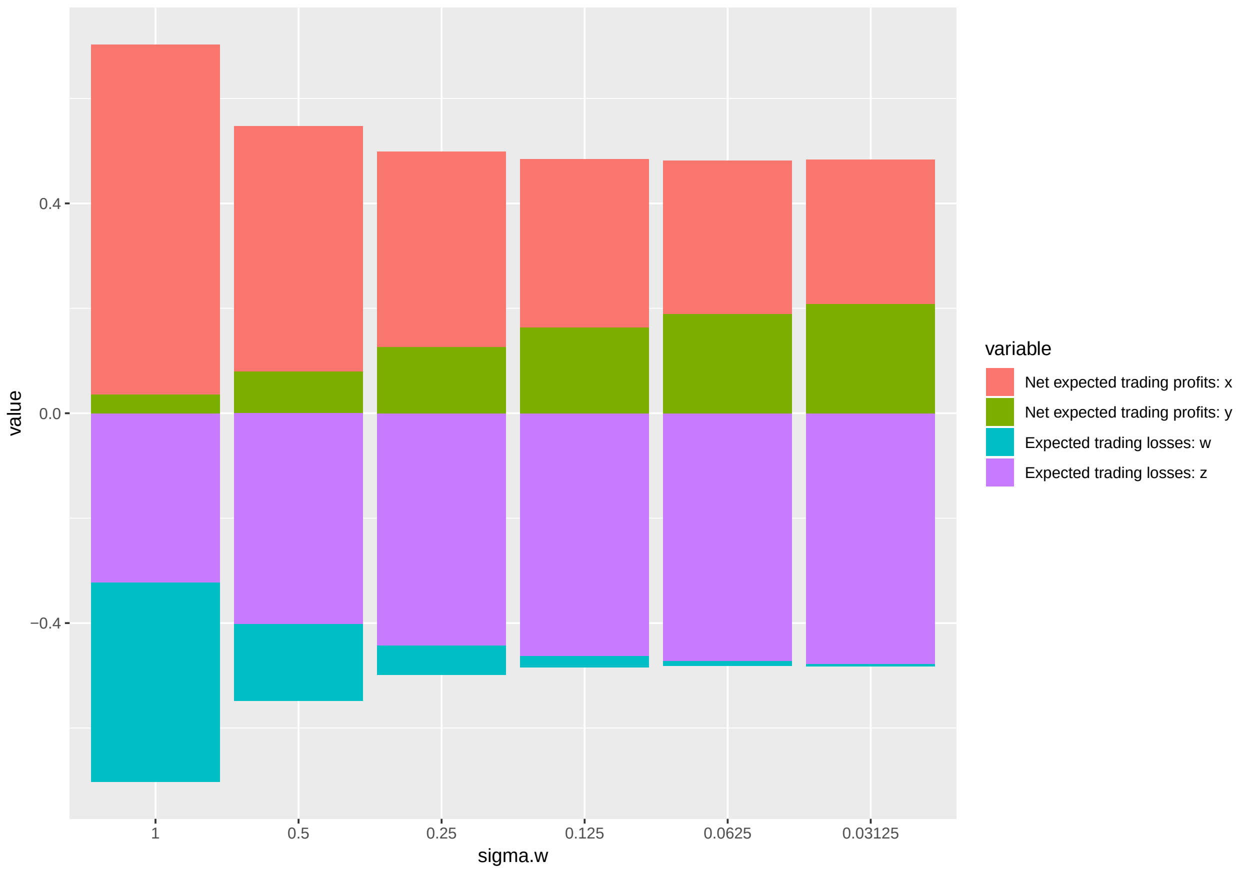
4.3 Market efficiency

In this subsection, I analyze market efficiency. First, I plot the price discovery – to what extent the market price allows to predict the fundamental value of the asset. Figure 15 reveals that in my model, the price is slightly less efficient than in the standard Kyle model in negative cases and slightly more efficient in the positive cases. The magnitude of these effects is directly related to the aggressiveness of UNINFORMED BORROWER. The lower price efficiency in the negative region comes from the noisier behavior of the LENDER. The decisions of the UNINFORMED BORROWER are copied by the LENDER only when $w < 0$. This makes prices noisier when they are negative.

The deviations of the conditional variance from the benchmark of the standard Kyle model are, in fact, quite paradoxical. It is important to point out that in a symmetric modification of the model presented here, the conditional variance is constant. It is exactly the same in

[13] Since my model is a zero-sum game, overall expected profits are equal to overall expected losses

Figure 14: Profit sharing between the INFORMED TRADER "x" and the LENDER, loss sharing between the UNINFORMED BORROWER "w" and the LIQUIDITY TRADER "z". The bars above zero depict the breakdown of profits. The bars below zero depict the breakdown of losses.



the standard Kyle model: $\mathrm{Var}(d|p_1) = \frac{1}{2}\sigma_d^2$.

Like in a preceding subsection 4.1, I refer to the same model described in Appendix B.[14] In the symmetric model, the noise created by the UNINFORMED BORROWER "w", is copied by the LENDERwith a linear coefficient. In Appendix B, I show that irrespective of the aggressiveness of the UNINFORMED BORROWER "w", in equilibrium, the price is as informative as in the standard Kyle model. Moreover, even if I set the LENDER's aggressiveness parameter exogenously, the level of price efficiency stays at the same level. Thus, the level of price efficiency is an intrinsic feature of the interaction between the INFORMED TRADER "x" and the MARKET MAKER. In the negative area ($d < 0$), my model is similar to the one presented in Appendix B, while in the positive region ($d > 0$), the is similar to the standard Kyle model. Both these models predict the same degree of inefficiency $\mathrm{Var}(d|p_1) = \frac{1}{2}\mathrm{Var}(d)$.

In the symmetric model with a "leader" (the INFORMED TRADER "x") and a "follower" (the LENDER) the extra noise created by the "follower" can be compensated by a more aggressive behavior of the "leader", and eventually, the signal-to-noise ratio remains the same, i.e., 1 : 1. It turns out that the deviations from the Kyle benchmark are due to the asymmetric nature of the noise component: the negative realizations of w (the UNINFORMED BORROWER "w") are reflected on the order flow s, while the positive realizations are not.

To prove the point that the deviations of market efficiency from the benchmark are due to the asymmetric nature of the noise and not due to the follower's choices, I provide another simplified model. I exclude the LENDER, and make the noise component z intentionally skewed. Let us suppose, $\tilde{z} \sim N(0,1)$ and define z in such a way that negative values have a more significant impact:

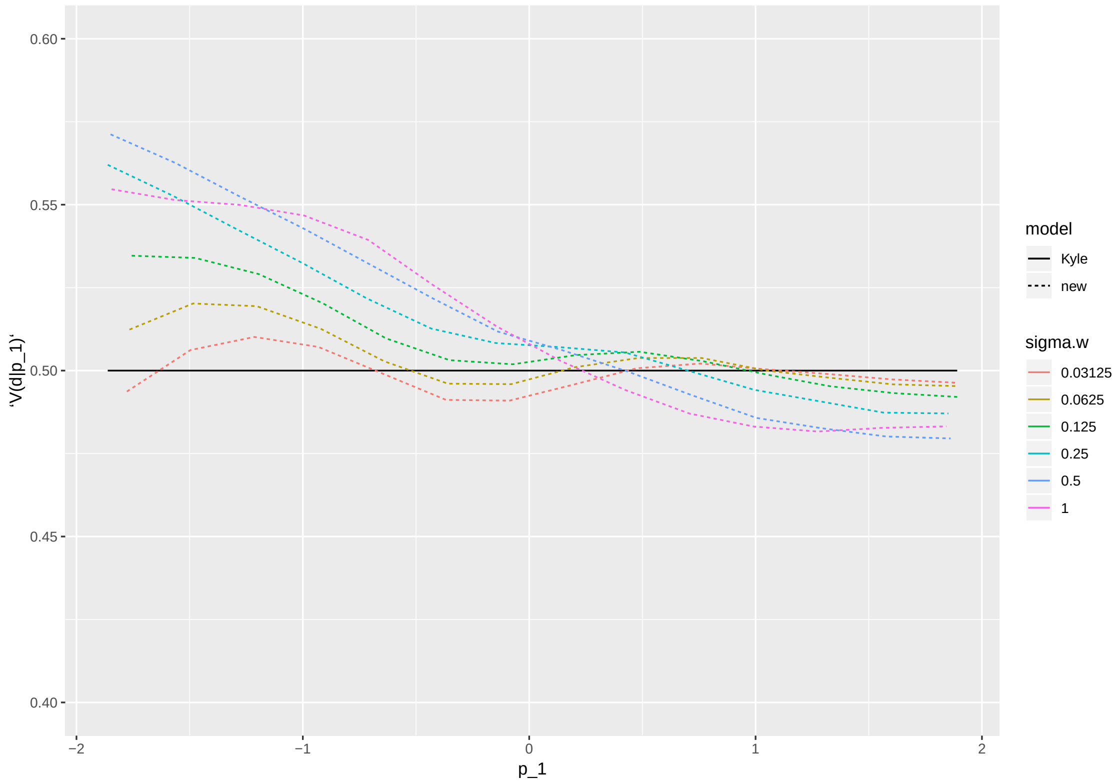
$$z = k\tilde{z}^- + \tilde{z}^+, \text{ where } k > 1.$$

Eventually, I end up having a standard Kyle model with non-Gaussian (left-skewed) noise. In Appendix C, I show that such a model features non-constant price efficiency $V(d|p_1)$. In the appendix, I explain in detail the mechanics behind this effect.The bottom line of the analysis provided in the appendix is that even though equilibrium in this model is similar to that in standard Kyle model, the trading aggressiveness in the buy side and sell side differs, and the mismatches in the buy side and sell side lead to the deviations from the benchmark market efficiency, as in the standard Kyle model.

Another main prediction is that the magnitude of the asymmetry in the price efficiency if commensurate the asymmetry in the distribution of the LIQUIDITY TRADER's order z. The magnitude of the effect in the main model – the model with "front-running" depends, therefore, on the extent to which the quantity of noise in the negative part exceeds that in the

[14]The LENDER observes $u = x + w$ instead of $u = x^- + w^-$. I retain the essence of the "leader-follower" essence of the model and eliminate the asymmetry.

Figure 15: Price discovery – conditional variance of the fundamental value $\mathrm{Var}(d|p_1)$
In the standard Kyle model $\mathrm{Var}(d|p_1) = \frac{1}{2}\sigma_d^2 = 0.5$

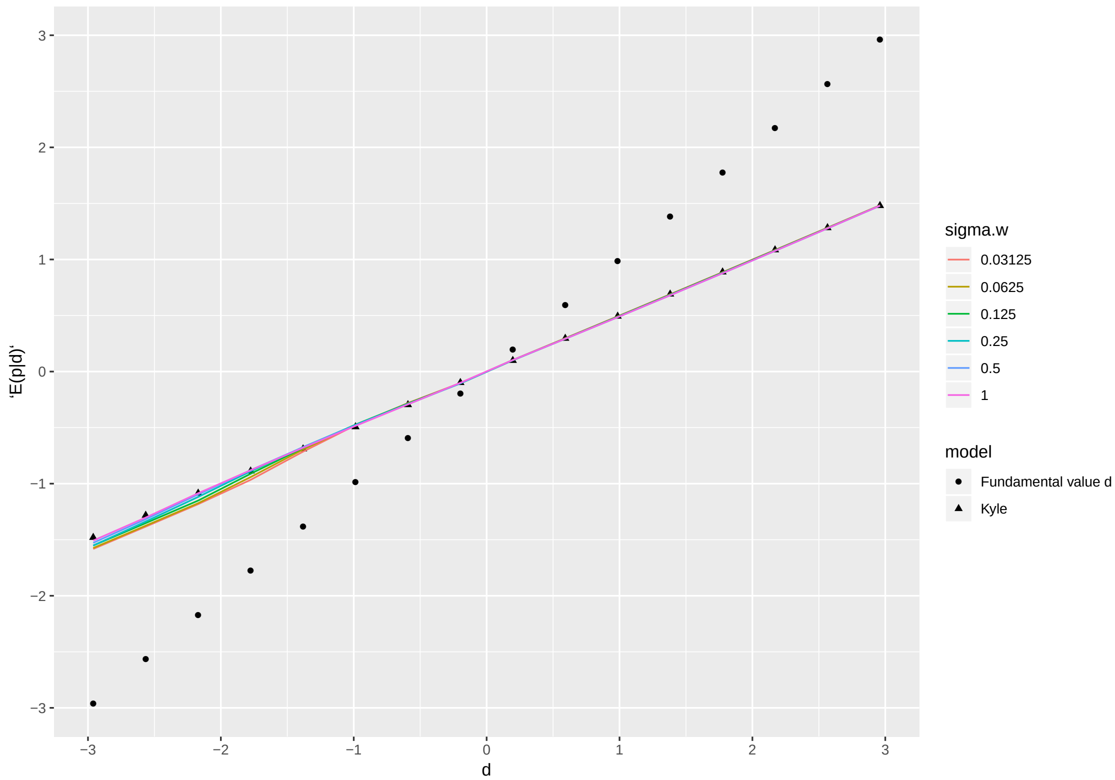


positive part. The amount of extra noise can be judged by looking at the linear "leader-follower" model in Appendix B. In Panel 25b of Figure 25, I show that in the linear model, the amount of extra noise is non-monotonic. If the UNINFORMED BORROWER "w" is very aggressive (ρ^2 close to 0 in the plot), then the LENDER does not "copy" the extra noise, so there should be little noise asymmetry in the full model. If, instead, the UNINFORMED BORROWER "w" is very inactive, this also leads to a relatively small amount of noise: as the activeness of the UNINFORMED BORROWER "w" shrinks, the magnification effect of the LENDER increases but not as fast. All-in-all, in two extreme cases, we have no extra noise while the maximum is attained somewhere between the two extremes.

Additionally, in Figure 16 I analyze the discrepancy between the expected price and the fundamental value from the perspective of the INFORMED TRADER. This discrepancy is what drives the profits of the INFORMED TRADER "x", to what extent the INFORMED TRADER can take advantage of her informational superiority.

Figure 16 highlights that the market price adjusts considerably in response to negative values of private signal d despite the fact that the INFORMED TRADER's short positions are reduced. This happens because the LENDER's behavior approximately compensates the fact

Figure 16: Expected price from the INFORMED TRADER's point of view $\mathrm{E}(p_1|d)$
In the standard Kyle model $\mathrm{E}(p_1|d) = \frac{1}{2}d$



the INFORMED TRADER reduces her short sales. The LENDER does part of the INFORMED TRADER's job in terms of revealing the negative signal to the MARKET MAKER, thereby moving the price downward.

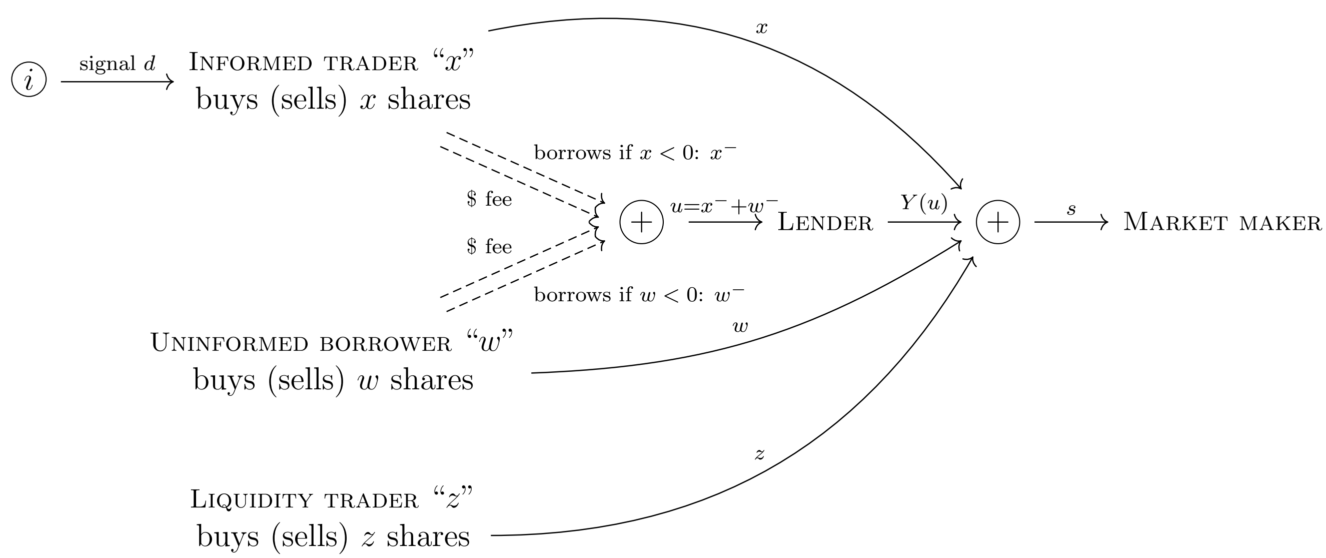
In my model, the expected price from the perspective of the INFORMED TRADER "x" is approximately the same as in the standard Kyle model. The INFORMED TRADER's expected profits per one unit of asset $d - \mathrm{E}(p_1|d)$ are also similar to the ones in the standard Kyle model. In the standard Kyle model, I have $\mathrm{E}(p_1|d) = d - \mathrm{E}(p_1|d) = \frac{1}{2}d$. In my model with front-running and without commission, these equations hold approximately.

5 Model with front-running and commission

In the two previous sections, I have considered the implications of commission and front-running separately. In this section, instead, I allow both mechanisms to work simultaneously. The LENDER can both raise the commission and front-run the INFORMED TRADER. Effectively, this section brings together sections 3 and 4 to understand how the two mechanisms work together.

The following diagram summarizes the model considered in this section.

Figure 17: Diagram of the model front-running



The sequence of actions in the variant of the model portrayed here is identical to the one considered in section 4 apart from the fact that prior to all other decisions, the LENDER sets the commission rate r. As a reminder, I provide a full description of the trading protocol.

1. The LENDER chooses size of commission rate r and commits to lend the securities at this rate.

2. Sequence of moves identical to one in section 4:

   - First move ($t = 1$):
     - True value of the asset d is revealed to the INFORMED TRADER "x";
     - The INFORMED TRADER "x" borrows if necessary $|x^-|$ shares and submits an order to buy (sell) $x(d)$ shares;
     - The UNINFORMED BORROWER "w" borrows if necessary $|w^-|$ shares and submits an order to buy (sell) w shares;
   - Second move ($t = 1$):
     - The LENDER observes $x^- + w^-$;
     - The LENDER submits an front-running order to buy (sell) $y\left[x^- + w^-\right]$ shares;
     - The LIQUIDITY TRADER "z" submits an order to buy (sell) z shares;
   - Clearing ($t = 1$):

  - The MARKET MAKER observes $x(d) + w + y\left[x^{-} + w^{-}\right] + z$;
  - The MARKET MAKER sets the price and executes all the orders;

- $t = 2$: The true value d is paid out.

Once the rate r has been chosen, the game studied in this section is almost identical to the one in section 4. The only difference is that the decisions of the INFORMED TRADER "x" are influenced by the fact that borrowing is no longer free-of-charge. As section 3 suggests, and as we shall see later, this discourages short-selling to the extent that for a certain range of values of d, the INFORMED TRADER "x" completely abstains from trading. Obviously, the strategy of front-running $Y(\cdot)$ and price-setting rule $G(\cdot)$ also adjust as a result.

In the model with both commission and front-running, the LENDER faces a trade-off between making money by charging high commission and extracting information from short-selling activity and front-running.

The higher is the commission rate, the more often does the INFORMED TRADER "x" refrains from short-selling and borrowing. In such cases, she hides her private signal from the LENDER. The probability that the INFORMED TRADER "x" does not borrow $\mathrm{P}\left(X_{r}^{-}(d) = 0\right) \equiv \mathrm{P}\left(X_{r}(d) \geq 0\right)$ generally increases as a function or interest rate r. As a result, if r is high enough, this hinders the LENDER's inferential ability about the liquidation value of the asset d. If r is high enough, the INFORMED TRADER "x" almost never opens short positions, and therefore, the LENDER becomes poorly informed as his signal $\left[u = U_{r}(d, w) \equiv X_{r}(d)^{-} + w^{-}\right]$ becomes almost independent of d. Note that the LENDER becomes less informed even in case the borrowing does not happen. To illustrate this, I compare the two cases – zero commission and a prohibitive rate of commission. If the commission is zero, then $u = 0$ is an insightful signal because it rules out possibilities $d < 0$. If instead, the commission is high, then the signal $u = 0$ becomes much less informative because it no longer rules out the moderate negative values of d.

Since the rate r is set before each of the agents (the INFORMED TRADER, the LENDER, and the MARKET MAKER) makes their decision, the equilibrium strategies $X(\cdot)$, $Y(\cdot)$, and $G(\cdot)$ can be suited only to each particular rate $r = r_1$. If the rate r changes (say takes on a different value, $r = r_2$), then all the strategies $X(\cdot)$, $Y(\cdot)$, and $G(\cdot)$ in equilibrium adjust.

In order to emphasize this, I augment the notations $X(\cdot)$, $Y(\cdot)$, and $G(\cdot)$, with subscript r. In the following subsection, I recast Eqs. 7, 11, 8, 9, and 10 from Section 4 and stipulate a notion of equilibrium with front-running and exogenous commission rate. I visually highlight the new parts of the modified equations.

Definition 5.1. Equilibrium with front-running and exogenous commission rate r is a triple of functions $X_{r}(\cdot)$, $Y_{r}(\cdot)$, and $G_{r}(\cdot)$ such that the following conditions are met:

1. $X_r(\cdot)$ is an optimal strategy of the INFORMED TRADER "x" who takes as given the rate r, the trading strategy of the LENDER ($Y_r(\cdot)$) and the price-setting mechanism of the MARKET MAKER ($G_r(\cdot)$):

$$X_r\left(\hat{d}\right) = \arg\max_x \left\{ \underbrace{x \cdot \left(\hat{d} - \mathrm{E}\left[G_r(x + w + Y_r(x^- + w^-) + z)\right]\right)}_{\text{expected gross profit}} \underbrace{+rx^-}_{\text{borr. costs}} \right\}. \tag{13}$$

2. $Y_r(\cdot)$ is an optimal strategy to front-run on the information available to the LENDER– signal u. In this context the LENDER takes r, $X_r(\cdot)$ and $G_r(\cdot)$ as given. In particular, this means that Y_r has to satisfy the following condition:

$$Y_r\left(\hat{u}\right) = \arg\max_y \left\{ \mathrm{E}\left[\Pi^{\text{front}}_{X_r,G_r}(y, d, w, z) \middle| U_r(d, w) = \hat{u}\right]\right\} \text{ where} \tag{14}$$

$$\Pi^{\text{front}}_{X_r,G_r}(y, d, w, z) = y \cdot \left[d - G_r(X_r(d) + y + z + w)\right], \text{ and} \tag{15}$$

$$U_r(d, w) = X_r{}^-(d) + w^- \equiv \min(0, X_r(d)) + \min(0, w). \tag{16}$$

3. The price setting rule that the MARKET MAKER follows ($G_r(\cdot)$) is consistent with the trading strategies of the INFORMED TRADER "x" and the LENDER. In particular, this means that G_r satisfies the following equation.

$$G_r(\hat{s}) = \mathrm{E}\left(d \middle| X_r(d) + w + Y_r(X_r{}^-(d) + w^-) + z = \hat{s}\right). \tag{17}$$

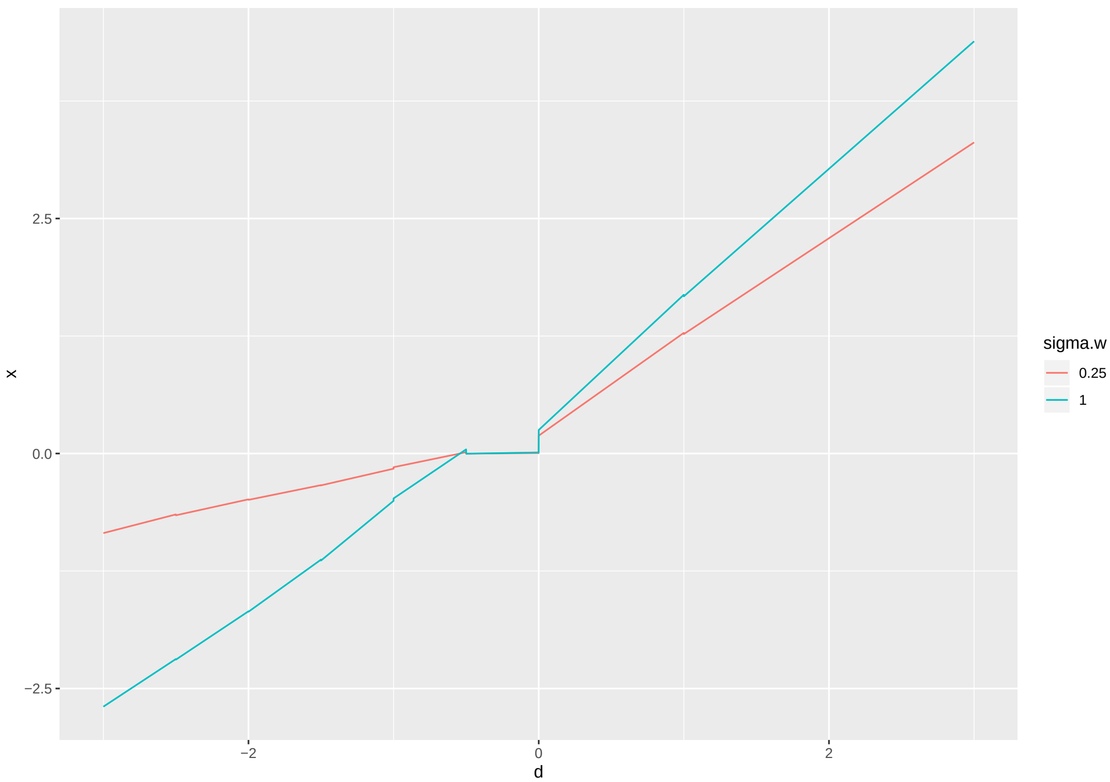
I find equilibrium numerically for a particular exogenous value of $r = 0.5$. In Figures 18, 19, and 20, I show equilibrium strategies of the INFORMED TRADER "x", the LENDER, and the MARKET MAKER, respectively.

Equilibrium trading strategy of the INFORMED TRADER inherits features both from the variants of the model with only commission (Section 3) and of the model with only front-running (Section 4).

The optimal order size $X_r(d)$ has a non-trade region just like in the model with commission: within a certain range of values of d, $X_r(d) = 0$, that is, the INFORMED TRADER abstains from trading because potential gains from short-selling are less than the commission fixed cost.

Similarly to the equilibrium strategy in Section 4 (front-running only), when the INFORMED TRADER does eventually short-sell, the magnitude of the short position is decreased quasi proportionally because the INFORMED TRADER takes into account the additional price pressure coming from LENDER's front-running.

Figure 18: INFORMED TRADER's order size as a function of private signal: $x = X_r(d)$



5.1 Profits/losses breakdown

In this section, I analyze the competition between the INFORMED TRADER "x" and the LENDER. The split of the profits between the two agents is affected by the aggressiveness of the UNINFORMED BORROWER "w".

First, as the aggressiveness of the UNINFORMED BORROWER "w" shrinks, she submits smaller order sizes, naturally, her losses also shrink. Even though LIQUIDITY TRADER's trading decisions do not change, her expected losses increase. The main reason is that she becomes the only looser in the economy. Academically speaking, she faces more adverse selection when the relative weight of the UNINFORMED BORROWER "w" shrinks.

Overall, as the aggressiveness of the UNINFORMED BORROWER "w" shrinks, the aggregate losers' activity moderates. Naturally, the net effect is that the aggregate losses in the economy moderate as well.

We are dealing with a zero-sum game. Thus, the two winners – the INFORMED TRADER "x" and the LENDER – share between them the prey from the two losers – the UNINFORMED BORROWER "w" and the LIQUIDITY TRADER "z".

As the aggressiveness of the UNINFORMED BORROWER "w" shrinks, the INFORMED TRADER "x" also becomes more cautious. Conversely, the LENDER becomes more informed and more

Figure 19: Front-running: LENDER's proprietary order size as a function of short interest: $y = Y(u)$, $u = x^{-} + w^{-}$

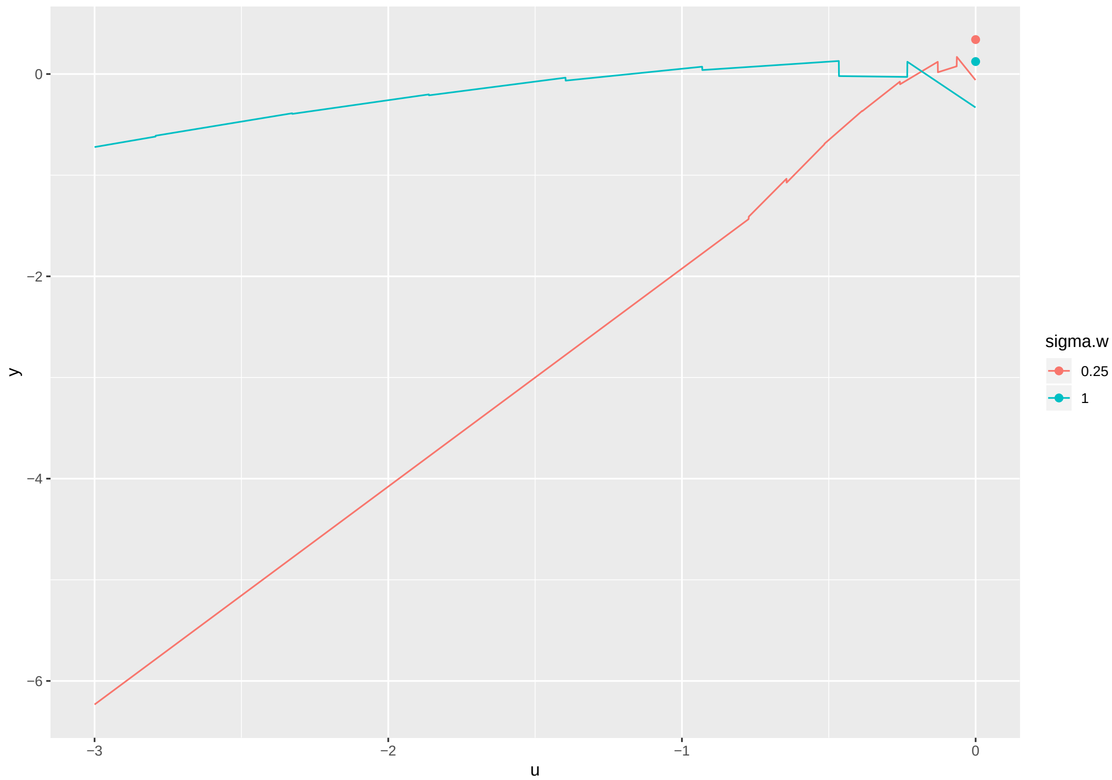
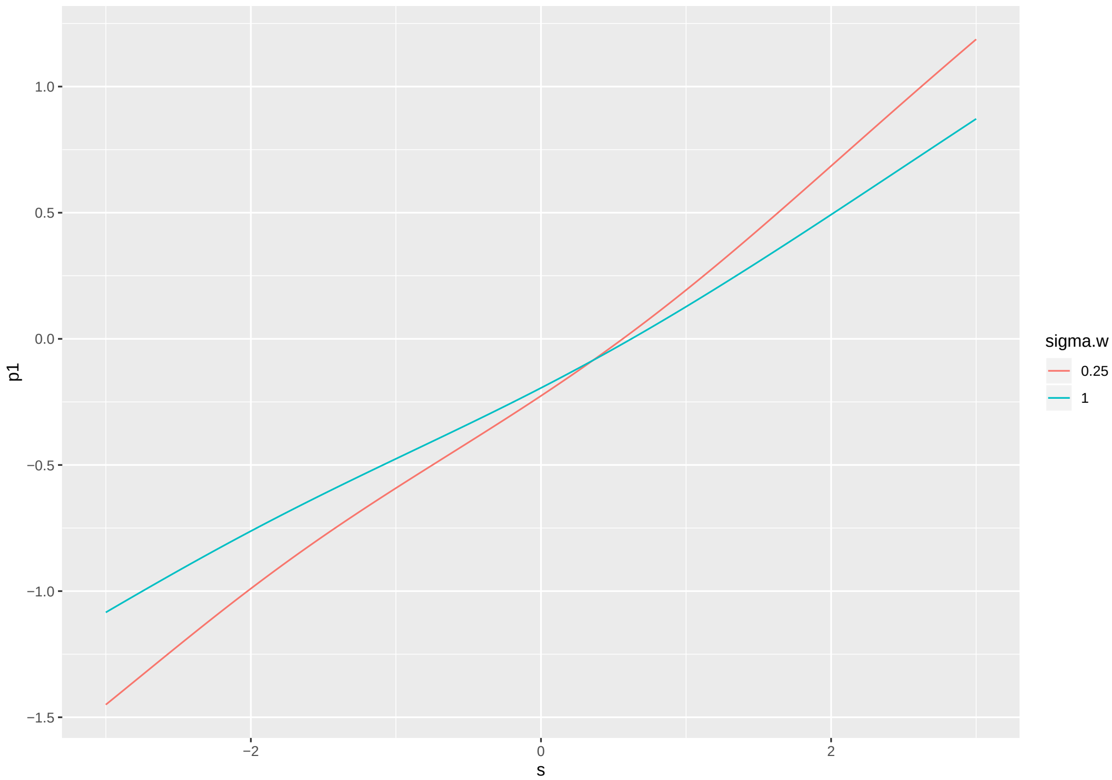


Figure 20: MARKET MAKER's price setting rule in equilibrium $p_1 = G(s), s = x + w + y + z$



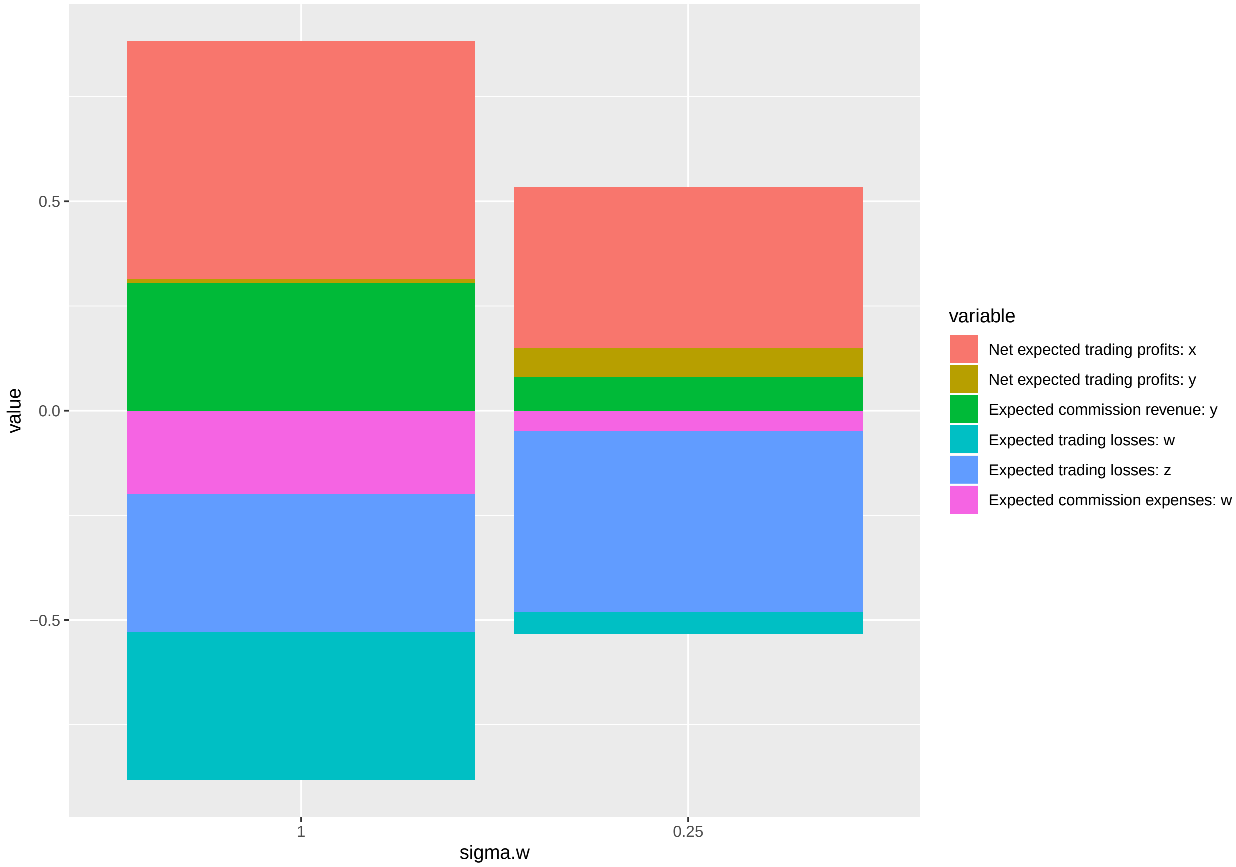
aggressive. As a result, a higher proportion of the expected profits is allocated to the LENDER.

In this section, I additionally analyze how exactly the agents win or lose. In particular, I distinguish between commission expenses and trading losses. Similarly, I distinguish between commission revenues and trading profits. I investigate in which cases commission revenues/expenses are positively related to trading profits/losses, and in which cases they are inversely related. As UNINFORMED BORROWER "w"becomes less aggressive, both her commission expenses and trading losses shrink. Regarding the INFORMED TRADER "x", as she also becomes less aggressive as UNINFORMED BORROWER curbs her trades. As the INFORMED TRADER "x" trades less, she earns less in terms of trading but at the same time spends less for the commission.

In equilibrium, the decrease of the INFORMED TRADER's trading profits is only partially offset by the decrease of the INFORMED TRADER's commission paid to the LENDER. Overall, the net expected profits of the INFORMED TRADER "x" decrease as σ_w shrinks.

As opposed to the INFORMED TRADER "x", the LENDER becomes and more aggressive when the UNINFORMED BORROWER becomes less aggressive because the LENDER is more informed. For her, the commission income and the trading profits move in opposite directions. As the UNINFORMED BORROWER becomes less aggressive, the LENDER's commission income shrinks while the trading profits increase.

Figure 21: Profit sharing between the INFORMED TRADER "x" and the LENDER, loss sharing between the UNINFORMED BORROWER "w" and the LIQUIDITY TRADER "z". The bars above zero depict the breakdown of the profit. The bars below zero depict the breakdown of the losses.



The net effect is that LENDER earns less overall, but it is noteworthy that the negative effect of commission dominates just because the commission rate that I chose for this example is rater high.

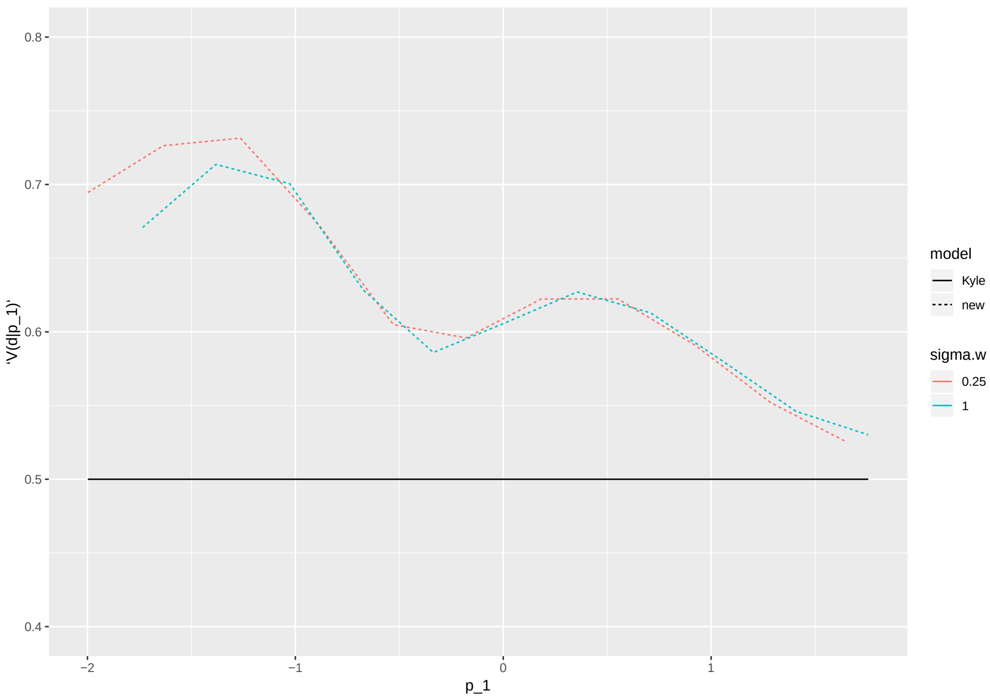
Figure 21 illustrates the overall expected profits/losses[15] in the model, how the profits are shared between the INFORMED TRADER "x" and the LENDER, and how the losses are shared between the UNINFORMED BORROWER "w" and the LIQUIDITY TRADER "z".

5.2 Market efficiency

In this subsection, I analyze the price efficiency in the setting where both commission and front-running are present. To this end, in Figure 22, I depict a measure of price efficiency – conditional variance of the fundamental value d given the price level p_1, i.e. how imprecise is my forecast about d given that I know p_1.

[15]Since my model is a zero-sum game, overall expected profits are equal to overall expected losses

Figure 22: Price discovery – conditional variance of the fundamental value $\mathrm{Var}(d|p_1)$
In the standard Kyle model $\mathrm{Var}(d|p_1) = \frac{1}{2}\sigma_d^2 = 0.5$



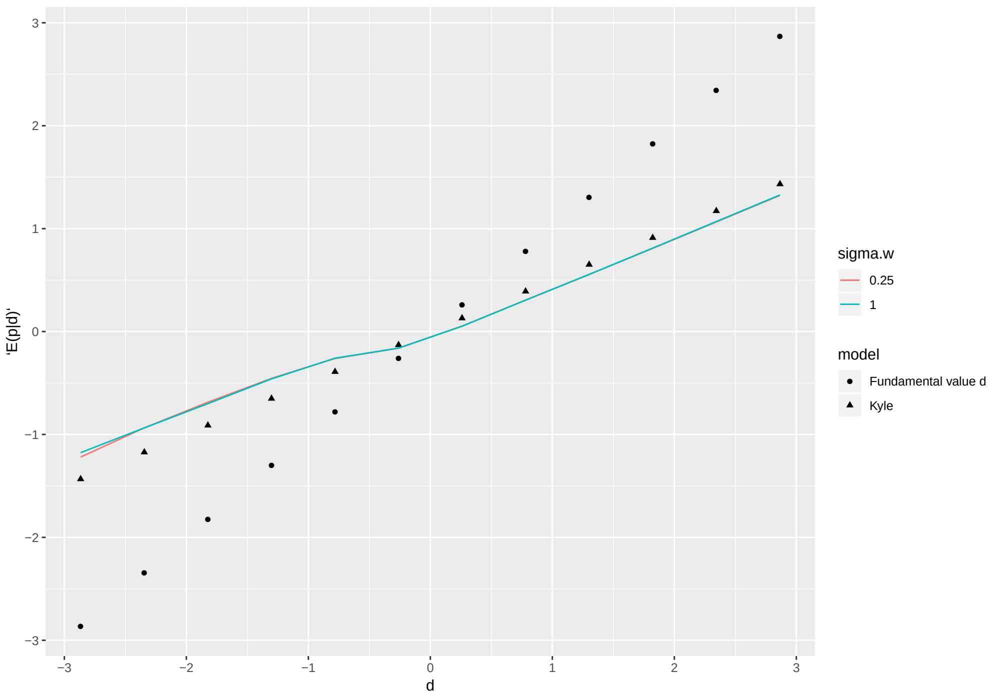
Compared to Subsection 4.3, the overall price efficiency is lower: the forecast of d given p_1 is less precise. The shape of the conditional variance curve inherits the properties of the corresponding curves in the model with “front-running” only (Section 4) and with commission only (Section 3).

The overall decline in the curves is due to the effects highlighted in Section 4.3 – model with “front-running” only. The upward-sloping part is due to the effects coming from the borrowing commission.

The extent to which the fundamental value d is reflected in the price can be also measured by looking at how close $\mathrm{E}(p_1|d)$ is to actual d. Figure 23 demonstrates that in the model with front-running and commission, the private signal d is reflected in price to a lesser extent than in the standard Kyle model: $\mathrm{E}(p_1|d)$ is farther from d than in the standard Kyle model.

It is also noteworthy that $\mathrm{E}(p_1|d = 0)$ is below zero. This effect also manifests itself in the commission-only model, and it is due to the non-trade region. Because of the commission, the INFORMED TRADER “x” abstains from trading or reduces the magnitude of sell orders. This makes the case $d = 0$ more similar to $d = -\frac{1}{2}$ than to $d = \frac{1}{2}$ from the perspective of the MARKET MAKER. This effect is described in detail in Pankratov (2018).

Figure 23: Expected price from the INFORMED TRADER's point of view $\mathrm{E}(p_1|d)$
In the standard Kyle model $\mathrm{E}(p_1|d) = \frac{1}{2}d$



6 Trade-off between front-running and commission

In this section, I focus on how the LENDER optimally solves the trade-off between extracting as much as possible commission revenues and extracting as much as possible from front-running.

Obviously, the two activities – charging commission and front-running – are mutually cannibalistic. When the LENDER charges a higher commission rate r, he forgoes a part of the information about the INFORMED TRADER's decisions and therefore part of the front-running profits. Conversely, when the LENDER front-runs the INFORMED TRADER "x", the LENDER discourages the INFORMED TRADER "x" from short-selling and therefore forgoes part of the commission.

It is a natural question therefore whether a monopolistic LENDER optimally chooses to charge any commission whatsoever given that he has a possibility to front-run, and if so, then whether the optimally chosen commission rate in the presence of front-running is significantly smaller than the optimally chosen commission rate in the absence of commission. In order to answer this question, I need to compare the implications of models with front-running [like in Section 5] and without front-running [like in Section 3] but otherwise comparable.

To achieve full comparability, I need to take another step. As I mention in Section 4, the presence of the UNINFORMED BORROWER in the model is necessary for the existence of equilibrium with front-running.

On the other hand, if I include the UNINFORMED BORROWER in the model as is, it becomes impossible to cast a sensible problem for optimal commission rate. The maximization problem breaks down because of the assumption that the UNINFORMED BORROWER's trading decisions do not depend on the magnitude of the commission: even if the commission is extremely high, the UNINFORMED BORROWER often borrows big stakes of the asset – behaves like a shoal of salmon during the mating run and becomes an easy victim of the LENDER. Clearly, this assumption is not suitable for studying the optimal commission rate. If the UNINFORMED BORROWER behaves incautiously enough, then the LENDER's optimal choice is to set $r = +\infty$.

To overcome this issue, I introduce a modification in the model: instead of UNINFORMED BORROWER "w" I consider a CAUTIOUS UNINFORMED BORROWER "W". The CAUTIOUS UNINFORMED BORROWER decreases the borrowing volume in response to increases in the commission rate. In particular, if the commission rate exceeds a short-selling private benefit perceived by the CAUTIOUS UNINFORMED BORROWER, he abstains from borrowing and from trading. One interpretation of the CAUTIOUS UNINFORMED BORROWER is an agent with a wrong signal about d.

Suppose, the CAUTIOUS UNINFORMED BORROWER believes that $d = d_{\text{fake}} := \frac{\sigma_d}{\sigma_w} w$ [16] and in the absence of commission he submits a market order $w = \frac{\sigma_w}{\sigma_d} d_{\text{fake}}$. In the pesence of commission rate at rate r, the CAUTIOUS UNINFORMED BORROWER chooses to moderate his short-selling activities as though his perceived expected gains from short selling shrink from $-d_{\text{fake}}$ to $-d_{\text{fake}} - r$:[17]

$$W := \begin{cases} \frac{\sigma_w}{\sigma_d} d_{\text{fake}} \text{ for } d_{\text{fake}} \geq 0, \\ \left[\frac{\sigma_w}{\sigma_d}[d_{\text{fake}} + r]\right]^-, \text{ otherwise.} \end{cases} \tag{18}$$

In terms of old notation, the CAUTIOUS UNINFORMED BORROWER's rule can be written as follows:

$$W = \begin{cases} w \text{ for } w \geq 0, \\ \left[w + \frac{\sigma_w}{\sigma_d} r\right]^-, \text{ otherwise.} \end{cases} \tag{19}$$

Such a behavior of the CAUTIOUS UNINFORMED BORROWER "W" can be interpretated as an optimal decision under wrong beliefs and risk-aversion. Let me additionally assume that that the CAUTIOUS UNINFORMED BORROWER "W" believes that the execution price is distributed normally $p_1 \sim N(0, \sigma_p^2)$ and her absolute risk aversion is γ, $u_W(\text{profits}) = -e^{-\gamma \cdot \text{profits}}$. In this case, the net price increase is $d + r\mathbb{I}_{W<0} - p_1$. According to the CAUTIOUS UNINFORMED BOR-

[16] With this notation, d and d_{fake} are independent and identically distributed. This means that the CAUTIOUS UNINFORMED BORROWER "W" is mistken about the realization of d but has a correct perception of the probability distribution of d.

[17] This implicitly assumes that the CAUTIOUS UNINFORMED BORROWER "W" is not strategic: he does not take into account price impact from his own trade. In reality, the non-trader region that arises due to commission must be even wider.

ROWER, this quantity is distributed normally $N(d_{\text{fake}} + r\mathbb{I}_{W<0}, \sigma_p^2)$. Therefore, the probability distribution of gains from trading is $N(W(d_{\text{fake}} + r\mathbb{I}_{W<0}), W^2\sigma_p^2)$, and the expected utility is $-\exp\left[-\gamma W(d_{\text{fake}} + r\mathbb{I}_{W<0}) + \frac{1}{2}\gamma^2 W^2 \sigma_p^2\right]$. The CAUTIOUS UNINFORMED BORROWER chooses W to maximize this quantity, which is equivalent to the maximization of

$$\gamma W(d_{\text{fake}} + r\mathbb{I}_{W<0}) - \frac{1}{2}\gamma^2 W^2 \sigma_p^2.$$

F.O.C. can be rewritten as follows:

$$W = \frac{d_{\text{fake}} + r\mathbb{I}_{W<0}}{\gamma\sigma_p^2}.$$

This is not a solution yet because W is on both sides. If $W > 0$, it boils down to $W = \frac{d_{\text{fake}}}{\gamma\sigma_p^2}$. If $W < 0$, it boils down to $W = \frac{d_{\text{fake}}+r}{\gamma\sigma_p^2}$.

In a nutshell, I obtain the following expression.

$$W = \begin{cases} \frac{d_{\text{fake}}}{\gamma\sigma_p^2} \text{ for } d_{\text{fake}} \geq 0, \\ \left[\frac{d_{\text{fake}}+r}{\gamma\sigma_p^2}\right]^{-}, \text{ otherwise.} \end{cases} \tag{20}$$

This is exactly in line with Eq. 18 as long as $\frac{\sigma_w}{\sigma_d} = \frac{1}{\gamma\sigma_p^2}$.

In the setting with the CAUTIOUS UNINFORMED BORROWER "W", I still have one degree of freedom – how aggressive is the CAUTIOUS UNINFORMED BORROWER compared to the LIQUIDITY TRADER. If we interpret the CAUTIOUS UNINFORMED BORROWER "W" as a misinformed agent, then this relative importance of the CAUTIOUS UNINFORMED BORROWER "W" depends on her risk aversion.

If the risk-aversion of the low, the CAUTIOUS UNINFORMED BORROWER "W" is willing to borrow aggressively, and thus, the LENDER is likely to charge a high commission rate. This in turn discourages the INFORMED TRADER from revealing negative signal.

Instead, if the CAUTIOUS UNINFORMED BORROWER's risk-aversion γ is high, she is reluctant to sell short and borrow. Therefore, her contribution to the LENDER's commission income is insignificant, and therefore the LENDER is more likely to set a low commission rate to extract more information from the INFORMED TRADER and trade. As a result of lower commission, the INFORMED TRADER reveals more information.

All-in-all, this is a mechanism through which the misinformed trading strongly decreases market efficiency. On top of the obvious mechanism that the misinformed trading creates additional noise, it also discourages informed trading indirectly because the LENDER sets a higher commission rate.

An additional implication of high aggressiveness of the UNINFORMED BORROWER is a stronger dependence between asset returns and volatility. If the UNINFORMED BORROWER is more aggressive then the market efficiency suffers, and particularly so in the case of "bearish" market ($d < 0$). The INFORMED TRADER is able to reveal positive signals, but refrains from revealing negative signals. As a result, the MARKET MAKER is able to judge the magnitude of a positive news but is blind with respect to the magnitude of negative news. This creates an asymmetric volatility (a.k.a. "leverage") effect roughly through the same mechanism that I explain in Pankratov (2018). From the perspective of a regulator this implies that the misinformation on the market should be very strongly discouraged because misinformed borrowers strongly and negatively affect the market efficiency.

7 Conclusion

I model a market with asymmetric information and short selling. The novelty of my approach is that I explicitly model the incentives and decisions of security lenders who can trade on the information impounded in the asset-borrowing demand (a.k.a. short interest) in addition to charging lending fees. This approach allows me to endogenize short-sale constraints. I show that in equilibrium the possibility of information leakages and short-selling costs discourage short selling.

In addition, I highlight that the explicit commission costs and the possibility of information leakages have different implications for equilibrium behavior of informed agents. In particular, the commission makes it purposeless to short-sell the asset if the potential trading gains do not exceed the commission. As a result, the informed traders sell short only if the signal that they observe is bad and strong enough. Instead, in the absence of commission, the possibility of leakages and front-running works in a different way. Unlike a security lender that only charges a commission, a front-running security lender does not entirely suppress the borrowing demand. Even if the informed trader observes a moderate bad signal about the asset payoff, she does not entirely abstain from trading, she borrows and sells short, but the size is reduced because the informed trader knows that her sales trigger lender's sales and therefore create extra price pressure.

The relative importance of the two profit channels of the security lender (commission and front-running) depends on the level of activity of the uninformed (or misinformed) borrowers. If there is a substantial borrowing demand coming from the uninformed traders, the security lender potentially can bargain on the borrowers by charging high commission without foregoing considerable profits from front-running because in this case the information content of the short interest is poor. If instead, the uninformed agents are reluctant to borrow, the front-running scenario becomes more important. In this case, the short interest becomes highly informative, and as a result, the security lender is more inclined to set a low commission rate, encourage the informed traders to reveal their information, and to use this information by front-running the informed traders. In the extreme case, in which the activity of the uninformed borrowers is negligible and the security lender does not charge a commission, she is able to extract formidable rent by front-running the informed trader. If the activity of the uninformed borrowers is minor, then the informed trader short-selling also becomes extremely cautious, and the dominant fraction of the profits from the asset's price drop is allocated to the front-running security lender. Overall, considering both directions of trading, the expected trading profits of the security lender can be almost as high as the expected profits of the informed trader.

From the perspective of the information efficiency of the market, the scenario in which front running plays a relatively big role is more desirable. With rather inactive uninformed borrowers, the asset security lenders can only rely on the informed borrowers as their source of income. In this setting, they are more likely to choose a moderate interest rate if they have the possibility to front-run the borrowers. In this case, the informed borrowers are more prone to reveal their negative signals to the security lenders and to the market in general. If the lending rate is low, then the informed trader differentiates between bad signals of different magnitude and eventually this information is revealed to the market. From the perspective of "asymmetric" volatility (a.k.a. leverage effect), this effect becomes stronger with high commission rates and dissipates as the commission rate tends to 0.

Also, my paper provides a theoretical background in which the security lenders that have the possibility to front-run charge a lower commission rate. The intuition is that when the lender has the possibility to trade and use the information impounded in the short interest, then a high commission rate makes the front-running activity less profitable. This prediction is in line with a recent empirical finding of Honkanen (2020) that actively managed mutual funds set lower commission rates for stock lending compared to the funds with passive investment strategies.

References

AKERLOF, G. A. (1970): "The Market for "Lemons": Quality Uncertainty and the Market Mechanism*," *The Quarterly Journal of Economics*, 84, 488–500.

ALLEN, F., S. MORRIS, AND A. POSTLEWAITE (1993): "Finite Bubbles with Short Sale Constraints and Asymmetric Information," *Journal of Economic Theory*, 61, 206–229.

BARBON, A., M. DI MAGGIO, F. FRANZONI, AND A. LANDIER (2019): "Brokers and Order Flow Leakage: Evidence from Fire Sales," *The Journal of Finance*, 74, 2707–2749.

BLACK, F. (1976): "Studies of Stock Price Volatility Changes," *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association, Washington DC*, 177–181.

BOEHMER, E., Z. R. HUSZAR, AND B. D. JORDAN (2010): "The good news in short interest," *Journal of Financial Economics*, 96, 80–97.

DANIELSEN, B. R. AND S. M. SORESCU (2001): "Why Do Option Introductions Depress Stock Prices? A Study of Diminishing Short Sale Constraints," *Journal of Financial and Quantitative Analysis*, 36, 451–484.

DANTHINE, J.-P. AND S. MORESI (1998): " Front-Running by Mutual Fund Managers: A Mixed Bag *," *Review of Finance*, 2, 29–56.

DIAMOND, D. W. AND R. E. VERRECCHIA (1987): "Constraints on short-selling and asset price adjustment to private information," *Journal of Financial Economics*, 18, 277 – 311.

HONKANEN, P. (2020): "Securities Lending and Trading by Active and Passive Funds," *Working paper.*

HUDDART, S., J. S. HUGHES, AND C. B. LEVINE (2001): "Public Disclosure and Dissimulation of Insider Trades," *Econometrica*, 69, 665–681.

JONES, C. M. AND O. A. LAMONT (2002): "Short-sale constraints and stock returns," *Journal of Financial Economics*, 66, 207–239.

KERVEL, V. V. AND A. J. MENKVELD (2019): "High-Frequency Trading around Large Institutional Orders," *Journal of Finance*, 74, 1091–1137.

KYLE, A. S. (1985): "Continuous auctions and insider trading," *Econometrica: Journal of the Econometric Society*, 1315–1335.

MELE, A. AND F. SANGIORGI (2020): "Trading Disclosure Requirements and Market Quality Tradeoffs," *Unpublished*.

PANKRATOV, A. (2018): "Informed trading, short-sale constraints, and leverage effect in equity returns," *Unpublished*.

——— (2019): "Information leakages, distribution of profits from informed trading, and last mover advantage," *Unpublished*.

SANNIKOV, Y. AND A. SKRZYPACZ (2016): "Dynamic Trading: Price Inertia and Front-Running," Research Papers 3487, Stanford University, Graduate School of Business.

SENECA, J. J. (1967): "Short Interest: Bearish Or Bullish?" *Journal of Finance*, 22, 67–70.

YANG, L. AND H. ZHU (2020): "Back-Running: Seeking and Hiding Fundamental Information in Order Flows," *Review of Finance*, 33, 1484–1533.

A Price discrimination by brokers

In this section I consider different way in which the LENDER can take advantage of the INFORMED TRADER. In particular, I assume that the LENDER cannot front-run the INFORMED TRADER. Instead, the LENDER is allowed to set the borrowing commission rate depending on the volume borrowed by the clients.

In this set-up, a sensible model can be cast without the UNINFORMED BORROWER. This Section I generalize Eq. 1 and assume instead the following quantity charged by the LENDER:

$$\text{commission}_{\text{inf}} = r\left[-X_r(d)^-\right] \cdot \left[-X_r(d)^-\right], \tag{21}$$

So that the commission rate depends on the borrowing volume.

This results in a more general expression for the INFORMED TRADER's trading profits. Consequently, the INFORMED TRADER solves a more general optimization problem compared to the one in Eq. 6:

$$X_{r(\cdot)}(d) = \arg\max_x \mathrm{E}\left[\underbrace{x \cdot \left[d - G_{r(\cdot)}(x+z+w)\right]}_{\text{gross profits}} \underbrace{+r\left[-x_r^-\right] \cdot x^-}_{\text{borrowing costs}} \middle| d\right] \tag{22}$$

This gives the LENDER more freedom and potentially allows her to earn larger profits.

Assumption A.1. $r(\cdot)$ is a linear function:

$$r(b) = \left[r_0 + k_r b\right]^+. \tag{23}$$

Assumption A.2. The LENDER chooses r_0 and k_r before knowing anything and commits to provide lending at the rate given by Eq. 23.

Theorem A.1. Optimal values for the intercept and slope coefficient of the commision rates are appoximately the following:

$$r_0 \approx 0.5 \cdot \sigma_d \tag{24}$$

$$k_r \approx 0 \cdot \frac{\sigma_d}{\sigma_z} \tag{25}$$

Proof. This result is obtained numerically using the following procedure:

1. Choose certain values of r_0 and k_r.

2. For a given commission rate schedule, find an equilibrium between the INFORMED TRADER and the MARKET MAKER.

3. In this equilibrium find the probability distribution of the order sizes submitted by the INFORMED TRADER.

4. Apply mapping

$$\text{Abs.commission} : x \mapsto \begin{cases} |x| \cdot [r_0 + k_r \cdot |x|]^+ & \text{for negative } x, \\ 0, \text{ otherwise} \end{cases} \tag{26}$$

   in order to obtain the probability distribution of the dollar commission raised by the LENDER.

5. Compute the expected value of the dollar commission raised by the LENDER.

6. Repeat these steps until I find the parameters $\hat{r}_0$, $\hat{k}_r$ that imply the highes possible expected value of the dollar commission.

□

B Linear proxy of front-running model[18]

In this appendix, I consider a proxy model of front running. The model resembles the original model for negative values of d. This model is intentionally made symmetric and has a closed-form solution. This allows to shed light on what happens as the intensity of uninformed borrowing decreases.

In this model the front-runner can judge about d irrespective of its sign: Instead of observing $u = X^-(d) + w^-$, he observes $u = X(d) + w$.

In this setting, the equilibrium conditions are as follows:

$$G(\hat{s}) := \mathrm{E}\left\{ \underbrace{d | X(d) + w + Y\left[X(d) + w\right] + z}_{s} = \hat{s} \right\},$$

$$X(d) = \arg\max_x \mathrm{E}\left[x \cdot \underbrace{(d - G(x + w + Y(x + w) + z))}_{\text{gains per unit of asset}} \middle| d \right],$$

$$Y(u) = \arg\max_y \mathrm{E}\left[y \cdot \underbrace{(d - G(u + y + z))}_{\text{gains per unit of asset}} \middle| X(d) + w = u \right]$$

[18] **Disclaimer: this appendix is not original and is entirely based on my own working paper Pankratov (2019). Notation and terminology have been adjusted to fit the needs of the present paper.**

There is a unique linear equilibrium: a unique triple of coefficients (λ, β, m) such that functions $G(s) = \lambda s$, $X(d) = \beta d$, and $Y(u) = mu$, satisfy the equilibrium conditions.

Figure 24: Amplification coefficient m as a function of $\rho^2 = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_w^2}$

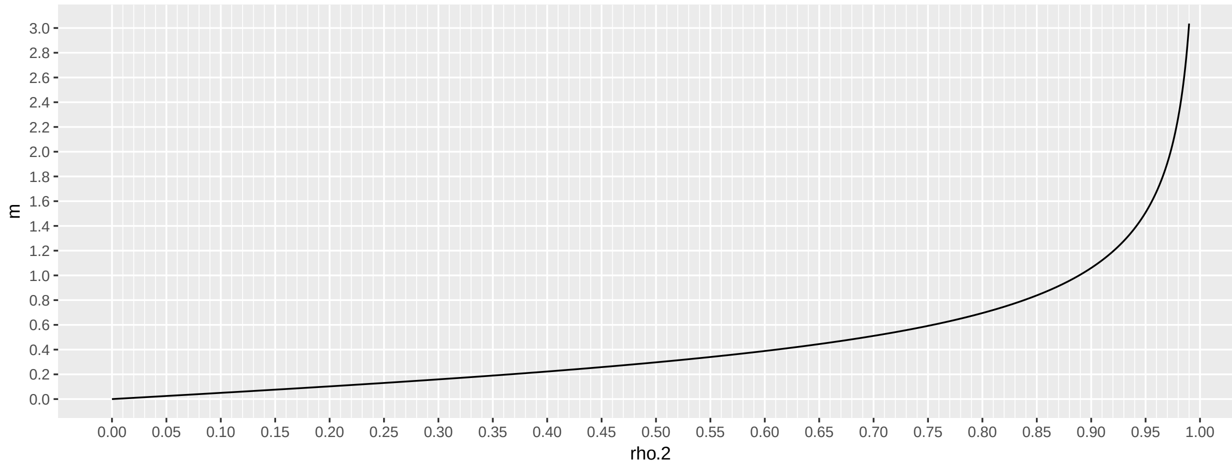
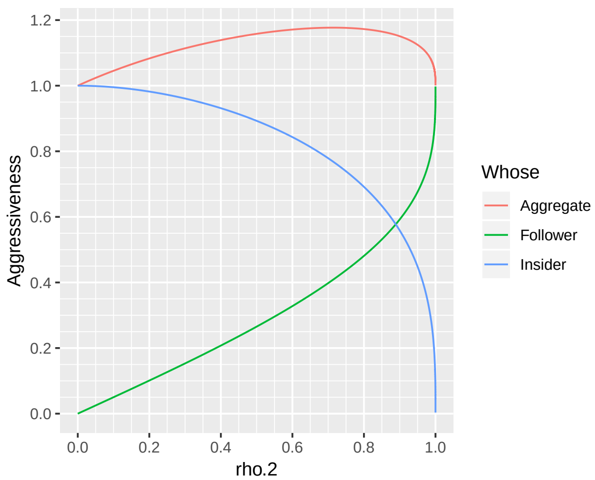
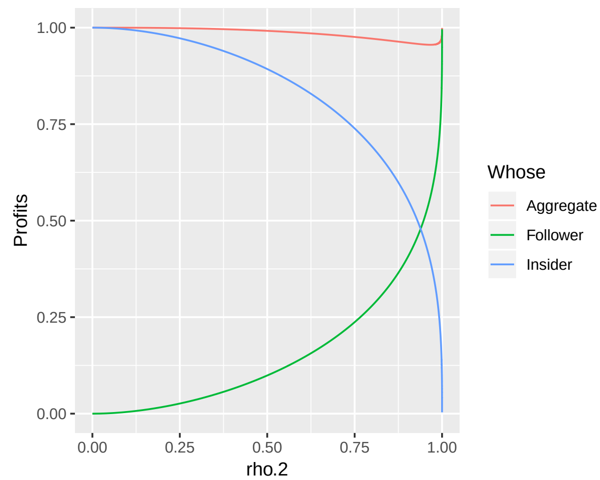


Figure 25: Aggressiveness, expected profits relative to benchmark (Kyle, 1985).

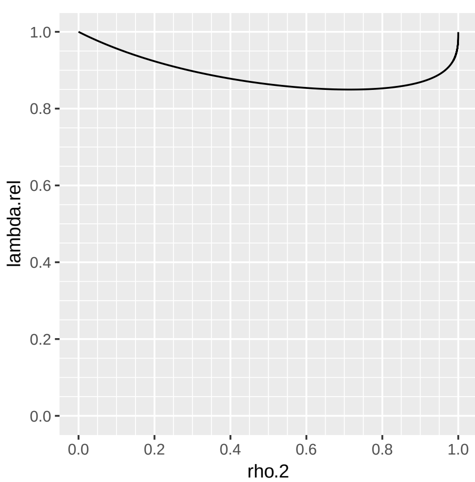
Aggressiveness $/\beta_{\text{Static Kyle}}$



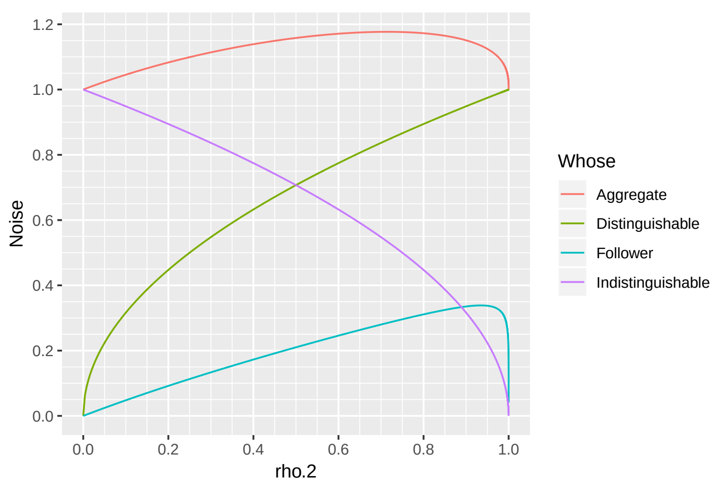
Expected profits $/\pi_{\text{Static Kyle}}$



(a) λ relative to $\lambda_{\text{Static Kyle}}$



(b) Noise breakdown: noise volatilities relative to σ_n



The coefficients are available in closed form. The punchline is that with the increasing precision of the front-runner's inference regarding d ($\sigma_w \to 0$), $\beta \to 0$ and $m \to +\infty$: the INFORMED TRADER "x" almost abstains from trading, while the front-runner strongly magnifies the INFORMED TRADER's orders. Overall, the front-runner's sensitivity to d converges to a finite number, moreover, this number corresponds to the equilibrium aggressiveness of the INFORMED TRADER in the static version of the Kyle model.

As a matter of fact, the behavior of the INFORMED TRADER becomes negligible from the perspective of the MARKET MAKER, while the front-runner becomes almost as knowledgeable as the INFORMED TRADER "x". Eventually, the front-runner assumes the role of the informed trader while the INFORMED TRADER "x" becomes merely a mole transmitting the signal to the front-runner.

B.1 Market efficiency

It is possible to show that the INFORMED TRADER's optimal behavior is:

$$X(d) = \beta d, \text{ where}$$

$$\beta = \frac{1}{2\lambda(1+m)} \tag{27}$$

provided that $G(s) = \lambda s$ and $Y(u) = mu$.

Also, it is possible to demonstrate that MARKET MAKER's rational price-setting rule is:

$$G(s) = \lambda s, \text{ where}$$

$$\lambda = \frac{\beta(1+m)\sigma_d^2}{\beta^2(1+m)^2\sigma_d^2 + \sigma_z^2 + (1+m)^2\sigma_w^2} \tag{28}$$

provided that $X(d) = \beta d$ and $Y(u) = mu$.

A natural substitution $B = \beta(1+m)$ leads to an elegant system of equations:

$$B = \frac{1}{2\lambda}, \tag{29}$$

$$\lambda = \frac{B\sigma_d^2}{B^2\sigma_d^2 + \sigma_z^2 + (1+m)^2\sigma_w^2}, \tag{30}$$

which is easy to solve by plugging Eq. 29 in to Eq. 30, and obtain an equation for B:

$$\frac{1}{2B} = \frac{B\sigma_d^2}{B^2\sigma_d^2 + \sigma_z^2 + (1+m)^2\sigma_w^2} \tag{31}$$

with a unique sensible (positive) solution:

$$B = \frac{\sqrt{\sigma_z^2 + (1+m)^2\sigma_w^2}}{\sigma_d}. \tag{32}$$

I measure market efficiency through the following quantity:

$$\mathrm{Var}(d|p_1) = \mathrm{E}\left[(d - \mathrm{E}(d|p_1))^2\right] = \mathrm{E}\left[(d - p_1)^2\right] = \mathrm{Var}(d - p_1)$$

I find it using the relationship:

$$\frac{\mathrm{Var}(d|p_1)}{\mathrm{Var}(d)} = \frac{\mathrm{Var}(s|d)}{\mathrm{Var}(s)}$$

The closer this value to zero, the more efficient is the price. This ratio can be computed as the ratio of the variance of the information part of s over the the entire variance of s:

$$s = x + w + m \cdot (x + w) + z = (1 + m)(\beta d + w) + z = Bd + (1 + m)w + z.$$

Finally, the variance ratio is:

$$\frac{(1+m)^2\sigma_w^2 + \sigma_z^2}{B^2\sigma_d^2 + (1+m)^2\sigma_w^2 + \sigma_z^2}$$

Using equation 32, I have $B^2\sigma_d^2 = \sigma_z^2 + (1+m)^2\sigma_w^2$.

Thus, the variance ratio is:

$$\frac{\mathrm{Var}(d|p_1)}{\mathrm{Var}(d)} = \frac{1}{2},$$

identical to that of the standard Kyle model.

It is noteworthy that in order to find the degree of price informativeness, the knowledge of the coefficient m is unnecessary. To obtain this result, I simply use the Bayesian updating rule employed by the MARKET MAKER and the optimality condition of the INFORMED TRADER "x". Note that Eq. 32 follows from Eqs. 27 and 28 that describe the behavior of the informed trader and the market maker. Even if the coefficient m were chosen exogenously, the price informativeness would have been the same.

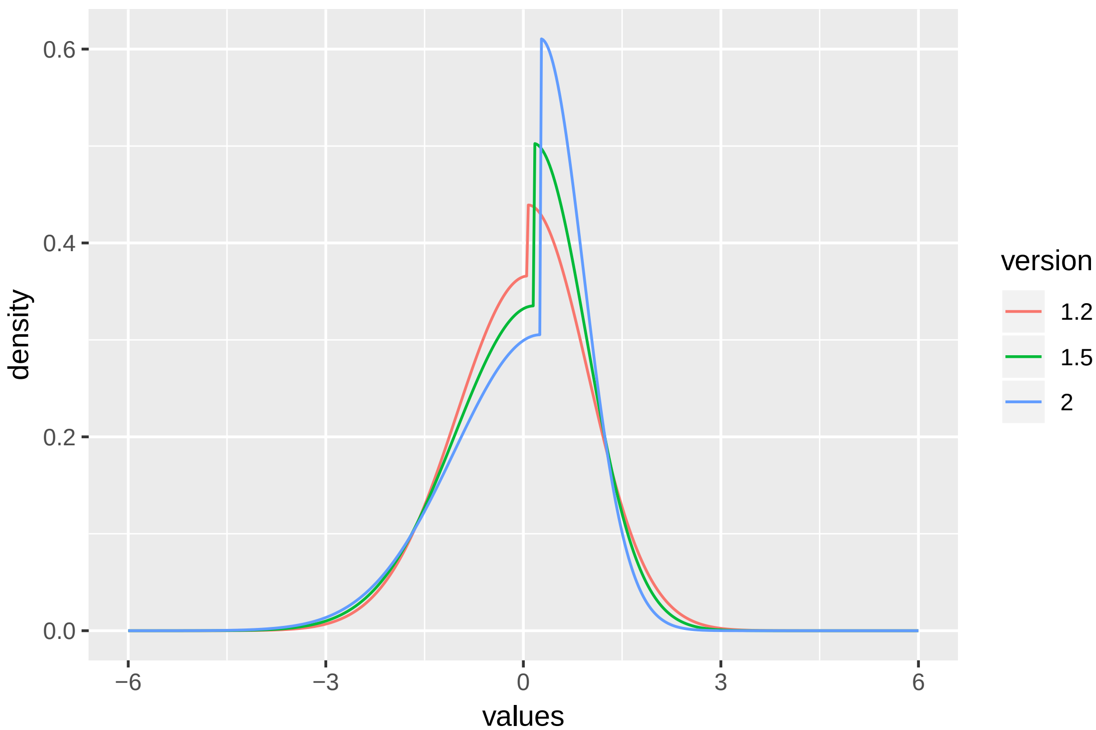
C Kyle model with asymmetric noise

In this Appendix, I drop the UNINFORMED BORROWER "w" and the LENDER, and assume that the distribution of LIQUIDITY TRADER's order size is asymmetric. Eventually, I have the standard Kyle model with just a single departure in assumptions - non-Gaussian noise.

In Figure 26, I depict the probability density function of z (three alternative versions). This probability density function is obtained in the following way.

- I take a standard normal variable $\tilde{z}$.
- I apply a monotonic transformation to $\tilde{z}$ so the negative tails have bigger impact: $z = k\tilde{z}^- + \tilde{z}^+$, where $k > 1$; in particular, I consider 3 cases: $k = 1.2, 1.5, \text{and } 2$.
- I apply a positive affine tansformation to z to make sure that $\mathrm{E}(z) = 0$ and $\mathrm{Var}(z) = 1$.

Figure 26: Density functions of the LIQUIDITY TRADER "z"



Unlike the model in the body of this paper, in this simplified setting, the order flow is given by:

$$s = x + z = X(d) + z,$$

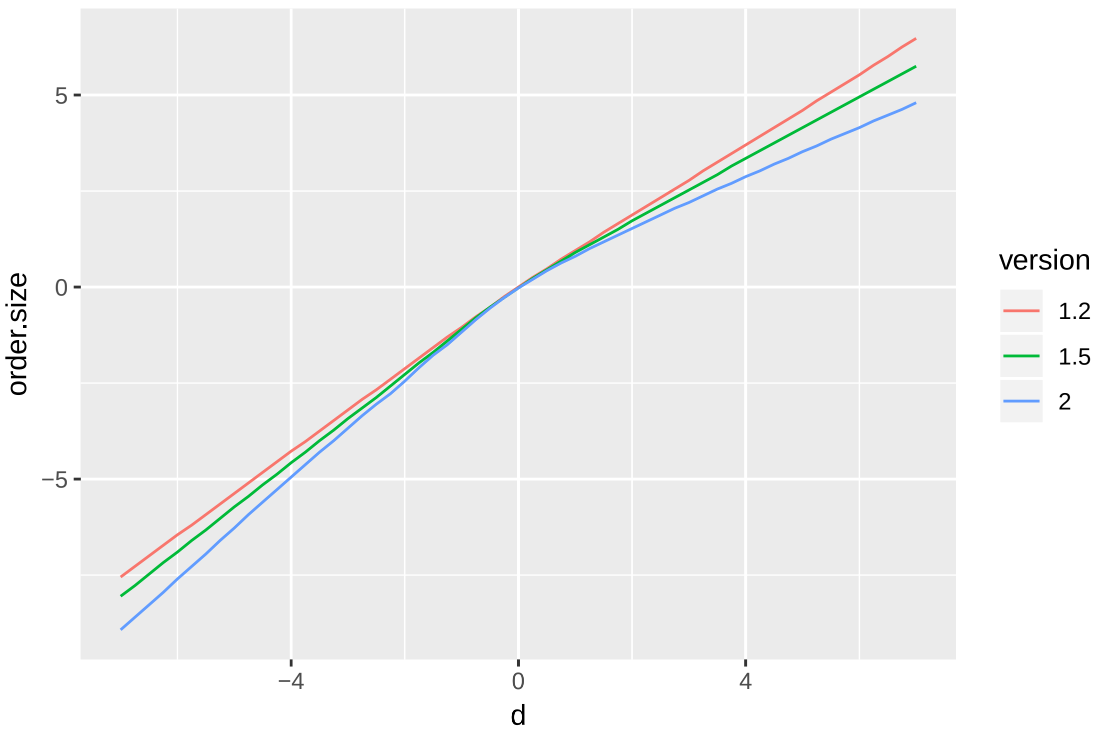
and the price is set accordingly:

$$p_1 = \mathrm{E}(d|X(d) + z).$$

In case $d \gg 0$, it is mostly the right half of the distribution of z that plays a role in the MARKET MAKER's inference process. Conversely, if $d \gg 0$, it is mostly the left half of the distribution of z that plays a role in the MARKET MAKER's inference process. This explains the equilibrium strategies of the INFORMED TRADER "x" are similar to the linear Kyle model with small noise in the right half, while they are similar to the linear Kyle model with large noise in the left half.

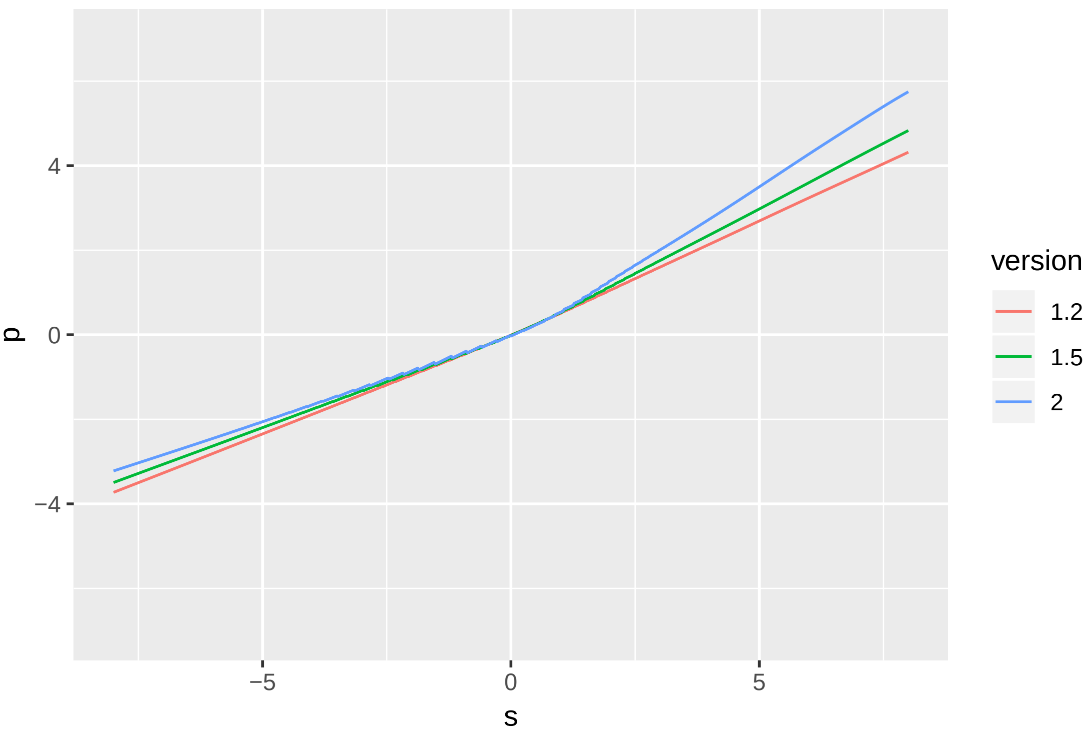
The INFORMED TRADER "x" is more cautious in the positive region because there is relatively little noise to hide behind. If the INFORMED TRADER "x" buys too much, hardly ever the MARKET MAKER will mistake him for the LIQUIDITY TRADER "z". Extremely large sales by the LIQUIDITY TRADER "z" are more plausible, so the INFORMED TRADER "x" is also more prone to sell more because he can plausibly mislead the MARKET MAKER. I show in Figure 27, that the magnitude of asymmetry between buying and selling is directly related to the degree of asymmetry in the distribution of z.

Figure 27: Equilibrium trading strategy of the INFORMED TRADER "x"



Both INFORMED TRADER "x" and UNINFORMED BORROWER "w" are more prone to sell extreme quantities than to buy extreme quantities. As a result, the price-setting mechanism also works to compensate for this effect. For $s \gg 0$, the price-setting rule works like in the linear Kyle model with smaller σ_z. Conversely, for $s \ll 0$, the price-setting rule works like in the linear Kyle model with larger σ_z. In Figure 28, I show that the degree of asymmetry in pricing is directly related to the degree of asymmetry in the distribution of z.

Figure 28: Price setting rule employed by the MARKET MAKER



In order to understand the reasons for the deviation of the price efficiency from the linear Kyle model, develop a theoretical expression for the conditional density function. First of all, instead of conditioning on price, I condition on s because it is equivalent and less tedious.

The density of $x = X(d)$ is given by

$$f_x(\hat{x}) = \frac{\phi(X^{-1}(\hat{x}))}{X'(X^{-1}(\hat{x}))}$$

The joint density of (x, z) is given by

$$f_{(x,z)}(\hat{x}, \hat{z}) = f_x(\hat{x}) f_z(\hat{z})$$

The joint density of $(x, s = x + z)$ is given by

$$f_{(x,s)}(\hat{x}, \hat{s}) = f_x(\hat{x}) f_z(\hat{s} - \hat{x})$$

The conditional density of $x|s = x + z$ is given by

$$f_{x|s}(\hat{x}, \hat{s}) = \frac{f_x(\hat{x}) f_z(\hat{s} - \hat{x})}{\int\limits_{\dot{x}} f_x(\dot{x}) f_z(\hat{s} - \dot{x}) \mathbf{d}\dot{x}} = \frac{\frac{\phi(X^{-1}(\hat{x}))}{X'(X^{-1}(\hat{x}))} f_z(\hat{s} - \hat{x})}{\int\limits_{\dot{x}} \frac{\phi(X^{-1}(\dot{x}))}{X'(X^{-1}(\dot{x}))} f_z(\hat{s} - \dot{x}) \mathbf{d}\dot{x}} =$$

$$= \frac{\frac{\phi(\hat{d})}{X'(\hat{d})} f_z(\hat{s} - X(\hat{d}))}{\int\limits_{\dot{d}} \phi(\dot{d}) f_z(\hat{s} - X(\dot{d})) \mathbf{d}\dot{d}}$$

The conditional density of $d|s$ is given by

$$f_{d|s}(\hat{d}, \hat{s}) = \phi(\hat{d}) \frac{f_z(\hat{s} - X(\hat{d}))}{\int\limits_{\dot{d}} f_z(\hat{s} - X(\dot{d}))\phi(\dot{d})\mathbf{d}\dot{d}} = \phi(\hat{d}) \frac{f_z(\hat{s} - X(\hat{d}))}{\mathrm{E}(f_z(\hat{s} - X(d)))}$$

Let me now focus on the numerator in the last equation and compare two cases:

1. $\hat{s}$ is moderately negative, say the value corresponding to $p_1 = -1$,

2. $\hat{s}$ is moderately positive, say the value corresponding to $p_1 = 1$.

In the first case the distribution $d - p_1$ has particularly fat right tail for the following reasons.

1. p_1 is negative because s is negative.

2. For positive $\hat{d}$, $X(\cdot)$ is relatively flat, so for rather big positive $\hat{d}$, $\hat{s} - X(\hat{d})$ not too big in absolute value.

3. Left tail of f_z fades out relatively slowly.

In the second case, I do not have that big positive outliers in the distribution of $d - p_1$, because now, $p_1 = 1$ instead of -1, so d has to be even bigger. At this stage, $\phi(\hat{d})$ kicks in and makes such extreme values very unlikely.

In the secon case, I could have had negative outliers. Neverthelwss, this does not happen because of the following reasons.

1. For negative $\hat{d}$, $X(\cdot)$ is relatively steep, so for rather big negative $\hat{d}$, $\hat{s} - X(\hat{d})$ also very big in absolute value.

2. Right tail of f_z fades out relatively quickly.

To sum up, even though, in the tails the model is asymptotically similar to the standard Kyle model, at the point when the negative area merges into the positive area, the mismatches of parameters in the two models leads to significant deviations from the standard Kyle model in terms of price efficiency.

The following two plots depict the conditional volatility and joint density in the model.

Figure 29: Price informativeness: conditional variance $\mathrm{Var}(d|p_1)$

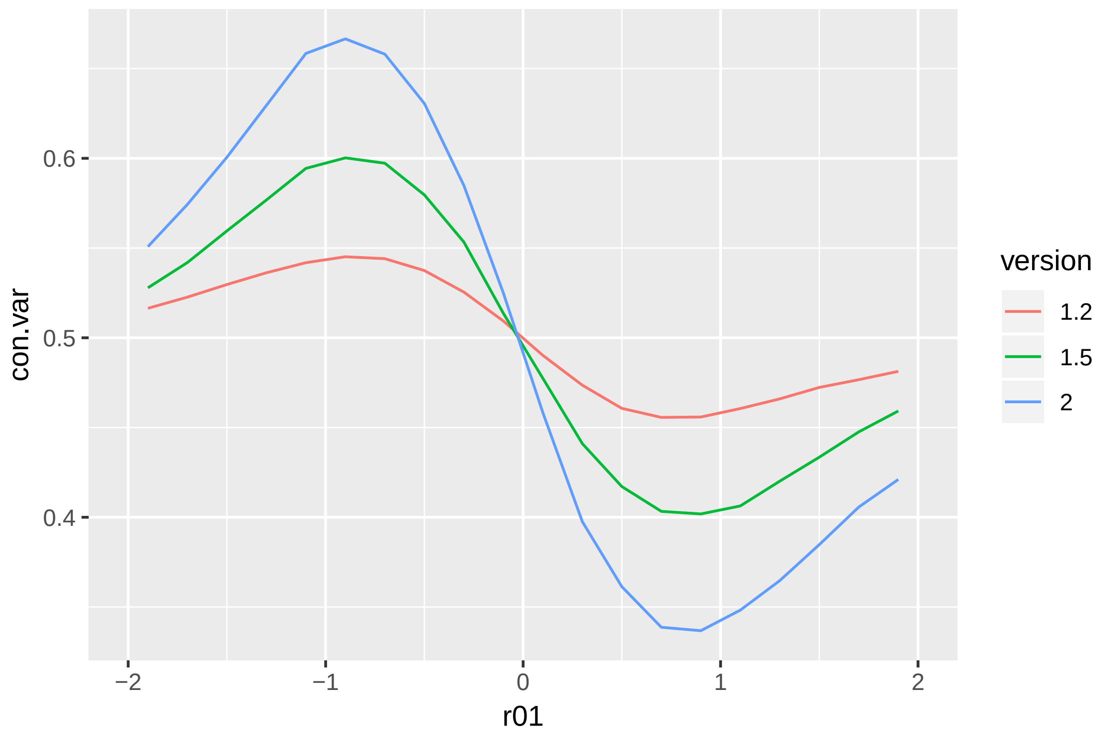
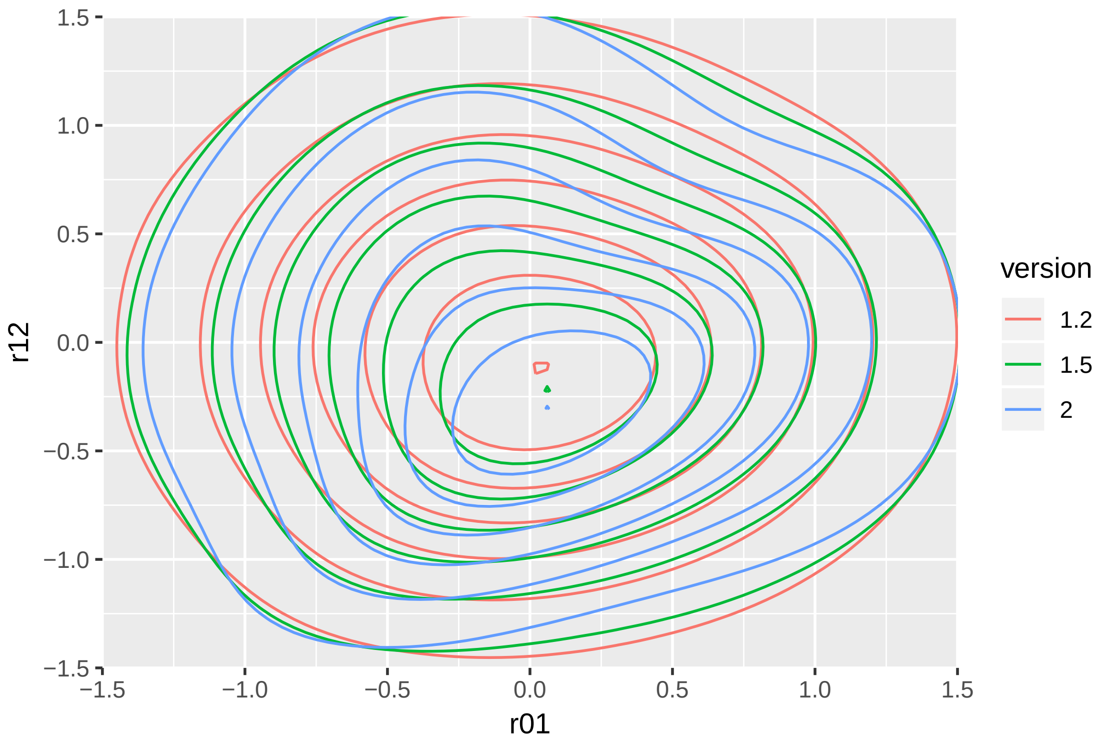


Figure 30: Joint density $(p_1, d - p_1)$



D Analysis of purchasing order size of the lender

For simplicity, I focus on the case $u = 0$. I consider the LENDER's optimization problem:

$$\max_y \{y \cdot [\mathrm{E}(d|u=0) - \mathrm{E}(P_1(d,w,z,y)|u=0)]\}, \text{ where} \tag{33}$$

$$\mathrm{E}(d|u=0) = \mathrm{E}(d|X(d) \geq 0), \text{ and} \tag{34}$$

$$P_1(d,w,z,y) = G(X(d) + w + y + z). \tag{35}$$

The first term $[\mathrm{E}(d|X(d) \geq 0)]$ is virtually independent of y, while the second term $[\mathrm{E}(p_1|u=0)]$ is dependent on y. I linearize the second term to get an insight into what drives the LENDER's behavior.

$$\mathrm{E}(p_1(d,w,z,y)|u=0) \approx p_1^{(0)} + \lambda y, \text{ where} \tag{36}$$

$p_1^{(0)}$ is expected execution price in case of inaction and λ is the sensitivity of expected execution price with respect to order size:

$$p_1^{(0)} := \mathrm{E}(p_1(d,w,z,y)|u=0)|_{y=0}, \text{ and} \tag{37}$$

$$\lambda := \frac{\partial}{\partial y}\mathrm{E}(p_1(d,w,z,y)|u=0)|_{y=0} = \mathrm{E}(G'(X(d)+w+z)|u=0) \quad ^{19} \tag{38}$$

In this notation, the maximization problem can be approximated follows:

$$\max_y \left\{y \cdot \left[\mathrm{E}(d|u=0) - p_1^{(0)} - \lambda y\right]\right\}. \tag{39}$$

Therefore, the optimal front-running response of the LENDER in case on no-borrowing can be approximated as follows:

$$y(0) \approx \arg\max_y \left\{y \cdot \left[\mathrm{E}(d|u=0) - p_1^{(0)} - \lambda y\right]\right\} = \frac{\mathrm{E}(d|u=0) - p_1^{(0)}}{2\lambda}. \tag{40}$$

In practice in equilibrium and $\mathrm{E}(d|u=0)$ does not depend strongly on σ_w because

$$\mathrm{E}[d|u=0] = \mathrm{E}[d|X(d) \geq 0] \approx \mathrm{E}[d|d \geq 0] = \sqrt{\frac{2}{\pi}} \approx 0.798.$$

Therefore the optimal choice $y(0)$ is driven by the interplay between $p_1^{(0)}$ and λ. As

[19] The interpretation of this quantity is "average" price sensitivity to order flow. Naturally, the notation is inherited form Kyle (1985).

σ_w shrinks towards zero, equilibrium λ increases, this causes a moderation of optimal $y(0)$ – induces the opposite effect of what we see in the equilibrium. On the other hand, as σ_w shrinks towards zero, the expected execution price in case of inaction $p_1^{(0)}$ also shrinks to towards zero and moves away from the expected value. That is, the return potential increases.

Overall, both return potential $[\mathrm{E}(d|u=0) - p_1^{(0)}]$ and price sensitivity $[\lambda]$ increase. The effect coming from return potential increase is stronger than the effect coming from price sensitivity increase. As a result, the optimal order size $y(0)$ increases as σ_w shrinks.

To understand the behavior of $p_1^{(0)}$ as σ_w shrinks I use the following approximation:

$$p_1^{(0)} = \mathrm{E}\left[G_{\boldsymbol{\sigma}}(X_{\boldsymbol{\sigma}}(d) + w + z) | X(d) \geq 0, w \geq 0\right] \approx G_{\boldsymbol{\sigma}}\left(\mathrm{E}[X_{\boldsymbol{\sigma}}(d) | X_{\boldsymbol{\sigma}}(d) \geq 0] + \mathrm{E}[w | w \geq 0]\right) \approx \tag{41}$$
$$\approx G_{\boldsymbol{\sigma}}\left(\sqrt{\frac{2}{\pi}}\left(k^+_{x/d}\sigma_d + \sigma_w\right)\right), \text{ where}$$

$k^+_{x/d}$ is the average slope of the function $X_{\boldsymbol{\sigma}}(d)$ for positive d, subscript "$\boldsymbol{\sigma}$" highlights that the equilibrium pricing rule $G(\cdot)$ and INFORMED TRADER's strategy $X(\cdot)$ depend on the model parameters $\boldsymbol{\sigma} = (\sigma_d, \sigma_w, \sigma_z)' = (1, \sigma_w, 1)'$.

This quantity shrinks as σ_w shrinks because

$$\overline{s}_0 = \sqrt{\frac{2}{\pi}}\left(k^+_{x/d}\sigma_d + \sigma_w\right)$$

is directly related to σ_w: as σ_w shrinks, the average public signal s in case of LENDER's inaction also shrinks.

Obviously, the value of

$$G_{\boldsymbol{\sigma}}\left(\sqrt{\frac{2}{\pi}}\left(k^+_{x/d}\sigma_d + \sigma_w\right)\right)$$

is not only driven by changes by changes in the argument (i.e. by changes in INFORMED TRADER's and UNINFORMED BORROWER's behaviors) but also by changes in the equilibrium MARKET MAKER's pricing rule $G_{\boldsymbol{\sigma}}(\cdot)$ as the subscript $\boldsymbol{\sigma}$ changes. Apparently, the effect coming from changes in the argument is stronger than the one coming from the changes in the subscript.

E Summary of continuous state space methodology

I look for a Nash equilibrium among three decision-making players:

1. The INFORMED TRADER "x";

2. The LENDER;

3. The MARKET MAKER.

I approach this task numerically. I use an iterative algorithm that can be interpreted as sequential mutual adjustments of strategies for each of the three agents. In particular, the sequence of adjustments is as follows:

1. Provide certain first-try conjectures for the strategies of the INFORMED TRADER and the LENDER.

2. Repeat the following adjustments until each agent's strategy converges to a certain fixed point:

   (a) MARKET MAKER chooses her rational strategy assuming the most up-to-date strategies of the INFORMED TRADER "x" and the LENDER,

   (b) LENDER adjusts her strategy assuming the most up-to-date strategies of the INFORMED TRADER "x" and the MARKET MAKER,

   (c) MARKET MAKER chooses her rational strategy assuming the most up-to-date strategies of the INFORMED TRADER "x" and the LENDER,

   (d) INFORMED TRADER "x" adjusts her strategy assuming the most up-to-date strategies of LENDER and MARKET MAKER.

The adjustment steps are described in detail and formalized below. For the sake of tractability, the strategies of the INFORMED TRADER, the LENDER, and the MARKET MAKERare restricted to certain families of functions.

In particular, I restrict the behavior of the INFORMED TRADER "x" such that she can only choose her order sizes $X(d)$ according to a piecewise-linear rule $x(\cdot)$ with prespecified nodes. From the economic perspective, this requirement is absolutely nonsensical: there is no economic reason why optimal choices made independently from each other for different values of d should constitute a piecewise-linear function of d. From the economic perspective, the decision made by the INFORMED TRADER "x" for each particular value of d is made irrespective of the optimal decisions made for all other decisions made for all other values of d. Nevertheless, if the true equilibrium $x(\cdot)$ is continuous, it is possible to approximate it with a piecewise-linear function. Moreover, the finer is the granularity, the more precise the approximation. Instead of maximizing the expected profits for each particular value of d and then finding the line of "best fit" (local linear approximation), I take a shortcut: I run a constrained optimization in the first place.

The original problem can be written as follows:

$$X(d) = \arg\max_{x} \mathrm{E}\left[\Pi^{Y(\cdot),G(\cdot)}_{\text{INFORMED TRADER ``}x\text{''}}(x,d,w,z)\middle|d\right] \tag{42}$$

Equivalently, I can define $X(\cdot)$ as follows

$$X(\cdot) = \arg\max_{X(\cdot)\in\mathcal{F}(\mathbb{R},\mathbb{R})} \mathrm{E}\left[\Pi^{X(\cdot),Y(\cdot),G(\cdot)}_{\text{INFORMED TRADER ``}x\text{''}}(d,w,z)\right]. \tag{43}$$

Economically, this equivalence means that optimizing the conditional expectations case-by-case (Eq. 42) is equivalent to maximizing unconditional expectation (Eq. 43), and follows from the law of iterated expectations.

In practice, instead of solving the unconstrained problem (Eq. 43) – which is purely notional – I solve an otherwise identical constrained maximization problem (Eq. 44).

$$X_{\text{constrained}}(\cdot) = \arg\max_{X(\cdot)\in\mathcal{X}} \mathrm{E}\left[\Pi^{X(\cdot),Y(\cdot),G(\cdot)}_{\text{INFORMED TRADER ``}x\text{''}}(d,w,z)\right], \text{ where} \tag{44}$$

$\mathcal{X}$ is the family of admissible functions.

Similarly, the original profit maximization problem of the LENDER is as follows. The original problem can be written as follows:

$$Y(u) = \arg\max_{y} \mathrm{E}\left[\Pi^{X(\cdot),G(\cdot)}_{\text{LENDER}}(y,d,w,z)\middle|u\right] \tag{45}$$

This problem is equivalent to the following problem.

$$Y(\cdot) = \arg\max_{X(\cdot)\in\mathcal{F}(\mathbb{R},\mathbb{R})} \mathrm{E}\left[\Pi^{X(\cdot),Y(\cdot),G(\cdot)}_{\text{INFORMED TRADER ``}x\text{''}}(d,w,z)\right]. \tag{46}$$

Finally, as a proxy for the unconstrained problem of the LENDER, (Eq. 46), I formally cast a tractable maximization problem for the LENDER.

$$Y_{\text{constrained}}(\cdot) = \arg\max_{Y(\cdot)\in\mathcal{Y}} \mathrm{E}\left[\Pi^{X(\cdot),Y(\cdot),G(\cdot)}_{\text{LENDER}}(d,w,z)\right], \text{ where} \tag{47}$$

$\mathcal{Y}$ as the family of admissible functions.

Definition E.1 (Sequential adjustment algorithm)**.** By sequential adjustment algorithm I mean the following steps.

1. I start with an arbitrary pair of functions $X^{(0)}(\cdot)$ and $Y^{(0)}(\cdot)$.

2. I iterate the adjustment steps $(i = 1, 2, 3, ...)$

(a) Find price-setting strategy $G(\cdot)$ that corresponds to the most recent trading strategy pair $(X(\cdot), Y(\cdot))$

$$G(\cdot)^{(i-1)} := \mathcal{B}_{\text{MARKET MAKER}}^{X^{(i-1)}(),\, Y^{(i-1)}()}(\cdot), \text{ where}$$

$\mathcal{B}_{\text{MARKET MAKER}}^{X(\cdot),\, Y(\cdot)}(\cdot)$ denotes the bayesian pricing rule [20] implemented by MARKET MAKER if she assumes that

- the INFORMED TRADER "x" follows strategy $X(\cdot)$,
- the LENDER follows strategy $Y(\cdot)$.

(b) Adjust $Y(\cdot)$ by using the relevant equilibrium condition: optimal trading strategy given the most recently computed strategy pair $(X(\cdot), G(\cdot))$,

$$Y^{(i)} := \alpha \operatorname*{arg\,max}_{Y^*(\cdot)\in\mathcal{Y}} \mathrm{E}\, \Pi_{\text{LENDER}}^{X^{(i-1)}(\cdot),\, Y^*(\cdot),\, G(\cdot)^{(i-1)}} + (1-\alpha) Y^{(i-1)}. \tag{48}$$

where $\alpha \in (0, 1)$ is a constant attenuation coefficient[21], $\Pi_{\text{LENDER}}^{X(\cdot),\, Y^*(\cdot),\, G(\cdot)}(d, w, z)$ denotes the LENDER's profits form trading, given that

- the INFORMED TRADER "x" follows strategy $X(\cdot)$,
- the LENDER follows strategy $Y^*(\cdot)$,
- the MARKET MAKER follows strategy $G(\cdot)$,
- particular values (d, w, z) are realized,

(c) Find price-setting strategy $G(\cdot)$ that corresponds to the most recent trading strategy pair $(X(\cdot), Y(\cdot))$

$$\hat{G}(\cdot)^{(i-1)} := \mathcal{B}_{\text{MARKET MAKER}}^{X^{(i-1)}(),\, Y^{(i)}()}(\cdot)$$

(d) Adjust $X(\cdot)$ by using the relevant equilibrium conditions: optimal trading strategy given the most recently computed strategy pair $(Y(\cdot), G(\cdot))$:

$$X^{(i)} := \alpha \operatorname*{arg\,max}_{X^*(\cdot)\in\mathcal{X}} \mathrm{E}\, \Pi_{\text{INFORMED TRADER}}^{X^*(\cdot),\, Y^{(i)}(\cdot),\, \hat{G}^{(i-1)}(\cdot)} + (1-\alpha) X^{(i-1)}, \text{ where} \tag{49}$$

$\Pi_{\text{INFORMED TRADER}}^{X^*(\cdot),\, Y(\cdot),\, G(\cdot)}(d, w, z)$ denotes the INFORMED TRADER's profits form trading given that

[20] Mathematically this function can be defined as:

$$\mathcal{B}_{X(\cdot),\, Y(\cdot)}^{\text{MARKET MAKER}}(\hat{s}) := \mathrm{E}\left(d \middle| X(d) + Y(X(d)^- + w^-) + z = \hat{s}\right), \forall \hat{s} \in \mathbb{R}.$$

[21] The bigger the coefficient the bigger the LENDER's adjustment.

- the INFORMED TRADER "x" follows strategy $X^*(\cdot)$,
- the LENDER follows strategy $Y(\cdot)$,
- the MARKET MAKER follows strategy $G(\cdot)$,
- particular values (d, w, z) realized.

I employ the following families of functions:

1. (a) $\mathcal{X} = \mathcal{X}_{\text{kink at 0}}$ is composed of all the functions that can be written in the following form:

$$X(d) = \begin{cases} x_0 + k_+ d \text{ for } d \geq 0 \\ x_0 + k_- d, \text{ otherwise.} \end{cases} \tag{50}$$

$\mathcal{X}_{\text{kink at 0}}$ is a set of continuous piecewise-linear functions with only one kink (at $d = 0$).

   (b) $\mathcal{Y} = \mathcal{Y}_{\text{jump at 0}}$ is composed of all functions that can be written in the following form:

$$Y(u) = \begin{cases} y_0 \text{ for } u = 0 \\ y_0' + ku \text{ for } u < 0. \end{cases} \tag{51}$$

$\mathcal{Y}_{\text{jump at 0}}$ is a set of functions that are defined only for $u \leq 0$, linear for negative u and with one discontinuity at $u = 0$.

2. (a) $\mathcal{X} = \mathcal{X}_{\text{jumps at nodes}}$ is composed of all functions that are linear between the nodes of a finite mesh $\{d_i\}$, see Assumption G.1. The values at the nodes do not matter because d is distributed continuously, and thus the probability of precisely hitting each individual node value is zero: $\mathrm{P}(d = d_i) = 0$.

   (b) $\mathcal{Y} = \mathcal{Y}_{\text{jumps at nodes}}$ is composed of all functions whose domain is $u \in (-\infty, 0]$ that are linear between the neighboring nodes of the grid $\{u_i\}$. At the nodes themselves, the function is defined as the arithmetic mean between the right and the left limits. The value at zero – where the right limit does not exist – may not be equal to the left limit at 0. As opposed to the strategy $X(\cdot)$, for the strategy $Y(\cdot)$ the individual values at the nodes may matter. It may be the case that $\mathrm{P}(u = u_i) > 0$ in some of the nodes u_i. In particular, this is very likely to be the case for the the rightmost node $u_n = 0$. This family satisfies the Assumption G.2.[22]

Definition E.2 ("Rough" equilibrium)**.** "Rough" equilibrium is a triple of mutually consistent strategies $(X(\cdot), Y(\cdot), G(\cdot))$ in which strategies $X(\cdot)$ and $Y(\cdot)$ are obliged to satisfy $X \in \mathcal{X}_{\text{kink at 0}}$ and $Y \in \mathcal{Y}_{\text{jump at 0}}$.

[22] The assumption is even slightly more general. In particular, the values at the nodes are not restricted to equal the arithmetic mean of the right and the left limits.

Remark E.1. Finding the "rough" equilibrium is essentially finding six coefficients of the strategies $X\cdot)$ and $Y(\cdot)$: two slopes and one intercept of the strategy $X(\cdot)$ plus one slope and one intercept of strategy $Y(\cdot)$, and plus the value $Y(0)$ (recall that Y may have a jump at 0).

Definition E.3 ("Transitional" ("Quasi") equilibrium based on conjecture (X_0, Y_0, G_0))**.** "Transitional" ("Quasi") equilibrium is a pair of strategies $(X(\cdot), Y(\cdot)) \in \mathcal{X} \times \mathcal{Y}$ that satisfies the following conditions:

1. Strategy $X(\cdot)$ is individually optimal for the INFORMED TRADER "x" if she assumes that
   - the LENDER follows strategy $Y_0(\cdot)$,
   - the MARKET MAKER follows strategy $G_0(\cdot)$.
2. Strategy $Y(\cdot)$ is individually optimal for the LENDER if she assumes that
   - the LENDER follows strategy $X_0(\cdot)$,
   - the MARKET MAKER follows strategy $G_0(\cdot)$.

Remark E.2. It is important to not that conjectured strategies X_0 and Y_0 do not have to come from the same family of functions. The essence of the "transitional" equilibrium is to make a transition (as the name implies) from some class of strategies $\mathcal{X}^* \times \mathcal{Y}^*$ to another class $\mathcal{X} \times \mathcal{Y}$.

In practice, I adopt the following procedure:

1. Run the sequential adjustment algorithm (see Def. E.1) subject to $X \in \mathcal{X}_{\text{kink at 0}}$, $Y \in \mathcal{Y}_{\text{jump at 0}}$ and find "rough" equilibrium within these families: $(X_{\text{rough}}(\cdot), Y_{\text{rough}}(\cdot), G_{\text{rough}}(\cdot))$
2. Use $(X_{\text{rough}}(\cdot), Y_{\text{rough}}(\cdot), G_{\text{rough}}(\cdot))$ as a conjecture, based on this conjecture find the "transitional" equilibrium $(X_{\text{quasi}}(\cdot), Y_{\text{quasi}}(\cdot)) \in \mathcal{X}_{\text{jumps at nodes}} \times \mathcal{Y}_{\text{jumps at nodes}}$.
3. Again, run the sequential algorithm, this time within a wider class of functions $X \in \mathcal{X}_{\text{jumps at nodes}}$, $Y \in \mathcal{Y}_{\text{jumps at nodes}}$ with starting point $(X_{\text{quasi}}(\cdot), Y_{\text{quasi}}(\cdot))$.

Luckily, there is convergence both times I apply the sequential adjustment algorithm.

E.1 Quasi-random simulations and maximization

Finally, I would like to make a few remarks on how I perform the maximization in each of the steps.

Both, from the perspective of the INFORMED TRADER "x" and the LENDER, I maximize the expectation of a random quantity that depends on three independent normal variables d, w, and z and on the choice of strategies $X(\cdot)$, $Y(\cdot)$, $G(\cdot)$. I use Quasi-Monte-Carlo method to compute these expectations.

1. I generate low-discrepancy-sequence (Halton sequence with parameters $(2, 3, 5)$) on a parallelepiped $[0, 1] \times [0, 1] \times [0, 1]$.

2. I rescale it to $[-k, k] \times [-k, k] \times [-k, k]$.

3. I truncate the "corners", by eliminating the observations that do not satisfy $\sqrt{d^2 + w^2 + z^2} \leq k$.[23]

4. I rescale the censored sample to $[-k\sigma_d, k\sigma_d] \times [-k\sigma_w, k\sigma_w] \times [-k\sigma_z, k\sigma_z]$ to obtain an ellipsis-shaped cloud.

5. I assign a weight to each observation this cloud that comes from the density of joint normal distribution. The weights are given by

$$\text{weight}_i = \frac{\phi\left(\frac{d_i}{\sigma_d}\right)\phi\left(\frac{w_i}{\sigma_w}\right)\phi\left(\frac{z_i}{\sigma_z}\right)}{\sum\limits_{i=1}^{N}\phi\left(\frac{d_i}{\sigma_d}\right)\phi\left(\frac{w_i}{\sigma_w}\right)\phi\left(\frac{z_i}{\sigma_z}\right)}, \text{ where} \tag{52}$$

   (d_i, w_i, z_i) are the observations generated at the previous steps.

Finally, the expectation $\mathrm{E}(f(d, w, z))$ is approximated as

$$\mathrm{E}(f(d, w, z)) \approx \sum_{i=1}^{N} \text{weight}_i \cdot f(d_i, w_i, z_i), \text{ where} \tag{53}$$

f is some quantity depending on the random outcome (d, w, z). In practice, f can mean either the realized profits of the INFORMED TRADER "x" or of the LENDER. Implicitly these quantities obviously depend on the strategies that each of the players follows: $(X(\cdot), Y(\cdot), G(\cdot))$.

E.2 Optimization

Instead of maximizing the expectations $(\mathrm{E}\,\Pi)$ in Eqs. 48 and 49, I maximize the approximation of these expectations following the approach in Eq. 53, subject to a relevant constraint ($x \in \mathcal{X}$ or $y \in \mathcal{Y}$). It is remarkable that when I preform maximizations subject to $x \in \mathcal{X}_{\text{jumps at nodes}}$ or to $y \in \mathcal{Y}_{\text{jumps at nodes}}$, I employ another trick that makes the maximization procedure more efficient and more reliable. In particular, I break down the overall

[23] This is roughly one half of the sample.

summation

$$\sum_{i=1}^{N} \text{weight}_i \cdot f(d_i, w_i, z_i) \tag{54}$$

into autonomous sub-summations. In the case of maximizing INFORMED TRADER's profits, the partitioning is made based on d_i. In each subgroup of indices I select those observations where $d_i \in (d_j^{\text{node}}, d_{j+1}^{\text{node}})$. If by chance some observations are exactly at the edge, then they are counted with half weight. Within each subgroup the function $X(\cdot)$ is linear. The maximization is separable: maximizing the overall summation is equivalent to maximizing the expectation and chose linear function withing each interval between the nodes. This trick reduces dimensionality: only two dimensions within each interval. As a by-product, it increases robustness because the coefficients pertaining to the outskirt regions of d, say $|d| > 3\sigma_d$ have a very small impact on the overall summation, and therefore the values of optimal strategy $X(\cdot)$ in the outskirts region may be poorly estimated by maximizing the overall expectation directly. The piecemeal maximization results in a more robust choice for the optimal order size for unlikely values of d, and therefore is more insightful for understanding the asymptotics of equilibrium strategies, which can be used to compare the predictions of the model with some benchmark cases for which closed-form solutions are known.

Finally, in order to find the numerical values of the coefficient, I use function "optim" from R "base" package, which in turn relies on the simplex Nead-Melder method.

E.3 Computation of function G

Since all the possible families $\mathcal{X}$, and $\mathcal{Y}$ are within the class of piecewise-linear function, I can use Corollary G.1.3, that is, I can express values of function G through functions $\exp(\cdot)$, $\Phi(\cdot)$, and $\Psi_{\cdot}(\cdot,\cdot)$ [24].

The computation of Ψ involves numeric integration, whose methodology is described in Internet Appendix. The final formula of $G(s)$ (see Eq. 108) is a fraction whose numerator and denominator tend to 0 as $|s| \to +\infty$. Moreover, when s takes on very unlikely values, the computation of g is challenged by the fact that numeric integration produces highly imprecise results because some of or many of the terms in summation become subnormal numbers, which leads to disproportionately high rounding errors. Eventually if the observed value of s is even more unlikely, the numerically computed value of the numerator in Eq. 108 can become 0 because all the terms in summation that approximates Ψ's turn 0 even though the true mathematical value of Ψ is far enough from 0.

[24] Ψ is the function whose values are de facto CDF of bivariate correlated normal random vector. Formally, Ψ is defined as $\Psi_k\left(\bar{\delta}, b\right) := \int\limits_{-\infty}^{\bar{\delta}} \psi_k\left(\delta, b\right) \mathbf{d}\delta$, where $\psi_k\left(\delta, b\right) := \Phi(k\delta + b)\phi(\delta)$.

In order to avoid imprecise results, I limit the potential values of s. In particular, I consider only s within an interval $[\underline{s}^*, \overline{s}^*]$ which I obtain as an extension of

$$[\underline{s}, \overline{s}] = \text{Range}\{s_0(d_i, w_i) + z_i\}_{i \in \overline{1,n}}. \tag{55}$$

I perform the extension as follows:

$$[\underline{s}^*, \overline{s}^*] = [\underline{s} - s_{\text{ext. coef.}}(\overline{s} - \underline{s}), \overline{s} + s_{\text{ext. coef.}}(\overline{s} - \underline{s})], \text{ which means that} \tag{56}$$

I move the margins away from the middle of the interval by $s_{\text{ext. coef.}}$ share of the length of the interval. I compute g within this interval and $[\underline{s}^*, \overline{s}^*]$ at all the nodes of a uniform partition including the ends of the interval.

$$s_{(i)} = \underline{s}^* + \frac{i(\overline{s}^* - \underline{s}^*)}{s_{\text{granularity}}} \tag{57}$$

Once, the values of G are computed at the nodes of the mesh [this is done only once for each conjectured function G], g is computed using linear interpolation between the nodes, only the two closest nodes are taken into consideration.

F Updating beliefs based on obseving the borrowed quantity

In this appendix, I analyze how the broker updates his beliefs about the probability distribution of (d, w) once he observes the signal $u \equiv X^-(d) + w^-$.

Suppose that the broker knows the functional form of X. $X(\cdot)$ is assumed to be a non-decreasing continuous function continuously differentiable for $d < d_{\text{short}}$ except perhaps a finite number of values, where the $X(\cdot)$ has kinks. Variables d and w are independent and continuously distributed around zero with well-behaved densities (here I do not require normality, however, in the body of the paper I have this assumption).

For $X(d) < 0 \iff d < d_{\text{short}}$, I have a simple relation between the densities of d and $X(d)$:

$$f_x(\hat{x}) = \frac{f_d\left(\hat{d}\right)}{X'\left(\hat{d}\right)}, \tag{58}$$

where $\hat{x} = X\left(\hat{d}\right)$.

$X^{-}(d)$ is distributed as follows:

$$X^{-}(d) \sim \begin{cases} 0 \text{ with probability } \pi = \mathrm{P}(X(d) \geq 0) = \mathrm{P}(d \geq d_{\text{short}}), \\ \hat{x} < 0 \text{ with density } f_x(\hat{x}) = \frac{f_d(\hat{d})}{X'(\hat{d})} = \frac{f_d(X^{-1}(\hat{x}))}{X'(X^{-1}(\hat{x}))} \end{cases} \tag{59}$$

Similarly, w^{-} is distributed as follows:

$$w^{-} \sim \begin{cases} 0 \text{ with probability } \rho = \mathrm{P}(w \geq 0) \\ \hat{w} < 0 \text{ with density } f_w(\hat{w}) \end{cases}, \tag{60}$$

where $\rho = \frac{1}{2}$ in case if w being continuously and symmetrically distributed.

I am interested in the joint distribution of (d, w) conditionally on $X^{-}(d) + w^{-} \equiv u = \hat{u}$. This conditional distribution depends on the level of $\hat{u}$. I treat separately two cases: strictly negative $\hat{u}$ and $\hat{u} = 0$.

F.1 Case of non-zero borrowing

I start with the negative case. In order to make sure that conditional probabilities well-defined, I weaken the constraint, considering the condition

$$x^{-} + w^{-} \equiv u \in B_{\frac{\varepsilon}{2}}(\hat{u}), \tag{61}$$

where $B_{\frac{\varepsilon}{2}}(\hat{u})$ denotes an open ball of radius $\frac{\varepsilon}{2}$ and ε is arbitrarily small.

In the plane (x, w) these areas look as in Panel a of Figure 31. In the plane (d, w) these areas look as in Panel b of Figure 31.

I am interested in conditional distribution of (d, w) given $u \in B_{\frac{\varepsilon}{2}}(\hat{u})$. I start analyzing this distribution from looking at the probability of $b \in B_{\frac{\varepsilon}{2}}\left(\hat{b}\right) \iff u \in B_{\frac{\varepsilon}{2}}(\hat{u})$.

$$\mathrm{P}\left[u \in B_{\frac{\varepsilon}{2}}(\hat{u})\right] = \varepsilon \cdot \left[\int_{\hat{u}}^{0} f_x(\hat{x}) f_w(\hat{u} - \hat{x})\, \mathbf{d}\hat{x} + \rho f_x(\hat{u}) + \pi f_w(\hat{u})\right] + o(\varepsilon), \tag{62}$$

where the quantity within the square brackets is the density of u evaluated at $\hat{u}$:

$$\int_{\hat{u}}^{0} f_x(\hat{x}) f_w(\hat{u} - \hat{x})\, \mathbf{d}\hat{x} + \rho f_x(\hat{u}) + \pi f_w(\hat{u}) = f_u(\hat{u}) \tag{63}$$

Note that u is not entirely a continuous random variable, u is distributed as follows:

$$u \sim \begin{cases} \hat{u} < 0 \text{ with density } f_u(\hat{u}) \\ 0 \text{ with probability } \pi\rho = \mathrm{P}(X(d) \geq 0) \cdot \mathrm{P}(w \geq 0) \end{cases}, \tag{64}$$

Joint distribution $(x^-, w^-)|u = \hat{u}(< 0)$ is as follows

$$(x^-, w^-)|b = \hat{b} \sim \begin{cases} x^- \text{ with density } f^{\text{both short}}_{x^-|u=\hat{u}}(\hat{x}) = \frac{f_x(\hat{x}) f_w(\hat{u}-\hat{x})}{f_u(\hat{u})} \text{ for } \hat{x} \in (\hat{u}, 0), w^- = \hat{u} - x^- < 0, \\ \left(\hat{b}, 0\right) \text{ with probability } \frac{\rho f_x(\hat{u})}{f_u(\hat{u})}, \\ \left(0, \hat{b}\right) \text{ with probability } \frac{\pi f_w(\hat{u})}{f_u(\hat{u})}. \end{cases} \tag{65}$$

From the point of view of the joint distribution of (d, w), the second and the third cases are further split into a continuum of possibilities. If $w^- = 0$, then w can be any non-negative number. Similarly, if $x^- = 0$, x can be any non-negative number while d can be any number with $d \geq d_{\text{short}}$.

Finally, the joint distribution is as follows:

$$(d, w)|u = \hat{u} \sim \begin{cases} d \text{ with density } f^{\text{both short}}_{d|b=\hat{b}}\left(\hat{d}\right) = \frac{f_d(\hat{d}) f_w(\hat{u} - X(\hat{d}))}{f_u(\hat{u})} \text{ for } \hat{d} \in \left(X^{-1}(\hat{u}), d_{\text{short}}\right), w = \hat{u} - X\left(\hat{d}\right) < 0, \\ \left(X^{-1}(\hat{u}), w\right) \text{ where } w \text{ has the density of } f^{\text{informed shorts}}_{w|b=\hat{b}}(\hat{w}) = \frac{f_w(\hat{w}) f_x(\hat{u})}{f_u(\hat{u})} = \frac{f_w(\hat{w}) f_d\left(X^{-1}(\hat{u})\right)}{f_u(\hat{u}) X'(X^{-1}(\hat{u}))} \text{ for } w \geq 0, \\ (d, \hat{u}) \text{ where } d \text{ has the density of } f^{\text{noise shorts}}_{d|b=\hat{b}}\left(\hat{d}\right) = \frac{f_d(\hat{d}) f_w(\hat{u})}{f_u(\hat{u})} \text{ for } d \geq d_{\text{short}}. \end{cases} \tag{66}$$

The first case in Eq. 66 corresponds to the downward sloping segments of the isoquants in Panel b of Figure 31. The second case corresponds to the vertical parts of the same isoquants, while the third case corresponds to the horizontal parts of those isoquants.

Since the lender is particularly interested in the expected value of d, I project the distribution given in Eq. 66 onto d-axis in order to obtain the conditional distribution of d.

$$d|u = \hat{u} \sim \begin{cases} X^{-1}(\hat{u}), \text{ with probability } p^{\text{informed shorts}}_{d|b=\hat{b}}\left(X^{-1}(\hat{u})\right) = \frac{\rho f_d\left(X^{-1}(\hat{u})\right)}{f_u(\hat{u}) X'(X^{-1}(\hat{u}))}, \\ \text{with density } f^{\text{both short}}_{d|b=\hat{b}}\left(\hat{d}\right) = \frac{f_d(\hat{d}) f_w\left(\hat{u} - X(\hat{d})\right)}{f_u(\hat{u})} \text{ for } \hat{d} \in \left(X^{-1}(\hat{u}), d_{\text{short}}\right), \\ \text{with density } f^{\text{noise shorts}}_{d|b=\hat{b}}\left(\hat{d}\right) = \frac{f_d(\hat{d}) f_w(\hat{u})}{f_u(\hat{u})} \text{ for } \hat{d} \geq d_{\text{short}}. \end{cases} \tag{67}$$

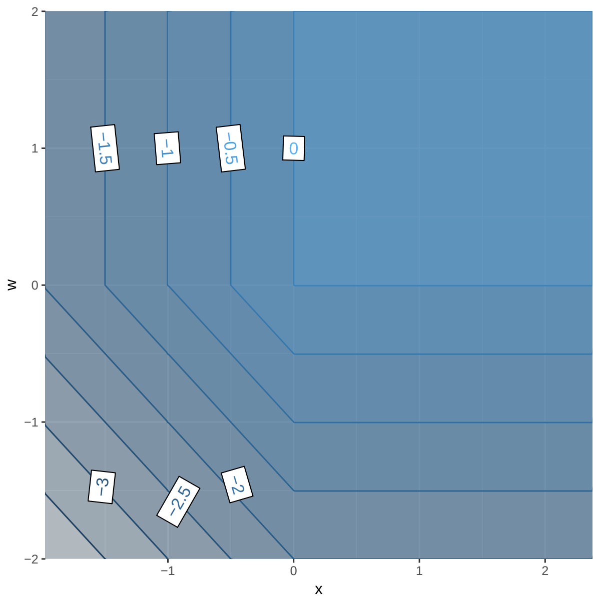
The definition of derivative in this case should be extended:

Definition F.1.

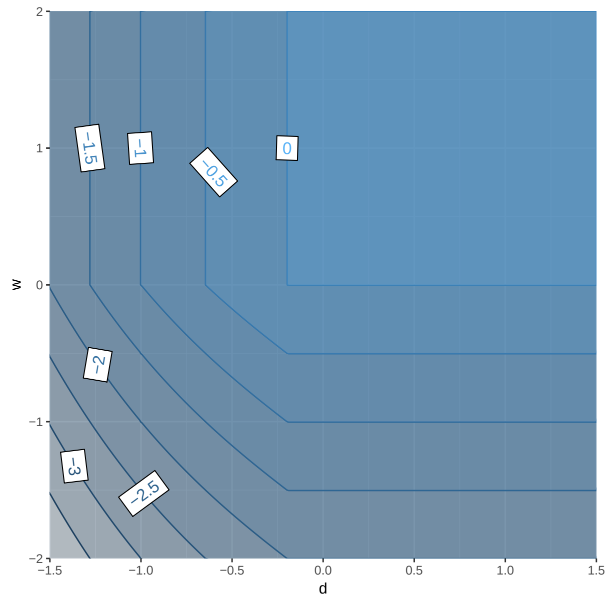
$$\psi'(h) = \lim_{dh \to 0} \frac{\psi\left(h + \frac{dh}{2}\right) - \psi\left(h - \frac{dh}{2}\right)}{dh}.$$

If function $X(\cdot)$ has kinks, then the generalized derivative is also well-defined at the kinks.

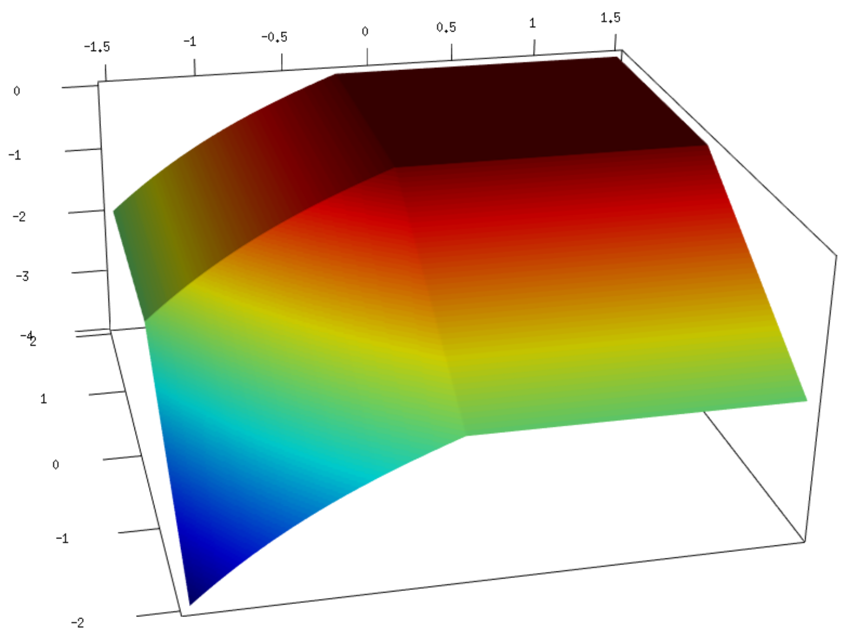
Figure 31: Signal $u = x^- + w^-$ (negative of borrowing volume b)



(a) as a function of x and w



(b) as a function of d and w



(c) as a function of d and w

Thus, the conditional expectation $\mathrm{E}\left[d|u=\hat{u}\right]$ is as follows:

$$X^{-1}(\hat{u})\cdot\frac{\rho f_d\left(X^{-1}\left(\hat{u}\right)\right)}{f_u\left(\hat{u}\right)X'\left(X^{-1}\left(\hat{u}\right)\right)}+\int\limits_{X^{-1}(\hat{u})}^{d_{\text{short}}}\hat{d}\cdot\frac{f_d\left(\hat{d}\right)f_w\left(\hat{u}-X\left(\hat{d}\right)\right)}{f_u\left(\hat{u}\right)}\mathbf{d}\hat{d}+\int\limits_{d_{\text{short}}}^{+\infty}\hat{d}\cdot\frac{f_d\left(\hat{d}\right)f_w\left(\hat{u}\right)}{f_u\left(\hat{u}\right)}\mathbf{d}\hat{d}. \tag{68}$$

F.2 Case of zero borrowing

So far, I have been dealing with the case where the broker observes strictly positive borrowing demand $u<0$ ($b>0$). The remaining case ($u=0$) corresponds to the the upper-right shaded area in Panel b of Figure 31 and to the flat part of the surface depicted in Panel c of Figure 31. In this case, d and w are conditionally independent and their densities are as follows:

$$f_{d|u=0}\left(\hat{d}\right)=f_{d|d\geq d_{\text{short}}}\left(\hat{d}\right)=\begin{cases}f_d\left(\hat{d}\right)\Big/\pi & \text{for } \hat{d}\geq d_{\text{short}},\\ 0, & \text{otherwise,}\end{cases} \tag{69}$$

$$f_{w|b=0}\left(\hat{w}\right)=f_{d|w\geq 0}\left(\hat{w}\right)=\begin{cases}f_w\left(\hat{w}\right)/\rho & \text{for } \hat{w}\geq 0,\\ 0, & \text{otherwise.}\end{cases} \tag{70}$$

Thus, the conditional expectation $\mathrm{E}\left[d|u=0\right]$ is as follows:

$$\frac{\int\limits_{d_{\text{short}}}^{+\infty}\hat{d}\cdot f_d\left(\hat{d}\right)\mathbf{d}\hat{d}}{\pi} \tag{71}$$

G Updating beliefs based on observing overall order flow

By assumption, d, w and z are continuously distributed random variables with densities defined everywhere apart from at most a finite number of values. $X(\cdot)$ and $Y(\cdot)$ are continuously differentiable functions apart from at most a finite number of values.

The market maker sets the price of the asset, based on the conditional distribution of d after observing

$$S_0(d,w)+z=:S(d,w,z)=:s=\hat{s},\text{ where} \tag{72}$$

$$S_0(d,w):=X(d)+w+Y(X^-(d)+w^-) \tag{73}$$

Theorem G.1. Joint density $f_{d,s}$ is given by:

$$f_{d,s}\left(\hat{d},\hat{s}\right) = f_d\left(\hat{d}\right)\int\limits_{-\infty}^{+\infty} f_w\left(\hat{w}\right) f_z\left(\hat{s} - S_0\left(\hat{d},\hat{w}\right)\right)\mathbf{d}\hat{w} \tag{74}$$

Proof.

$$\mathrm{P}\left[d \in B_{\frac{\varepsilon}{2}}\left(\hat{d}\right), s \in B_{\frac{\varepsilon}{2}}\left(\hat{s}\right)\right] =$$

$$\frac{1}{\varepsilon}\int\limits_{-\infty}^{+\infty} \mathrm{P}\left[d \in B_{\frac{\varepsilon}{2}}\left(\hat{d}\right), w \in B_{\frac{\varepsilon}{2}}\left(\hat{w}\right), s \in B_{\frac{\varepsilon}{2}}\left(\hat{s}\right)\right]\mathbf{d}\hat{w} =$$

$$\frac{1}{\varepsilon}\int\limits_{-\infty}^{+\infty} \mathrm{P}\left[d \in B_{\frac{\varepsilon}{2}}\left(\hat{d}\right), w \in B_{\frac{\varepsilon}{2}}\left(\hat{w}\right), z \in B_{\frac{\varepsilon}{2}}\left(\hat{s} - S_0\left(\hat{d},\hat{w}\right)\right)\right]\mathbf{d}\hat{w} =$$

$$\varepsilon^2\int\limits_{-\infty}^{+\infty} f_d\left(\hat{d}\right) f_w\left(\hat{w}\right) f_z\left(\hat{s} - S_0\left(\hat{d},\hat{w}\right)\right)\mathbf{d}\hat{w}.$$

□

Corollary G.1.1. Conditional density $f_{d|s}$ is given as follows:

$$f_{d|s}\left(\hat{d},\hat{s}\right) = \frac{f_{d,s}\left(\hat{d},\hat{s}\right)}{f_s\left(\hat{s}\right)} = \frac{f_{d,s}\left(\hat{d},\hat{s}\right)}{\int\limits_{-\infty}^{+\infty} f_{d,s}\left(\hat{d},\hat{s}\right)\mathbf{d}\hat{d}} = \frac{f_d\left(\hat{d}\right)\int\limits_{-\infty}^{+\infty} f_w\left(\hat{w}\right) f_z\left(\hat{s} - S_0\left(\hat{d},\hat{w}\right)\right)\mathbf{d}\hat{w}}{\int\limits_{-\infty}^{+\infty} f_d\left(\hat{d}\right)\int\limits_{-\infty}^{+\infty} f_w\left(\hat{w}\right) f_z\left(\hat{s} - S_0\left(\hat{d},\hat{w}\right)\right)\mathbf{d}\hat{w}\mathbf{d}\hat{d}}. \tag{75}$$

Corollary G.1.2. Conditional expectation $\mathrm{E}\left[d|s=\hat{s}\right]$ is given as follows:

$$\mathrm{E}\left[d|s=\hat{s}\right] = \frac{\int\limits_{-\infty}^{+\infty}\hat{d}\cdot f_{d,s}\left(\hat{d},\hat{s}\right)\mathbf{d}\hat{d}}{\int\limits_{-\infty}^{+\infty} f_{d,s}\left(\hat{d},\hat{s}\right)\mathbf{d}\hat{d}} = \frac{\int\limits_{-\infty}^{+\infty}\hat{d}\cdot f_d\left(\hat{d}\right)\int\limits_{-\infty}^{+\infty} f_w\left(\hat{w}\right) f_z\left(\hat{s} - S_0\left(\hat{d},\hat{w}\right)\right)\mathbf{d}\hat{w}\mathbf{d}\hat{d}}{\int\limits_{-\infty}^{+\infty} f_d\left(\hat{d}\right)\int\limits_{-\infty}^{+\infty} f_w\left(\hat{w}\right) f_z\left(\hat{s} - S_0\left(\hat{d},\hat{w}\right)\right)\mathbf{d}\hat{w}\mathbf{d}\hat{d}}. \tag{76}$$

I additionally impose certain assumtions on the functional forms of $X(\cdot)$ and $Y(\cdot)$ and on the prior distribitions of d, w, and z.

In particular, I assume the following about $X(\cdot)$.

Assumption G.1.

$$X(d) = X^{(i)}(d) \text{ for } d \in I_d^i, \tag{77}$$

$$X^{(i)}(d) = k_x^i d + b_x^i, \tag{78}$$

$$I_d^i = (d_i, d_{i+1}) \text{ for } i \in 1, ..., N_x,\ d_1 = -\infty,\ d_{N_x+1} = +\infty.^{25} \tag{79}$$

Besides, I assume the following about $Y(\cdot)$.

Assumption G.2.

$$Y(u) = Y^{(i)}(u) \text{ for } u \in I_u^i, \tag{80}$$

$$Y^{(i)}(u) = k_y^i d + b_y^i, \tag{81}$$

$$I_u^i = (u_i, u_{i+1}) \text{ and } I_u^{-i} = \{u_{i+1}\} = [u_{i+1}, u_{i+1}] \tag{82}$$

$$\text{for } i \in 1, ..., N_y - 1,\ d_1 = -\infty,\ d_{N_y+1} = 0.^{26}$$

Obviously, slope coefficients are redundant for the degenerate intervals I_u^{-i}. Nevertheless, I would like to keep the generality of notation. For this reason I specify $k_y^{-i} = 0$.

Regarding the prior distributions, I assume the following.

Assumption G.3. $d \sim N(0, \sigma_d^2)$, $z \sim N(0, \sigma_z^2)$, and $w \sim N(0, \sigma_w^2)$.

Corollary G.1.3. Under these three assumptions the conditional expectation in the previous Corollary can be computed according to the following steps.

Consider separately numerator and denominator that both can be written as:

$$\int_{-\infty}^{+\infty} \hat{d}^{\nu} \cdot f_d\left(\hat{d}\right) \int_{-\infty}^{+\infty} f_w\left(\hat{w}\right) f_z\left(\hat{s} - S_0\left(\hat{d}, \hat{w}\right)\right) \mathbf{d}\hat{w}\mathbf{d}\hat{d}, \text{ where } \nu \in \{0, 1\} \tag{83}$$

1. Split the integration area $\mathbb{R}^2$ into a finite set of non-overlapping and closure-exhaustive trapezoids [tiles] within which function s_0 is linear in $\hat{d}$ and $\hat{w}$.[27] I denote this finite set of tiles as $\Theta = \{\tau\}$, where each tile τ ia a subset of $\mathbb{R}^2$ for $\tau \in \Theta$. Each of these tiles has four sides, two of them are vertical: $\underline{d}(\tau) < d < \overline{d}(\tau)$, while the other two are either horizontal or sloped straight lines: $\underline{w}_\tau(d) < w < \overline{w}_\tau(d)$, where $\underline{w}_\tau(d) = \underline{K}_{w/d}(\tau) \cdot d + \underline{B}_{w/d}(\tau)$ and $\overline{w}_\tau(d) = \overline{K}_{w/d}(\tau) \cdot d + \overline{B}_{w/d}(\tau)$. Any of the sides can degenerate to $\pm\infty$. If, say, the upper boundary $\overline{w}_\tau(\cdot)$ degenerates to $+\infty$, then I say that $\overline{K}_{w/d}(\tau) = 0$ and $\overline{B}_{w/d}(\tau) = +\infty$.

[25] For pracrical purposes the openness of the intervals is innocuous. In the computation of the repeated integrals $\iint \cdots \mathbf{d}\hat{d}\mathbf{d}\hat{w}$ so it is acceptable that a function is not defined at a finite number of straight lines $\{\{d_i\} \times \mathbb{R}\}_{(i)}$.

[26] I deliberately allow for discontinuities. If function $X(\cdot)$ has negative shelves, then values of function $Y(\cdot)$ in individual points may matter: it cannot be defined up to a set of measure 0. Minus ("$-$") in front of "i" has a purely symbolical meaning. It is only meant for distinguishing intervals from the points between them.

[27] The description of the algorithm is provided in Internet Appendix.

2. For each tile τ, determine the slope and intercept coefficients that are effective for functions $X(\cdot)$, $X^{-}(\cdot)$, $Y(\cdot)$ and the slope of w^{-} as a function of w.

$$k_x(\tau), b_x(\tau), k_{x^-}(\tau), b_{x^-}(\tau), k_y(\tau), b_y(\tau), \text{ and } \mathbb{I}_{w<0}(\tau) \tag{84}$$

3. For each tile, explicitly articulate the linear form of s_0:

$$\begin{aligned} s_0\left(\hat{d}, \hat{w}\right) &= \underbrace{b_x(\tau) + k_y(\tau) b_{x^-}(\tau) + b_y(\tau)}_{:=B_{s_0}(\tau)} + \underbrace{\left[k_x(\tau) + k_y(\tau) k_{x^-}(\tau)\right]}_{:=K_{s_0/d}(\tau)} \cdot \hat{d} + \underbrace{\left[1 + \mathbb{I}_{w<0}(\tau) k_y(\tau)\right]}_{:=K_{s_0/w}(\tau)} \cdot \hat{w} \\ &= B_{s_0}(\tau) + K_{s_0/d}(\tau) \cdot \hat{d} + K_{s_0/w}(\tau) \cdot \hat{w} \end{aligned} \tag{85}$$

4. Provide a closed-form integral over $\hat{w}$ (the internal one) within a particular tile intersected with a particular value of $\hat{d}$.

$$\begin{aligned} &\int_{\underline{K}_{w/d}(\tau)\hat{d}+\underline{B}_{w/d}(\tau)}^{\overline{K}_{w/d}(\tau)\hat{d}+\overline{B}_{w/d}(\tau)} f_w(\hat{w}) f_z\left(\hat{s} - s_0\left(\hat{d}, \hat{w}\right)\right) \mathbf{d}\hat{w} = \\ &\int_{\underline{K}_{w/d}(\tau)\hat{d}+\underline{B}_{w/d}(\tau)}^{\overline{K}_{w/d}(\tau)\hat{d}+\overline{B}_{w/d}(\tau)} f_w(\hat{w}) f_z\left(\hat{s} - B_{s_0}(\tau) - K_{s_0/d}(\tau) \cdot \hat{d} - K_{s_0/w}(\tau) \cdot \hat{w}\right) \mathbf{d}\hat{w} = \\ &\frac{\exp\left[-\frac{\left(\hat{s} - B_{s_0}(\tau) - K_{s_0/d}(\tau) \cdot \hat{d}\right)^2}{2\Sigma^2(\tau)}\right]}{\sqrt{2\pi}\Sigma(\tau)} \cdot \Phi(\Delta)\Bigg|_{\frac{\underline{K}_{w/d}(\tau)\hat{d}+\underline{B}_{w/d}(\tau)-\lambda(\tau)\cdot\left(\hat{s}-B_{s_0}(\tau)-K_{s_0/d}(\tau)\cdot\hat{d}\right)}{\sigma(\tau)}}^{\frac{\overline{K}_{w/d}(\tau)\hat{d}+\overline{B}_{w/d}(\tau)-\lambda(\tau)\cdot\left(\hat{s}-B_{s_0}(\tau)-K_{s_0/d}(\tau)\cdot\hat{d}\right)}{\sigma(\tau)}} \equiv \\ &\frac{\exp\left[-\frac{\left(\hat{s} - B_{s_0}(\tau) - K_{s_0/d}(\tau) \cdot \hat{d}\right)^2}{2\Sigma^2(\tau)}\right]}{\sqrt{2\pi}\Sigma(\tau)} \times \\ &\times \left[\Phi\left(\frac{\overline{K}_{w/d}(\tau)\hat{d} + \overline{B}_{w/d}(\tau) - \lambda(\tau) \cdot \left(\hat{s} - B_{s_0}(\tau) - K_{s_0/d}(\tau) \cdot \hat{d}\right)}{\sigma(\tau)}\right) - \right. \\ &\left. \Phi\left(\frac{\underline{K}_{w/d}(\tau)\hat{d} + \underline{B}_{w/d}(\tau) - \lambda(\tau) \cdot \left(\hat{s} - B_{s_0}(\tau) - K_{s_0/d}(\tau) \cdot \hat{d}\right)}{\sigma(\tau)}\right)\right], \text{ where} \end{aligned} \tag{86}$$

$$\Sigma(\tau) = \sqrt{K_{s_0/w}(\tau)^2 \sigma_w^2 + \sigma_z^2}, \tag{87}$$

$$\sigma(\tau) = \frac{\sigma_w \sigma_z}{\Sigma(\tau)}, \tag{88}$$

$$\lambda(\tau) = \frac{K_{s_0/w}(\tau)\sigma_w^2}{\Sigma^2(\tau)}, \tag{89}$$

5. Rewrite the internal integral with a new notation:

$$\int\limits_{\underline{K}_{w/d}(\tau)\hat{d}+\underline{B}_{w/d}(\tau)}^{\overline{K}_{w/d}(\tau)\hat{d}+\overline{B}_{w/d}(\tau)} f_w(\hat{w}) f_z\left(\hat{s} - s_0\left(\hat{d}, \hat{w}\right)\right) \mathbf{d}\hat{w} =$$

$$\frac{\exp\left[-\frac{\left(\hat{s}-B_{s_0}(\tau)-K_{s_0/d}(\tau)\cdot\hat{d}\right)^2}{2\Sigma^2(\tau)}\right]}{\sqrt{2\pi}\Sigma(\tau)} \times \left[\Phi\left(\overline{\varkappa}(\tau)\cdot\hat{d} + \overline{\beta}(\tau)\right) - \Phi\left(\underline{\varkappa}(\tau)\cdot\hat{d} + \underline{\beta}(\tau)\right)\right], \text{ where} \tag{90}$$

$$\underline{\varkappa}(\tau) = \frac{\underline{K}_{w/d}(\tau) + \lambda(\tau)K_{s_0/d}(\tau)}{\sigma(\tau)} = \frac{\underline{K}_{w/d}(\tau)}{\sigma(\tau)} + \frac{K_{s_0/w}(\tau)K_{s_0/d}(\tau)\sigma_w}{\Sigma(\tau)\sigma_z}, \tag{91}$$

$$\overline{\varkappa}(\tau) = \frac{\overline{K}_{w/d}(\tau) + \lambda(\tau)K_{s_0/d}(\tau)}{\sigma(\tau)} = \frac{\overline{K}_{w/d}(\tau)}{\sigma(\tau)} + \frac{K_{s_0/w}(\tau)K_{s_0/d}(\tau)\sigma_w}{\Sigma(\tau)\sigma_z}, \tag{92}$$

$$\underline{\beta}(\tau) = \frac{\underline{B}_{w/d}(\tau)}{\sigma(\tau)} + \frac{K_{s_0/w}(\tau)\sigma_w}{\Sigma(\tau)\sigma_z}\left(B_{s_0}(\tau) - \hat{s}\right), \tag{93}$$

$$\overline{\beta}(\tau) = \frac{\overline{B}_{w/d}(\tau)}{\sigma(\tau)} + \frac{K_{s_0/w}(\tau)\sigma_w}{\Sigma(\tau)\sigma_z}\left(B_{s_0}(\tau) - \hat{s}\right). \tag{94}$$

6. Use the simplified version of the inner integral (with respect to $\hat{w}$) obtained at the previous step to plug into the outer integral.

$$\mathcal{I}^{(\nu)}(\tau) := \int\limits_{\underline{d}(\tau)}^{\overline{d}(\tau)} \hat{d}^{\nu}\cdot f_d\left(\hat{d}\right) \int\limits_{\underline{K}_{w/d}(\tau)\hat{d}+\underline{B}_{w/d}(\tau)}^{\overline{K}_{w/d}(\tau)\hat{d}+\overline{B}_{w/d}(\tau)} f_w(\hat{w}) f_z\left(\hat{s} - s_0\left(\hat{d}, \hat{w}\right)\right) \mathbf{d}\hat{w}\mathbf{d}\hat{d} =$$

$$\frac{1}{\sqrt{2\pi}\Sigma(\tau)} \int\limits_{\underline{d}(\tau)}^{\overline{d}(\tau)} \hat{d}^{\nu}\cdot f_d\left(\hat{d}\right) \exp\left[-\frac{\left(\hat{s}-B_{s_0}(\tau)-K_{s_0/d}(\tau)\cdot\hat{d}\right)^2}{2\Sigma^2(\tau)}\right] \times \tag{95}$$

$$\times\left[\Phi\left(\overline{\varkappa}(\tau)\cdot\hat{d} + \overline{\beta}(\tau)\right) - \Phi\left(\underline{\varkappa}(\tau)\cdot\hat{d} + \underline{\beta}(\tau)\right)\right]\mathbf{d}\hat{d}$$

7. Rewrite the integral by transforming

$$f_d\left(\hat{d}\right)\exp\left[-\frac{\left(\hat{s}-B_{s_0}(\tau)-K_{s_0/d}(\tau)\cdot\hat{d}\right)^2}{2\Sigma^2(\tau)}\right]=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_d}\exp\left[-\frac{1}{2}\left(\frac{\hat{d}^2}{\sigma_d^2}+\frac{\left(\hat{s}-B_{s_0}(\tau)-K_{s_0/d}(\tau)\cdot\hat{d}\right)^2}{\Sigma^2(\tau)}\right)\right]=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_d}\exp\left[-\frac{1}{2}\left(\frac{\hat{d}^2}{\sigma_d^2}+\frac{\left(K_{s_0/d}(\tau)\cdot\hat{d}-\tilde{s}(\tau)\right)^2}{\Sigma^2(\tau)}\right)\right]=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_d}\exp\left[-\frac{1}{2}\left(\frac{\left(\hat{d}-\tilde{\lambda}(\tau)\tilde{s}(\tau)\right)^2}{\tilde{\sigma}^2(\tau)}+\frac{\tilde{s}^2(\tau)}{\tilde{\Sigma}^2(\tau)}\right)\right]=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_d}\exp\left[-\frac{\left(\hat{d}-\tilde{\lambda}(\tau)\tilde{s}(\tau)\right)^2}{2\tilde{\sigma}^2(\tau)}\right]\exp\left[-\frac{\tilde{s}^2(\tau)}{2\tilde{\Sigma}^2(\tau)}\right],\ \text{where}$$

$$\tilde{s}(\tau)=\hat{s}-B_{s_0}(\tau)\tag{96}$$

$$\tilde{\Sigma}^2(\tau)=\Sigma^2(\tau)+\sigma_d^2K_{s_0/d}^2(\tau)\tag{97}$$

$$\tilde{\sigma}^2(\tau)=\Sigma^2(\tau)\sigma_d^2/\tilde{\Sigma}^2(\tau)\tag{98}$$

$$\tilde{\lambda}(\tau)=\sigma_dK_{s_0/d}/\tilde{\Sigma}^2(\tau)\tag{99}$$

In particular, obtain

$$\begin{aligned}\mathcal{I}^{(\nu)}(\tau)=\frac{1}{\sqrt{2\pi}\Sigma(\tau)}\int\limits_{\underline{d}(\tau)}^{\overline{d}(\tau)}\hat{d}^{\nu}\cdot\frac{1}{\sqrt{2\pi}\sigma_d}\exp\left[-\frac{\left(\hat{d}-\tilde{\lambda}(\tau)\tilde{s}(\tau)\right)^2}{2\tilde{\sigma}^2(\tau)}\right]\exp\left[-\frac{\tilde{s}^2(\tau)}{2\tilde{\Sigma}^2(\tau)}\right]\times\\ \times\left[\Phi\left(\overline{\varkappa}(\tau)\cdot\hat{d}+\overline{\beta}(\tau)\right)-\Phi\left(\underline{\varkappa}(\tau)\cdot\hat{d}+\underline{\beta}(\tau)\right)\right]\mathbf{d}\hat{d}\end{aligned}\tag{100}$$

8. Standardize the integral by applying a change of variable:

$$\delta=\frac{\hat{d}-\tilde{\lambda}(\tau)\tilde{s}(\tau)}{\tilde{\sigma}(\tau)}\iff\hat{d}=\tilde{\lambda}(\tau)\tilde{s}(\tau)+\tilde{\sigma}(\tau)\delta\Rightarrow\mathbf{d}\hat{d}=\tilde{\sigma}\mathbf{d}\delta\tag{101}$$

$$\mathcal{I}^{(\nu)}(\tau) = \frac{\phi\left[\frac{\tilde{s}(\tau)}{\tilde{\Sigma}(\tau)}\right]}{\tilde{\Sigma}(\tau)} \times \int\limits_{\underline{\delta}(\tau)}^{\overline{\delta}(\tau)} \left(\tilde{\lambda}(\tau)\tilde{s}(\tau) + \tilde{\sigma}(\tau)\delta\right)^{\nu} \cdot \phi(\delta) \times$$
$$\times \left[\Phi\left(\overline{\varkappa}(\tau)\tilde{\sigma}(\tau)\cdot\delta + \overline{\beta}(\tau) + \overline{\varkappa}(\tau)\tilde{\lambda}(\tau)\right) - \Phi\left(\underline{\varkappa}(\tau)\tilde{\sigma}(\tau)\cdot\delta + \underline{\beta}(\tau)\underline{\varkappa}(\tau)\tilde{\lambda}(\tau)\right)\right]\mathbf{d}\delta =$$
$$\frac{\phi\left[\frac{\tilde{s}(\tau)}{\tilde{\Sigma}(\tau)}\right]}{\tilde{\Sigma}(\tau)} \times \int\limits_{\underline{\delta}(\tau)}^{\overline{\delta}(\tau)} \left(\lambda(\tau)\tilde{s}(\tau) + \tilde{\sigma}(\tau)\delta\right)^{\nu} \cdot \phi(\delta) \times$$
$$\times \left[\Phi\left(\overline{K}(\tau)\cdot\delta + \overline{B}(\tau)\right) - \Phi\left(\underline{K}(\tau)\cdot\delta + \underline{B}(\tau)\right)\right]\mathbf{d}\delta, \text{ where} \tag{102}$$

$$\underline{K}(\tau) = \underline{\varkappa}(\tau)\tilde{\sigma}(\tau), \quad \overline{K}(\tau) = \overline{\varkappa}(\tau)\tilde{\sigma}(\tau), \tag{103}$$

$$\underline{B}(\tau) = \underline{\beta}(\tau) + \underline{\varkappa}(\tau)\tilde{\lambda}(\tau)\tilde{s}(\tau), \quad \overline{B}(\tau) = \overline{\beta}(\tau) + \overline{\varkappa}(\tau)\tilde{\lambda}(\tau)\tilde{s}(\tau), \tag{104}$$

$$\underline{\delta}(\tau) = \frac{\underline{d}(\tau) - \tilde{\lambda}(\tau)\tilde{s}(\tau)}{\tilde{\sigma}(\tau)}, \quad \overline{\delta}(\tau) = \frac{\overline{d}(\tau) - \tilde{\lambda}(\tau)\tilde{s}(\tau)}{\tilde{\sigma}(\tau)}, \tag{105}$$

9. Rewrite the integral in terms of standardized integrals Ψ that I define in Internet Appendix.

$$\mathcal{I}^{(0)}(\tau) =$$
$$\frac{\phi\left[\frac{\tilde{s}(\tau)}{\tilde{\Sigma}(\tau)}\right]}{\tilde{\Sigma}(\tau)} \left\{\left[\Psi^{(0)}_{\overline{K}(\tau)}\left(\overline{\delta}, \overline{B}(\tau)\right) - \Psi^{(0)}_{\overline{K}(\tau)}\left(\underline{\delta}, \overline{B}(\tau)\right)\right] - \left[\Psi^{(0)}_{\underline{K}(\tau)}\left(\overline{\delta}, \underline{B}(\tau)\right) - \Psi^{(0)}_{\underline{K}(\tau)}\left(\underline{\delta}, \underline{B}(\tau)\right)\right]\right\} \tag{106}$$

$$\mathcal{I}^{(1)}(\tau) = \tilde{\lambda}(\tau)\tilde{s}(\tau)\mathcal{I}^{(0)}(\tau) + \tilde{\sigma}\frac{\phi\left[\frac{\tilde{s}(\tau)}{\tilde{\Sigma}(\tau)}\right]}{\tilde{\Sigma}(\tau)} \times$$
$$\times \left\{\left[\Psi^{(1)}_{\overline{K}(\tau)}\left(\overline{\delta}, \overline{B}(\tau)\right) - \Psi^{(1)}_{\overline{K}(\tau)}\left(\underline{\delta}, \overline{B}(\tau)\right)\right] - \left[\Psi^{(1)}_{\underline{K}(\tau)}\left(\overline{\delta}, \underline{B}(\tau)\right) - \Psi^{(1)}_{\underline{K}(\tau)}\left(\underline{\delta}, \underline{B}(\tau)\right)\right]\right\}. \tag{107}$$

10. Compute the conditional expectation as

$$\frac{\sum_{\tau}\mathcal{I}^{(1)}(\tau)}{\sum_{\tau}\mathcal{I}^{(0)}(\tau)} \tag{108}$$