Front-runners: predators or liquidity providers?

Erik Hapnes*

February 10, 2019

Abstract

I develop a dynamic model of strategic trading where front-running occurs in equilibrium. I show that the presence of a front-runner may uniformly improve liquidity in all states, while episodic presence of a front-runner shifts liquidity towards the states where he is present. Improved liquidity increases welfare, but state-contingent restrictions of front-running may be beneficial because it moves liquidity to those periods where it is most needed. The effect of restricting front-running without taking into account the effect on liquidity is different and may improve welfare for some traders. This reconciles the conventional ”predatory” view of front-running with the ”welfare- and liquidity-improving” view of this paper. The model provides novel empirical predictions on how temporary and permanent price impacts are affected differently by the expected length of a liquidity event. I provide empirical support for key predictions of the model using an event where the departure of a non-fundamental trader caused a long-lasting reduction in liquidity.

*EPFL and Swiss Finance Institute. Email: erik.hapnes@epfl.ch
1 Introduction

Do front-runners reduce market quality? Front-runners are order anticipators who try to profit from the trading intentions of other traders. They can profit from trading and later reversing their orders if they correctly anticipate the actions of other traders. [Harris (2003)] describes order anticipators as parasitic traders that only profit when they can prey on other market participants, in particular large traders, but do not provide any benefits for the market. The goal of this paper is to study the welfare implications of front running in the context of a dynamic equilibrium model of strategic trading.

I show that, contrary to the conventional wisdom, front running can be beneficial for both liquidity and welfare. Front-runners profit from anticipating the trading needs of other traders and in particular large traders. On the other hand, they also affect the equilibrium behavior in a way that increase liquidity. Traders care about both the price they can buy and sell a given quantity, but also how long it will take to execute an order. The price of trades is a zero-sum game whereas the execution time affects total welfare. Improved liquidity increases total welfare, but the gain can potentially be distributed unequally among traders with some traders loosing from the presence of front-runners.

When I refer to a front-runner in this paper, I consider a trader who trades away from his long-run position and later reverses his trades expecting to do so at a better price. In addition, whenever relevant, I will assume that front-runners do not have their own trading needs for tractability. One issue with the way I define front-runners is that it would also be a suitable definition for a financial intermediary/market maker and the conventional wisdom is more positive to the presence of them than front-runners. The model has this distinction in equilibrium and whenever some traders see a trader as a front-runner, other traders see him as a liquidity provider.
The effect of liquidity provision and front-running is relevant in multiple markets. An important recent debate is the effect of high frequency traders (HFTs). Some of the algorithms used by HFTs gives trading behavior that is in line with the definition of front-runners. While the evidence generally suggest that the large increase of algorithmic traders has improved liquidity in financial markets, there is a concern about their behavior in times of stress. The results of this paper suggests that there will be a large effect on liquidity if HFTs or other traders temporarily stop trading, yet welfare is still higher with their presence in some periods. Another example of front-running is a strategy used by Citigroup, internally called “Dr. Evil”, where they quickly sold a large amount of sovereign bonds and later reversing parts of their trades at a large profit(FT Aug. 22. 2006, The day Dr Evil wounded a financial giant). The strategy exploited a particular feature of the market structure and was largely seen as predatory. I provide empirical support for the model from European electricity derivatives. Liquidity deteriorated after a large trader, whose behavior is consistent with my definition of front-runners, went bankrupt and the effect lasted for several weeks after his bankruptcy.

This paper has two main contributions. The first contribution is methodological. It provides new insights into the implications of time-varying number of participants in financial markets. There are both testable predictions to evaluate the relevance of the model and policy implications for market design. Trade can potentially break down if sufficiently many traders are absent from the market at certain periods or if these periods are sufficiently short and liquidity moves to the periods with more traders. The second contribution is to evaluate how front-runners affect welfare in a tractable way. The first alternative is to evaluate the effect partially out-of-equilibrium such that the liquidity effect is not taken into account, but other values are equilibrium values. The second alternative is to evaluate the effect with the full equilibrium implications. The out-of-equilibrium approach
connects the model in this paper with the previous literature and conventional wisdom. The equilibrium approach gives a more positive view of front-running. The reason is that front-runners earn their profits due to imperfect competition and their presence increases competition. There are however situations where restricting front-runners can be welfare improving if it can move liquidity into states where the need for liquidity is greater.

The question has been examined before in Carlin, Lobo, and Viswanathan (2007). They consider an asset where trading has both a temporary and a permanent price impact. The temporary price impact gives rise to the possibility to gain from front running another trader who needs to buy or sell a given quantity. They show that it can be an equilibrium for traders to collude to not front-run each other in a repeated game. In this paper, I will revisit this question in a model where the price and behavior endogenously have similar dynamics as in the aforementioned paper. Modelling the price as an equilibrium value rather than specifying the dynamics offers two new insights related to the effect of front-runners. The first is that the utility of all traders is considered. In terms of the model, a front-runner will always provide liquidity to at least a subset of the traders and these are the traders with the largest trading needs. The second insight is a bit more subtle and relates to the temporary price impact. The temporary price impact decreases with more traders present and the effect is particularly strong with a time-varying number of traders. This effect is the main mechanism behind the different conclusions of this paper compared to the previous literature.

I find empirical support for some of the novel theoretical results in a recent natural experiment. A large and for a long period very successful trader recently went bankrupt. He was active in markets for European electricity derivatives. I find that liquidity deteriorated after his bankruptcy in a large number of derivatives with a stronger and more long lasting effect in some securities as predicted by the
theoretical model of this paper. He was a trader without any fundamental trading needs who made significant gains over a long period, which is consistent with the behavior of the front-runners in this paper.

2 Related literature

The paper most closely related to this paper is Carlin et al. (2007). As explained in the introduction, they analyze the effect on welfare when front-runners are present. They find that front-running reduces total welfare and introduces a scope for cooperation in a repeated game. There exist an equilibrium in repeated games where traders do not front-run each other except when a trader has to liquidate a very large position. The main difference between the model in this paper and the one developed in Carlin et al. (2007) is the price dynamics. They have exogenously specified price dynamics whereas the model in this paper has endogenous price dynamics. The equilibrium price dynamics are similar in the two papers, but the effect of the presence of more traders is different and this difference changes the welfare implications of front-runners. Brunnermeier and Pedersen (2005) propose a similar model where distressed traders have to liquidate a position. The remaining traders can take advantage of the distressed traders by trading in the same direction and later reversing their positions, causing momentum and consequent reversal. Predatory trading can endogenously put more traders into distress increasing the gains for the traders who are not distressed. An important difference from the model in this paper is that the price is not forward looking and this is the mechanism that allows the front-runners to gain from the momentum and reversal. Another difference is that the number of traders do not affect liquidity whereas this paper shows that a reduction in the number of traders can cause a large drop in liquidity. This effect increases the cost of offloading the positions gained and re-
duces the incentives for predatory trading. While the aforementioned papers take a predatory view on front-running, another area of the literature is the one starting with Grossman and Miller (1988). They propose a model where there are market makers who provide immediate liquidity and gain from trading at better prices in the future. More competition among market makers improves liquidity for traders with liquidity needs. This paper provides a link between the two strands of the literature. The front-runners provide liquidity as in Grossman and Miller (1988), but the equilibrium behavior looks similar to the trading behavior in Carlin et al. (2007).

The paper builds on the literature where traders with market power trade slowly. Among them are Vayanos (1999) and Du and Zhu (2017) where they examine the optimal trading frequency. Duffie and Zhu (2017) shows how mechanisms such as workup trades can improve welfare. The scope of these mechanisms is however limited due to their effect on liquidity between mechanisms as shown by Antill and Duffie (2017). They find that the market breaks down between mechanisms if they are sufficiently frequent. The model in this paper have a closely related result where trading breaks down if the market switches between states with different number of traders sufficiently often. The policy implications are however different. In their paper, mechanisms move liquidity, but have a neutral or negative effect on welfare. In this paper, the mechanism provides insights into how it can be possible to move liquidity into states where it is most needed. The aforementioned papers have homogeneous traders whereas Sannikov and Skrzypacz (2016) develop a tractable model with heterogeneous traders. One implication of the model with heterogeneous traders is that there will be front-running in equilibrium. In this paper, I extend Sannikov and Skrzypacz (2016) to include time-varying heterogeneity. Time-varying heterogeneity brings the model closer to the one of Carlin et al. (2007) and also enables me to evaluate the equilibrium effect of restricting
3 Model setup

The setup builds on Sannikov and Skrzypacz (2016). There are $N \geq 3$ traders who receive inventory shocks and trade with other traders to reduce their inventory costs. The payoff for trader $i$ is given by

$$E[\int_0^\infty e^{-rt}(-\frac{b_i(s_t)(X_i^t)^2}{2} + p_tq_i^t)dt]$$

where $b_i(s_t)$ is the state dependent inventory cost, $p_t$ is the price and $q_i^t$ is the trading intensity of trader $i$ at time $t$. The inventories follow

$$dX_t = \Sigma(s_t)dB_t - q_tdt$$

where $\Sigma(s_t)$ is a covariance matrix with rank $N$, $B_t$ an $N$-dimensional Brownian motion, $X_t$ a vector of unwanted inventories and $q_t$ a vector of trading intensities. I assume that there exist a random state $s_t \in \{s_1, s_2\}$ and the state will change with intensity $\lambda(s_t)$\footnote{That is, $\text{Probability}(s_{t+dt} = s_2 | s_t = s_1) = \lambda(s_1)dt$ and $\text{Probability}(s_{t+dt} = s_1 | s_t = s_2) = \lambda(s_2)dt$.}. The difference from the model in Sannikov and Skrzypacz (2016) is the introduction of the two states. I will consider three differences between the two states. The holding cost $b_i(s_t)$ can be state dependent. This extension changes the model from one where the same traders are being front-run to one where all traders can potentially be front-run part of the time. This change makes the setup more similar to Carlin et al. (2007) where both traders can receive inventory shocks and risk to be front-run. The second difference is to have a state dependent covariance matrix of inventory shocks. This enables me to answer when it is beneficial to restrict front-running by introducing time-varying liquidity needs.
The third dimension is a time-varying number of traders and allows me to evaluate the effect of restricting trading by a front-runner in some, but not all states.

The trading mechanism will be a uniform price conditional double auction. That is, each trader can observe and condition his demand on the trading rates of other traders. This simplification is necessary to avoid a complex filtering problem. It differs from the models such as the ones presented by Vayanos (1999) or Du and Zhu (2017) where traders are homogeneous. The complexity arrives from heterogeneous traders and each trader is interested in not only knowing the aggregate unwanted inventory, but also its distribution among the other traders. This assumption is is made to simplify the model and avoid difficult filtering problems where the traders try to figure out the distribution of inventories by observing total order flow. In addition, at least for sophisticated traders, there is evidence that they spend significant resources to identify the traders behind the order flow.

The book by Lewis (2014) explains some of these methods. It is also not necessary to observe the trades of every trader in the model. If there are M groups of homogeneous traders, it is sufficient to observe the aggregate trades of each of the groups. In the examples I will use in this paper, there are two or three groups of homogeneous traders. In the numerical examples with two groups, it is possible to infer the inventories of each group of traders by observing the price and volume.

The assumption of observable trades is not necessarily as strong as it initially looks. The reason is that it is possible that some traders are better off by letting other traders observe their order flow. Hapnes (2018) analyze a simpler, but economically similar model to explain this intuition. The intuition from the simplified model is as follows: suppose there is a subset of N traders where only the aggregate order flow is observable. The trader with the highest inventory cost will have an incentive to reveal his order flow and receive a smaller price impact. Of the remaining N – 1 traders, the trader with the highest inventory cost will now have an incentive to reveal his order flow. This continues until there is only one trader(or several homogeneous traders) left in the subset and hence his(their) order flow(s) will also be observable.
The conditional double auction is defined as:

**Definition 1.** Conditional double auction: *At every time* \( t \), *every active trader* \( i \) *announces a supply-demand function*

\[
p = \bar{\pi}^i + \sum_{i \neq j} \pi^{ij} q^j
\]

*where* \( p \) *is the price agent* \( i \) *is willing to pay to buy/sell the residual supply. The market maker selects* \( p \) *to satisfy*

\[
\sum_i q^i = 0 \quad \text{and} \quad p = \bar{\pi}^i + \sum_{i \neq j} \pi^{ij} q^j \quad \forall i.
\]  

(1)

I will restrict the attention to *acceptable linear stationary* equilibria. An equilibrium is *linear* and *stationary* if \( \pi^{ij} \) only depends on the state \( s \) and \( \bar{\pi}^i = \hat{\pi}^i X^i - \bar{p} \) where \( \hat{\pi} \) depends on the state \( s \) and \( \bar{p} \) depends on the state and inventory of the traders who do not trade. An equilibrium is *acceptable* if there is a unique solution to the system (1) for a given set of slopes \( \pi^{ij} \) and \( X = 0 \) when \( \hat{\pi}^i \neq 0 \forall i \).

A simpler way to solve the problem is to implement the equilibrium of the conditional double auction as a direct revelation mechanism defined as:

**Definition 2.** Direct revelation mechanism: *A stationary linear mechanism is a pair* \((P, Q)\) *where* \( P \) *is an* \( N \)-*dimensional vector and* \( Q \) *is an* \( N \times N \) *matrix where the columns sum to 0. \( P \) and \( Q \) *are constant between state changes. The market maker asks every active trader to report his inventory* \( X^i_t \) *at every time* \( t \). *The price is* \( p_t = \bar{p} + PX_t \) *and trading rates are* \( q_t = QX_t \). *The mechanism is truth-telling if telling the truth is an equilibrium.*

A mechanism is defined as *acceptable* if \( QX = 0 \) and \( PX = 0 \) only if \( X = 0 \). That is, there is a one-dimensional space of inventories \( X \) that results in no-trade and all those allocations result in different prices.
Proposition 1. There is a one-to-one map between acceptable stationary linear strategies and acceptable mechanisms such that for any non-zero allocation $X$ that leads to no trade, every element of $X$ is non-zero that lead to the same map $X \rightarrow (p, q)$. Conditional on an allocation $X^{-i}$, trader $i$ can attain the same price-flow pairs $(p, q)$ in the conditional double auction equilibrium and the direct revelation mechanism.

Proof: See appendix.

Proposition 1 shows that rather than solving a model with the conditional double auction, we can instead solve for an equilibrium in the direct revelation mechanism, a significantly simpler problem to analyze. The proposition is almost exactly the same as theorem 1 in Sannikov and Skrzypacz (2016). The only difference is that the price can be different from 0 for $X = 0$ if there are traders who do not trade at the given time who hold inventories. The differences only affect the definition of an acceptable and stationary equilibrium/mechanism and otherwise the proofs are exactly the same.

There is one important difference between observability of inventories between my model and the one presented in Sannikov and Skrzypacz (2016). I allow for equilibria where some of the traders only trade in one of the states. In these equilibria, we have to account for the uncertainty in the inventories of the traders who do not trade. In equilibrium, this is not going to affect trading behavior, but will have an impact on the realized price since it does not take into account the unobserved inventory shocks of the traders who do not trade.

4 Model solution

Define the value function of trader $i$ as

$$ f^i(X, s, t) = \max_{y} E \left[ \int_{0}^{\infty} e^{-rt} \left( -\frac{b_i(s_t)(X^t)^2}{2} + p_t q^t \right) dt \right]. $$
The variables of the value function are the inventories \( X \), the current state \( s \) and \( \hat{t} \) that is defined as the time since the last change in \( s \). The inventories \( X \) are the inventories of the traders who trade in state \( s \) and the expected value of inventories of traders who do not trade in state \( s \). The inventories of traders who do not trade in state \( s \) will be normal and have a variance that is linear in \( \hat{t} \) and hence this is a state variable. The mean of the unknown inventories consist of one observable part if inventory shocks are correlated with the ones of the active traders and a mean zero part. Trader \( i \) announces \( \hat{X}_t^i = X_t^i + y^i \) at time \( t \) and in the equilibrium, every active trader announces his position truthfully, i.e. \( y^i = 0 \). Define \( Q^i(s) \) as the \( i \)th column, \( Q^i(s) \) as the \( i \)th row and \( q^{ii} \) as the \( i \)th diagonal element of \( Q(s) \) and \( p^i(s) \) the \( i \)th element of \( P(s) \). When every other trader announces his inventory and trader \( i \) announces \( \hat{X}_t^i \), the inventories have the dynamics

\[
\frac{dX_t}{dt} = -Q(s)X_tdt - Q^i(s)y^i dt + \sum(s)dB_t.
\]

The HJB-equation for trader \( i \) is given by

\[
rf^i(X, s, \hat{t}) = f^i_t(X, s, \hat{t}) + \sup_y \left\{ -\frac{b^i(s)(X^i)^2}{2} + (P(s)X + p^i(s)y)(Q^i(s)X + q^{ii}(s)y) \right. \\
- f^i_s(X, s, \hat{t})(Q(s)X + Q^i(s)y) + \frac{1}{2} \text{tr}[\Sigma(s)\Sigma(s)^T f^i_{xx}(X, s, \hat{t})] \\
\left. + \lambda(s)(E[f^i(X, \hat{s}, 0)] - f^i(X, s, \hat{t})) \right\}
\]

and it has to hold for state \( s \) and \( \hat{s} \). Compared to Sannikov and Skrzypacz (2016), there are two differences. The possibility that some of the traders do not trade in one of the states makes it necessary to include a state variable \( \hat{t} \) to account for the unobserved inventory they have received while they do not trade. The second difference is the introduction of the jump between states where the expectation takes into account that the inventory of some traders can be unknown before the jump, but all inventories will be revealed at every jump. This is a quadratic optimization problem and we can conjecture that the solution will be on the form
\( f^i(X, s, \hat{t}) = X^T A^i(s)X + k_0^i(s) + k_1^i(s)\hat{t} \) where \( A^i(s) \) is a symmetric \( N \times N \) matrix and \( k_0^i(s) \) and \( k_1^i(s) \) are constants. With this ansatz, the HJB-equation becomes

\[
r(X^T A^i(s)X + k_0^i(s) + k_1^i(s)\hat{t}) = k_1^i(s) + \sup_y \frac{-b_i(s)(X^t)^2}{2} + (P(s)X + p^i(s)y)(Q^i(s)X + q^i(s)y)
- 2X^T A^i(s)(Q(s)X + Q^i(s)y) + \frac{1}{2} \text{tr}[\Sigma(s)\Sigma(s)^T A^i(s)]
+ \lambda(s)(X^T A^i(s)X) + E[(\hat{X} - X)^T A^i(\hat{s})(\hat{X} - X)] + k_0^i(\hat{s}) - X^T A^i(s)X + k_0^i(s) + k_1^i(s)\hat{t})
\]

where \( \hat{X} \) is the true inventories and \( X \) is the observed inventories. There is only a difference for the traders who do not trade. If we match coefficients, take the first order condition and impose truth-telling \( y = 0 \) we obtain the set of equations

\[
p^i(s)Q^i(s) + q^i(s)P(s) = 2(A^i(s)Q^i(s))^T,
\]

\[
(r + \lambda(s))A^i(s) + A^i(s)Q(s) + Q^T(s)A^i(s) = \frac{P^T(s)Q^i(s) + (Q^i(s))^T P(s)}{2} - \frac{b_i(s)}{2} 1_i^i + \lambda(s)A^i(\hat{s}),
\]

\[
k_1^i(s)\hat{t} = \frac{\lambda(s)}{r + \lambda(s)} E[(\hat{X} - X)^T A(\hat{s})(\hat{X} - X)]
\]

and

\[
(r + \lambda(s))k_0^i(s) = k_1^i(s) + \frac{1}{2} \text{tr}[\Sigma(s)\Sigma(s)^T A^i(s)] + \lambda(s)k_0^i(\hat{s}).
\]

We need to solve the sets of equations (2) and (3) for every trader and for both states to find \( Q(s), P(s) \) and \( A^i(s) \) for all traders and states. The second order condition is satisfied if \( p^i(s)q^i < 0 \). Stability of trading behavior requires that all eigenvalues of \( Q(s) \) have negative real parts. The equations (4) and (5) only affect the utility function, but not the trading behavior.

**Proposition 2.** If \( (Q(s), P(s), k_0^i(s), k_1^i(s) \text{ and } A^i(s), i = 1..N, s = 1, 2) \) solves the system (2)-(5) and satisfies the second order condition for all traders in both states
and where all eigenvalues of $Q(s)$ have negative real parts, then all traders prefer to tell the truth rather than any other action that satisfies the no-Ponzi condition $E[e^{-rt}X_t^2] \rightarrow 0$.

Proof: See appendix.

Equation (4) takes into account the variance and covariance of inventories of the traders who do not trade in the current state. The covariance matrix of shocks is constant in each state and hence the shocks are linear in $\hat{t}$ which is the time since the last time-change. All positions of traders who trade in at least one state will be known whenever the state changes since they either traded right before or right after the state change. If all traders trade all the time, it will be equal to 0. With the exception of a few special cases, I have not been able to find a closed form solution to the system and have to either use approximations or numerical solutions. The system has multiple solutions. Among them one equilibria where $M$ traders trade all the time where $3 \leq M \leq N$. In this paper I will focus on three kinds of equilibria. In all of them, a subset of traders will trade all the time, the remaining traders will trade in (i) both states, (ii) no state or (iii) only one of the states. This is to evaluate the effect of restricting the trade of a front-runner, while still keeping the tractable solution given by equations (2)-(5).

The model gives both a temporary and a permanent price impact. Define $I^i(s)$ as the temporary and $\Lambda^i(s)$ as the permanent price impact of trader $i$ in state $s$. Let $Z_0(s)$ be the eigenvector corresponding to the eigenvalue 0 of $Q(s)$ normalised such that it sums to 1 and define $\eta_i(s) = \sum_{j \neq i} Z_{0j}^i(s)$ where $Z_{0j}^i(s)$ is the $j$th element of $Z_0(s)$.

**Proposition 3.** The temporary and permanent price impact of trader $i$ are given by $I^i(s) = -\frac{p_i(s)}{\eta_i(s)}$ and $\Lambda^i(s) = \frac{P(s)Z_0(s)}{\eta_i(s)}$.

Proof: See appendix
The temporary price impact is related to the trading intensity of the trader because his current trading intensity is informative about the total amount he will buy or sell. The permanent price impact is related to the capacity to hold inventories. This difference offers testable predictions of the model as certain shocks will have a strong effect on the equilibrium trading intensity, but not on the capacity to hold inventory and vice versa. If some traders are restricted from trading for a short period of time, it will cause a strong effect on \( q_i(s) \) in the equilibrium, but the ability to hold inventory is not affected much because they will start trading again soon. A longer lasting shock will have the opposite effect.

There are only a few special cases where there are closed form solutions to the system of equations given by equations (2) and (3). In the situations that are relevant for the questions of this paper I need to use approximate or numerical solutions. The following result is useful in both situations:

**Result 1.** The unique equilibrium with homogeneous traders is characterised by:

\[
P = -\frac{b}{Nr} \mathbf{1} \quad \text{and} \quad Q = \frac{N-2}{N} (I - S/N)
\]

and matrices \( A^i \) with

\[
a_{ii} = -\frac{b}{2r} \frac{3N-2}{N^2}, \quad a_{ij} = -\frac{b}{2r} \frac{N-2}{N^2} \quad \text{and} \quad a_{jk} = \frac{b}{2r} \frac{N-2}{(N-1)N^2}
\]

where \( \mathbf{1} \) is a vector with ones and \( S \) is an \( N \times N \) matrix with ones.

Result 1 is from Proposition 12 of Sannikov and Skrzypacz (2016). It will be useful in this paper as it serves as a starting point for approximate solutions and numerical algorithms.

5 Equilibrium with varying participants

A trader might act as a front-runner in some periods and as a liquidity provider in other periods. Because of this, it could potentially be welfare improving to restrict
his trading in some periods. It turns out that this may have a strong effect on liquidity.

Throughout this section, I assume that all $N$ traders have homogeneous inventory costs, and the random state $s_t \in \{s_1, s_2\}$ defining the composition of traders participating in the market. Namely, only $m$ traders (henceforth, group 1) can trade all the time, whereas the other $N - m$ traders (group 2) can only trade in the state $s_1$. I assume that $s_t$ follows a two-state Markov chain with the transition intensities from $s_1$ to $s_2$ and from $s_2$ to $s_1$ both equal to $\lambda$. The Markov chain assumption simplifies the analysis because the transitions between states are independent of the current inventories. Hence, the same Ansatz used to solve the partial differential equation in the previous sections is still valid.\footnote{Ideally, one would like to study the problem where a trader is restricted from participation only at the times when he acts as a front-runner. However, this problem is significantly harder because the decision to restrict one or several traders depends on the current inventories.} I will restrict attention to equilibria that are symmetric within each group of traders. That is, all traders within each of the two groups follow the same (group-specific) strategy. I will denote by $q_i(s_j), i = 1, 2$ the trading intensity of group $i$ in the state $s_j, i, j = 1, 2$.

**Proposition 4.** Define $\bar{\lambda} = \frac{rN(N-1) m - 2}{N - m}$. There exist an equilibrium such that all traders trade in state $s_1$ and no traders trade in state $s_2$ if $\lambda = \bar{\lambda}$. There are no other values for $\lambda$ where there is such an equilibrium unless this structure is assumed. In this equilibrium, the trading intensities $q_i(s_j, \lambda)$ satisfy:

- $q_1(s_1, 0) = q_2(s_1, 0)$ and $\frac{q_1(s_1, \bar{\lambda})}{q_1(s_1, 0)} = \frac{q_2(s_1, \bar{\lambda})}{q_2(s_1, 0)} = \frac{r + 2\lambda}{r + \lambda}$
- $q_2(s_2, \lambda) = 0$, while $0 = q_1(s_2, \bar{\lambda}) < q_1(s_2, 0)$
- $q_2(s_2, \lambda) = q_1(s_2, \lambda) = 0$ for all $\lambda \geq \bar{\lambda}$. 
If \( N = 4 \) and \( m = 3 \), there is a unique equilibrium where the unrestricted traders always trade if \( \lambda \in [0, \bar{\lambda}) \). There are no such equilibria for \( \lambda \geq \bar{\lambda} \). \[4\]

Proof: see appendix.

I conjecture that the result of Proposition 4 also holds for general values of \( N \) and \( m \) as the same economic mechanisms will be present for other values of these parameters. By Proposition 3, equilibrium liquidity (as measured by the temporary price impact) is linked to trading intensities. The following is true for \( N = 4 \) and \( m = 3 \):

**Corollary 1.** Equilibrium price impacts satisfy: \( \frac{\partial I(s_1)}{\partial \lambda} < 0 \), \( \frac{\partial I(s_2)}{\partial \lambda} > 0 \), \( \frac{\partial I(s_1)}{\partial \lambda} = 0 \) and \( \frac{\partial I(s_2)}{\partial \lambda} < 0 \).

Proof: Numerical solution to the model with a unique equilibrium. See Figure 1.

Due to market power, the number of traders naturally has an effect on equilibrium liquidity (see Sannikov and Skrzypacz (2016) and the special case shown in Result 1). Time varying number of participants exacerbates this liquidity effect. The mechanism for increased liquidity (as captured by the drop in the temporary price impact or equilibrium trading speed) in state \( s_1 \) is as follows: There is a direct effect from the number of traders such that liquidity is higher in the state with more traders. However, in addition to this direct effect, there is also an indirect effect from future shocks to liquidity: Each trader knows that the state can switch

\[4\] If the solution is unique, it can be found numerically with the following algorithm:

1. Solve the system for \( \lambda_0 = 0 \).
2. Iteratively solve the system for \( \lambda_n = nh \) with Newton’s method. Use the solutions from \( \lambda_{n-1} \) as starting values.

This algorithm guarantees to find the solution of the problem for sufficiently small \( h \).

\[5\] My current proof requires significant amount of computing time for particular values of \( N \) and \( m \) and much more for general values. It holds for other values tested.
Figure 1: The top left plot shows the trading intensity in each state. The blue and yellow lines are in state 1 whereas the green line is state 2. Top right is the permanent price impact in blue and the permanent price impact conditional on staying in state 2 in yellow. Bottom left is temporary price impact in state 1 and bottom right is temporary price impact in state 2. X-axis is $\lambda \in [0, 6r)$. There are 3 unrestricted and one restricted trader.

to a state with lower liquidity in the future and will thus trade slightly faster when $\lambda > 0$, compared to the case when there is a fixed number of traders (i.e., when $\lambda = 0$). This further improves liquidity and every trader will again trade faster. The effect is the exact opposite in state 2. If the switching intensity is sufficiently large (i.e., when $\lambda > \bar{\lambda}$), it is optimal for unrestricted traders to wait until the transition to state $s_1$ happens, and then trade with a larger number of traders. As a result, liquidity in state $s_2$ may evaporate completely.

The threshold $\bar{\lambda}$ can be separated into components: The direct component, $\frac{rN(N-1)}{2}$, and the indirect component, $\frac{m-2}{N-m}$. The direct component reflects the fact
that liquidity is increasing in the total number of market participants. The indi-
rect component is proportional to the ratio of the number of unrestricted traders relative to the number or restricted traders. As a result, it is likely to see strong
time-varying values of liquidity if the market has few participants (small \( N \)), or if there are periods where a large fraction of traders are unable to trade. Interestingly enough, for the case when \( \lambda = \bar{\lambda} \), we can characterize the indirect impact of time-varying liquidity explicitly: In this case, agents trade \( \frac{r+2\lambda}{r+\lambda} \) times faster in state \( s_1 \) than they would otherwise do if liquidity were constant.

Empirical evidence suggests that a drop in the number of traders indeed may have a significant impact on market liquidity. For example, a study by Bogous-
slavsky, Collin-Dufresne, and Sağlam (2017) finds that trading costs increased after a market maker (Knight Capital) experienced major losses from a trading glitch. This affected liquidity in all the stocks where Knight Capital was a designated market maker and lasted until the stocks received new market makers.

Another empirical prediction of Corollary 1 is the difference between the effect of time-varying liquidity on temporary and permanent price impacts. A liquidity event that is expected to be short-lasting will not influence the permanent price impact much, but will have a stronger effect on the temporary price impact and vice versa. The intuition is that the permanent price impact depends on the capacity of the traders to hold inventories whereas the temporary price depends on the equilibrium trading speed of the traders. Investigating this empirical prediction is an interesting topic for future research.

\(^{6}\text{The minus 2 accounts for the necessity of at least three traders to have a solution with trade.}\)
6 Equilibrium and front-running

This section shows how front running is present in the unrestricted version where all the traders trade all the time. There are four traders and two states. Trader one and two have holding cost \( b_1 = b(1 + e) \) in state one and \( b_2 = b(1 - e) \) in state two where \( b \) and \( e \) are constants with \( e \in (0, 1) \). The holding costs for trader three and four are the opposite. In state \( s_i \), the state will change with intensity \( \lambda_i = \alpha L \) and \( \lambda_2 = (1 - \alpha)L \). The only difference between the two groups is the amount of time they expect to have high and low holding costs. At time 0 the inventories are given by \( X_0 = (1, 0, 0, -1) \). This makes it easy to consider if any of the traders are trading away from their long run position which is 0 for all traders. There is no simple closed form solution to the system of equations given by equations (2) and (3). It can however be approximated well for small values of \( e \) by considering a linear approximation around the solution for \( e = 0 \).

**Proposition 5.** For small values of \( e \), the matrix \( Q \) can be approximated by

\[
Q_1 = r(I - \frac{S}{4}) + r(I - \frac{S}{4})D_{1e} + \tilde{Q}_1 D_{1e}
\]

where \( I \) is the identity matrix, \( D_{1e} = \text{diag}(e, e, -e, -e) \), \( S \) is a \( 4 \times 4 \) matrix of ones and \( \tilde{Q}_1 = \frac{L\alpha}{2(L + 3r)(L + 6r)} \)

\[
\left( \begin{array}{cccc}
-3(L + 4r) & L & L + 6r & L + 6r \\
L & -3(L + 4r) & L + 6r & L + 6r \\
L + 6r & L + 6r & -3(L + 4r) & L \\
L + 6r & L + 6r & L & -3(L + 4r)
\end{array} \right)
\]

The price vector can be approximated by

\[
P_1 = -\frac{b}{4r} \frac{1}{4} (I + \frac{2r}{3} D_{1e}) - \frac{4\alpha L}{3} \frac{L + 5r}{(L + 2r)(L + 6r)} D_{1e}
\]

where \( 1 \) is a vector of ones. The results are symmetric for \( Q_2 \) and \( P_2 \) with \( 1 - \alpha \) instead of \( \alpha \) and \( D_{2e} = -D_{1e} \).

Proof: See appendix
The approximation in proposition 5 has both $Q$ and $P$ separated into three separate terms where each of them have a clear intuition. The first term is the solution to the model without heterogeneity in holding cost ($e=0$). The second term is the approximate solution to the model when the state is constant and the third term accounts for time-varying state. The effect from heterogeneous holding costs and time-varying holding costs have the opposite sign. The intuition is that a trader who expects his holding costs to increase will be less inclined to hold inventories today and vice versa. The heterogeneity between the two groups is captured by $\alpha$. A lower $\alpha$ means that we expect to spend more time in state 1 and hence the difference in inventory costs has a stronger effect in state 1.

Define $\bar{X}_t = E_0[X_t|s_r = 1 \forall r \leq t]$. This is the expected inventories conditional on being in state 1 from time 0 to time $t$ and they follow

$$d\bar{X}_t = -Q_1\bar{X}_t dt. \quad (6)$$

We can use equation (6) to evaluate when we will have front-running and how the different parameters affect the magnitude of front-running. From the definition of a front-runner, a trader is defined as a front-runner if he trades when he is at his long-run position. The only situation where no trader is defined as a front-runner is when the inventories are allocated efficiently between the groups. Otherwise there will be at least one trader who trades away from his long-run position at least in some periods.

Trader 2 and 3 will be defined as front-runners if they trade because the aggregate inventories is 0. The solution to $\bar{X}_{3t}$ is given by

$$\bar{X}_{3t} = \frac{1}{2}e^{-rt}(-1 + e^{\frac{r(L(1-2\alpha) + 6r)t}{L+6r}}).$$

$\bar{X}_{2t}$ has a slightly messier expression, but has the same implications. The first insight is that there is no front-running if $L(1-2\alpha) + 6r = 0$. If $L(1-2\alpha) + 6r > 0$, trader 2 and 3 will front-run trader 4. This is similar to Sannikov and
where the trader with low holding cost (trader 4) trades slowly and other traders front-run him. When the jump intensity $L$ is sufficiently large and asymmetric with $\alpha > \frac{1}{2}$, we can have the opposite result. Trader 2 and 3 will front-run trader 1 if $L(1 - 2\alpha) + 6r < 0$. The intuition is that the trader who has a low holding cost most of the time will be front-run. Figure 2 shows how the position of trader 3 evolves over time and how the peak value depends on the different parameters. It has the feature that the trader will first trade away from his long-run position and gradually revert back to 0. The amount of front-running depends on how frequently the state changes and more frequent state changes reduces the amount of front-running in equilibrium. Higher holding cost differences increases the amount of front-running and it will be stronger when the group that expects to have low holding costs more of the time has low holding costs.

The above analysis has a fixed number of traders. One way to understand the effect of front-runners is to extend it to a setting with $2n$ traders with $n \geq 2$ in each group as described above. To understand the effect of additional traders we can calculate the discounted cash-flows and inventory costs for different values of $n$. Figure 3 and shows the discounted total cash-flow and inventory cost for each type of trader. The plots for the front-runners is the aggregate cash-flow and inventory cost for the $(2n - 2)$ front runners. The main observation is that more front-runners reduce the cash-flow for trader 1 and for the front-runners with one exception. The total discounted cash-flow for the front-runners is slightly higher for $n = 3$ than for $n = 2$. On the other hand it increases the cash-flow of trader 2 since the aggregate cash-flow sums to 0. The total effect of increases in $n$ is that the traders with inventories at $t = 0$ are better off due to faster trading speed. There is only one type of trader who is hurt by more front-runners in this setting; other front-runners.
Figure 2: The plots show the expected inventory of trader 3. The top left has time at the x-axis. Top right e at the x-axis, bottom left L and bottom right l. The parameters that are not on the x-axis are $e = 0.1$, $L = 10$ and $l = \frac{L}{2}$. The value of $t$ is taken as the one that maximize $\tilde{X}_{3t}$.

7 Front-running and welfare

I will consider two different approaches to evaluate the effect of front-running. The first one is to consider the effect of restricting front-running by out-of-equilibrium behavior. This is done be reducing the magnitude of trading by one trader who is defined as a front-runner relative to his equilibrium behavior. The second approach is to take into account the equilibrium effect of restricting one of the traders in one of the states. The difference between the two approaches can be thought of in terms of as a stock price with both a temporary and permanent price impact defined as

$$p_t = p_0 + \sigma B_t + I_{q_t} + \Lambda \int_0^t q_s ds$$  (7)
where \( q_t \) is the trading rate. The temporary price impact is given by \( I q_t \) and the permanent by \( \Lambda \int_0^t q_s ds \). Price dynamics on this form is often used in practice. Almgren and Chriss (2001) among others analyze optimal trading behavior in such a setting. The first approach is similar to considering price dynamics similar to (7) with constant parameters whereas the second approach takes into account that the presence of more traders influence the parameters for temporary and permanent price impact. The first approach gives similar results as the previous literature on the topic and the main reason for the difference in my results is the effect of equilibrium trading speed that affects the temporary price impact.

The simplest way to consider the effect of front-running is to keep the setup from the previous section with two groups of traders and initial positions linear in \( X_0 = (1, 0, 0, -1) \) and let the parameters be such that trader 2 and 3 front-run trader 1 in state 1. For the equilibrium effect, we also have to account for the future inventory shocks and need to specify \( \Sigma(s) \). I will use

\[
\Sigma(s) = \sigma(s) \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & \varepsilon & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

where \( \sigma(s) \) is a constant that takes into account the possibility of different distributions of inventory shocks in the two states and \( \varepsilon \to 0 \). The choice if of \( \Sigma(s) \) has three elements. The possibility of state dependent \( \sigma(s) \) enables me to answer the question of when it can be beneficial to restrict a front-runner. Trader 2 is hit by small inventory shocks and trade "only" due to inventory shocks of other traders. It is however necessary to have \( \varepsilon \neq 0 \) to satisfy the full rank requirement of \( \Sigma(s) \). The last choice is to have independent inventory shocks and this choice is made solely for simplicity.
7.1 Out-of-equilibrium

Consider the effect of changing the trading behavior of trader 2 by $z_2 dt$ to reduce front-running. The change is distributed among the other traders according to $Q(s)$. The effect is that trader 1 will be closer to his long-run position, trader 2 will front-run less, trader 3 will front-run more and trader 4 will be further away from his long-run position. Define $Z = \begin{pmatrix} z_1 dt & z_2 dt & z_3 dt & z_4 dt \end{pmatrix}^t$. The utility at time 0 for trader $i$ is given by

$$U^i(X_0) = -\frac{b_i}{2} X_0 dt + p_i(q^i_t) dt + (1 - rdt)(1 - \lambda dt) E[(X_{dt})'A^i(s)(X_{dt})]$$

$$+ (1 - rdt)\lambda dt E[(X_{dt})'A(s')(X_{dt})|s]$$

and out of equilibrium by

$$\hat{U}^i(X_0) = -\frac{b_i}{2} X_0 dt + p_i(q^i_t + Z^i) dt + (1 - rdt)(1 - \lambda dt) E[(X_{dt} + Z)'A^i(s)(X_{dt} + Z)]$$

$$+ (1 - rdt)\lambda dt E[(X_{dt} + Z)'A(s')(X_{dt} + Z)|s].$$

If we cancel the terms that enter in both expressions and terms of the order $d^2_t$ and $dt dW_t$ we obtain

$$\Delta^i_U = \hat{U}^i(X_0, s) - U^i(X_0, s) = p_i Z^i + 2 * Z' * A^i(s)X_0.$$

The first term is related to the net gain from trade at time 0. This has to be 0 in aggregate due to market clearing. The second term is the effect on the change of the allocations at time $dt$. I use the approximate solution rather than the numerical solution to obtain closed form expressions. Numerical examples give the same results.

**Proposition 6.** With the approximate solution in proposition 3 we have

$$\Delta^1_U > 0, \Delta^2_U < 0, \Delta^3_U > 0, \Delta^4_U < 0$$

and

$$\sum_i \Delta^i_U < 0.$$
For sufficiently small $\epsilon$, we have

$$\Delta^1_U + \Delta^2_U > 0$$

Proof: see appendix.

The result of proposition 6 has the same welfare effects as Carlin et al. (2007). Welfare improves for trader 1 and deteriorates for trader 2 and the total welfare for trader 1 and 2 improves. In addition, it takes into account the welfare of all traders rather than a subset and the total welfare for all traders is reduced. The reason is that trader 4 is worse off and this is the trader who has the strongest need to reduce his inventory.

7.2 Equilibrium and constant inventory shocks

Next consider the equilibrium effect of restricting front-running. To evaluate the equilibrium effect of restricting front-running, I will compare two equilibria. One where all traders trade all the time and one where trader 2 only trades in state 1. It is also possible to compare it with an equilibrium where trader 2 never trades, but this is strictly worse than one where he always trades. Let initial positions be given by $X_0 = (y, 0, 0, -y)$. The utility for any trader in the equilibrium where every trader trades in both states will be on the form $U^i = c^i_1 y^2 + c^i_2$ and in the equilibrium where trader two only trades in state one will be on the form $\hat{U}^i = \hat{c}^i_1 y^2 + \hat{c}^i_2$ where $j$ is the state. Define the differences by

$$\delta^i_{1j} = \hat{c}^i_{1j} - c^i_{1j}$$
$$\delta^i_{2j} = \hat{c}^i_{2j} - c^i_{2j}.$$  

For the parameters where this condition fails to hold, the approximation will probably not be appropriate for answering this question.

25
The difference in utility from initial positions for trader \(i\) is given by \(\delta_{i1}y^2\) and for future inventory shocks by \(\delta_{2j}\). First consider the situation where inventory costs \(b_i\) are constant to isolate the effect of time-varying number of traders.

**Proposition 7.** For any \(\lambda \in (0, \bar{\lambda}]\) with homogeneous inventory cost, the difference in utilities satisfy:

- \(\delta_{11}^i > 0\) for \(i = 1, 4\).
- \(\delta_{12}^i < 0\) for \(i = 1, 4\).
- \(\delta_{1j}^i = 0\) for \(i = 2, 3\) and \(j = 1, 2\).
- \(\delta_{2j}^i < 0\) for all \(i\) and \(j\).

**Proof:** see appendix

In the simple model, we see that the effect of restricting trading by trader 2 in state 2 reduces the welfare from future inventory shocks for all traders. This is because the effect of reduced liquidity in state \(s_2\) dominates the increased liquidity in state \(s_1\) in terms of the welfare effect. The effect of initial inventories can however have the opposite effect. If initial inventories are sufficiently large, it can be beneficial to restrict trading in the future to increase liquidity today. If we restrict trading today on the other hand, the effect is negative for all traders. In this setting, there is no front-running given the initial conditions due to homogeneous holding cost. In a setting with heterogeneous holding costs, the results are similar as in proposition 7.

**Proposition 8.** For \(\lambda \) close to 0 or \(\bar{\lambda}\) and small \(\epsilon\), the results from Proposition 7 holds with the following exceptions:

- \(\delta_{11}^4 < 0\) if \(\epsilon\) is sufficiently large and \(\lambda\) sufficiently close to \(\bar{\lambda}\).
\[ \delta_{12}^3 > 0 \text{ if } \lambda \text{ is sufficiently close to } 0. \]
\[ \delta_{21}^3 = \delta_{21}^4 > 0 \text{ if } \epsilon \text{ is sufficiently large and } \lambda \text{ close to } 0. \]
\[ \delta_{22}^2 > 0 \text{ if } \lambda \text{ is close to } 0. \]

The total welfare effect is \( \sum_i \delta_{11}^i > 0 \) and \( \sum_i \delta_{kj}^i < 0 \) otherwise.

Proof: See appendix.

The proposition has two elements. The effect from initial inventories and the effect from future inventory shocks. The traders who initially hold inventory can sell it faster in state 1 and this improves the welfare from initial inventories in state 1. The opposite is the case in state 2. In addition, trader 3 will be able to front-run more when trader 2 is restricted and thus his welfare from initial inventories is improved. Trader 4 can potentially also have a negative effect from the initial inventories in both states depending on the size of \( \epsilon \). The effect on future inventory shocks is clear. Welfare from future inventory shocks is always negative when inventory shock intensity is constant. In addition to restricting trader 2 from trading (and hence front-running) in state 2, it also endogenously reduces the front-running by trader 2 in state 1. I conjecture that this last element is due to trader 2 being more hesitant to take on inventory if he knows he will be restricted in the future, but I have not found any simple way to investigate if there are other potential mechanisms.

Proposition 8 is based on approximating the solution close to the parameters for \( \lambda \) where it is easy to find a closed form solution. If I solve the model numerically, I obtain the same results as above for all \( \lambda \) where the second order conditions are satisfied. That is, all values have the same sign, the \( \delta \)s that change sign do so once and when the approximate \( \delta \)s are equal to 0, the numerical solutions have a magnitude much closer to 0 compared to the approximate \( \delta \)s that are different from 0.
7.3 Asymmetric liquidity needs

The effect of restricting one trader always has a negative welfare effect from future inventory shock with the setup above. It can however be beneficial to restrict some traders in some periods due to the effect it has on moving liquidity between states. If inventory shocks are state dependent, it can potentially be beneficial to move liquidity to periods with larger inventory shocks. To allow for time-varying liquidity, the simplest option is to have \( \sigma(s_1) > \sigma(s_2) \).

**Proposition 9.** There exist \( \sigma(s_1) > \sigma(s_2) \) such that \( \delta_{2j}^i > 0 \) for \( i = 1, 3, 4 \) if:

- \( \lambda \) is sufficiently close to 0.
- \( \lambda \) is sufficiently close to 6r and \( \epsilon \) sufficiently small for \( i = 3, 4 \).
- \( \lambda \in (0, \bar{\lambda}) \) and \( \epsilon = 0 \).

Proof: see appendix.

The result of proposition 9 is that if the inventory shocks are sufficiently asymmetric, it can improve welfare if one trader is restricted when inventory shocks are small. We can potentially improve the welfare of trader \( i = 1, 3, 4 \) by restricting trader 2 in state \( s_2 \). The conditions in Proposition 9 are the approximate solutions close to \( \lambda = 0, \lambda = 6r \) or the numerical solution where the solution is unique from Proposition 4. It also holds in the numerical examples for \( \epsilon > 0 \) and \( \lambda \in (0, 6r) \) where the second order condition holds. The unrestricted traders will have their welfare increased whereas the restricted trader will have his welfare reduced. The reason is that liquidity is improved when it is needed and reduced when there is less use for it. Empirically, this will be similar to Carlin et al. (2007), but for a different reason and with the opposite welfare implications. In Carlin et al. (2007), traders cooperate to not front-run each other when the trading needs are small, but the cooperation breaks down when a trader has to buy or sell a large quantity.
are large. In my model, it can be beneficial to restrict trading in states with small inventory shocks to improve liquidity when it is most needed. One example of the exact opposite is during Black Monday, the crash of 1987, when many traders experienced that their brokers did not pick up their phones leaving the traders unable to reach them. Regulators and market participants are concerned that HFTs will have similar behavior and only be present in the market during calm periods. This is the opposite of what is considered in Proposition 9. With similar calculations I consider the effect of a trader who is only trading in one state compared to if he is never present. I have not managed to find examples where some of the traders have their welfare reduced if some traders are only present in one state compared to never present. This suggest that, in terms of my model, it is better with HFTs who stop trading in some periods compared to no HFTs even if they leave the market when liquidity is needed the most.

8 Evidence from Nordic electricity derivatives

There is a relatively liquid market for derivatives in Nordic electricity markets. Among the participants, there are both electricity producers and other firms with hedging needs and traders without any fundamental reason for trading in electricity derivatives (e.g., investment banks). One particular former trader in this market is Einar Aas. After almost two decades of consistent and large trading gains a single large position resulted in a loss that whipped out his wealth and a large fraction of the safety fund of the clearing house (FT Sep. 13 2018, “Trader blows €100M hole in Nasdaq’s Nordic Market”). The losses came after a large and increasing spread between Nordic and German power markets cause by forecasts of heavy rainfall

---

*His trading gains were large enough to earn the highest income of all Norwegian tax payers for 2006, 2009, 2013, 2014 and 2016.*
reducing the price of Norwegian hydropower and an increase in the price of carbon allowances increasing the price of German electricity. Neither of the two factors are likely to have long lasting effects on the liquidity in the market for electricity derivatives. As a result, this setting has a potential to evaluate the effect of one trader disappearing. It differs from the model by being permanent for the trader who disappears, but it also increases profits for potential entrants.

To measure the impact of this event on market liquidity, I use two datasets. The most detailed dataset is a panel of derivatives (liquidity is concentrated in futures with delivery) from August 20, 2018. This gives me one month of data before the date of the event and a longer sample afterwards. The dataset has daily information on bid and ask prices, trading volume, and the number of trades for every derivative. In addition, I have a longer sample of daily aggregate volume. I use the detailed dataset to estimate the effect on liquidity and the longer dataset to estimate the effect on volume. I have each dataset for both Nordic electricity markets and German electricity markets where the Nordic market has substantially more liquidity.

To test the impact of the event on liquidity, I use four samples. The samples are constructed as follows: The first sample is all derivatives with delivery and at least 5 observations before the event and at least 5 observations after the event. The second sample is constructed by removing all derivatives except futures from the first sample. The third sample is constructed by furthermore removing all derivatives with average bid-ask spreads above €0.4. The last sample is restricted to the most liquid futures contract in terms of the average number of trades and the average bid-ask spread. As a measure of liquidity, I use the normalized quoted

9Some derivatives exist both with and without delivery. They are highly correlated and would affect the inference if both were included. Liquidity is concentrated in the derivatives with delivery.
bid-ask spread. That is, I do a linear transformation such that the bid-ask spread has mean 0 and variance 1 for each derivative. This normalization is used because the different derivatives have different magnitudes of their bid-ask spreads, and this is a simple way to make them comparable. The results are similar if I use bid-ask spreads or relative bid-ask spreads for the samples that are not too heterogeneous. To capture time-varying liquidity, I split the sample into six sub-periods: One pre-event period, the event period, and 4 two-week periods after the event. The event window includes one week before the event, the week of the event and the following week. I include the week before the event because of media reports that the trader tried to offload his position over several days before the event. Another trader obtained the portfolio and I include the week following the event to account for trading by this trader to offload parts of his unwanted inventory. The two-week length of the following intervals is a compromise between efficiency and the ability to see the speed of reversal to the pre-event levels of liquidity. I obtain similar results for one-week intervals.

Table 1 shows that there is a strong effect during the event for every sample and this effect is long lasting. There is still a significant effect in the period that starts 6 weeks after the bankruptcy of the large trader for all samples except the one that only contains one (most liquid) futures contract, partially due to larger standard errors.

The key prediction of my theory (see Proposition 4) that I am testing here is that liquidity may (almost) evaporate when a large trader vanishes, but only when people anticipate that liquidity will be (partially) restored during the life span of the derivative contract. The derivatives in my sample have different horizons. Thus, my theory implies that derivatives with longer horizons should experience a stronger reduction in liquidity due to the possibility of increased future liquidity
over their life span. To test this prediction, I split the sample into four parts: weekly, monthly, quarterly, and yearly futures. The empirical prediction that I test is that the effect on liquidity is stronger for the longer horizon futures. The samples consist of assets that are more homogeneous than in the previous analysis and thus I use a slightly different measure. I normalize the bid-ask spread within each sample by the sample mean to keep the results comparable between samples.

As one can see from Table 2, liquidity is lower during the event period for all four sub-samples, but in subsequent periods, significant reductions in liquidity are only observed for the futures with a longer maturity. For the yearly futures there is still a significant effect even at the last period (i.e., almost three months after the event). The effect is also large in terms of economic significance. The significant differences between the effects of the event on liquidity across different futures maturities is thus perfectly in line with the theoretical prediction.

The data for trading volume is available for a longer period and I am thus able to compare the volume with previous years. I split the sample into sub-periods and include an interaction term for 2018. I find no significant effects on trading volume during the event period, which makes sense because there are at least two traders who had to offload large positions in that period. At the same time, trading volumes drop significantly after the event, consistent with the prediction that trading intensities are much lower in the state with a disappeared trader. This follows because the estimated coefficient on the interaction term is large, both in economic terms and in terms of statistical significance for all sub-samples except for the off balance sheet volume in the German market when compared to previous years.

---

10 Normalizing to mean 0 and variance 1 would not be suitable for the weekly futures because they are only traded close to maturity and it would drive the estimates towards 0.

11 Every year is matched by weeks rather than exact date and the starting week is taken as the
Thus, all empirical results are in line with the predictions of my model. Both liquidity and volume are reduced after a large trader disappears from the market. The effect is stronger and more long-lasting for securities for which traders have the option to wait for a potential future increase in liquidity, consistent with one of the main theoretical predictions of this paper.

9 Conclusion

I develop a model of strategic trading with an endogenously time-varying liquidity and show that the presence of a front-runner may improve liquidity and total welfare. Restricting front-running only improves welfare in two situations: (1) if initial inventories are sufficiently large, then it can be beneficial to commit to restricting front-running in the future, but not today and (2) if traders anticipate time-varying shocks to their trading needs. I study welfare implications for state-contingent regulation of front-running. Optimal regulation may feature restricting front running when trading needs are small, while permitting front-running when trading needs are larger. Restricting front-running part of the time increases liquidity in the state without restrictions and reduces it in the state with restrictions. The model offers new empirical predictions: A time-varying number of traders has different effects on the temporary and permanent price impacts depending on how long the liquidity is expected to be affected. The mechanism is that traders would trade slower and wait until future periods with more liquidity. I illustrate key mechanisms of my model with data from a natural experiment in the market for Nordic electricity derivatives.

Monday closest to August 15.
Bibliography


10 Appendix 1

10.1 Tables
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All derivatives</td>
<td>All futures</td>
<td>Liquid futures</td>
<td>Most liquid future</td>
</tr>
<tr>
<td>Sep 3-Sep 23 2018</td>
<td>0.45***</td>
<td>0.62***</td>
<td>0.86***</td>
<td>0.66**</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.119)</td>
<td>(0.163)</td>
<td>(0.304)</td>
</tr>
<tr>
<td>Sep 24-Oct 7 2018</td>
<td>0.23***</td>
<td>0.40***</td>
<td>0.63**</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.099)</td>
<td>(0.243)</td>
<td>(0.689)</td>
</tr>
<tr>
<td>Oct 8-Oct 21 2018</td>
<td>0.16***</td>
<td>0.13*</td>
<td>0.33*</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.075)</td>
<td>(0.173)</td>
<td>(0.133)</td>
</tr>
<tr>
<td>Oct 22-Nov 4 2018</td>
<td>0.11**</td>
<td>0.14**</td>
<td>0.42**</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.063)</td>
<td>(0.168)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>Nov 5 -</td>
<td>0.05</td>
<td>0.08</td>
<td>0.52***</td>
<td>-0.17*</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.086)</td>
<td>(0.126)</td>
<td>(0.101)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.18***</td>
<td>-0.24***</td>
<td>-0.46***</td>
<td>-0.18**</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.031)</td>
<td>(0.052)</td>
<td>(0.087)</td>
</tr>
<tr>
<td>N</td>
<td>5572</td>
<td>1548</td>
<td>330</td>
<td>75</td>
</tr>
<tr>
<td>R²</td>
<td>0.029</td>
<td>0.056</td>
<td>0.092</td>
<td>0.131</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses clustered by time.

* p < 0.10, ** p < 0.05, *** p < 0.01
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Weekly</td>
<td>Monthly</td>
<td>Quarterly</td>
<td>Yearly</td>
</tr>
<tr>
<td>Sep 3-Sep 23 2018</td>
<td>0.83**</td>
<td>0.43*</td>
<td>0.36***</td>
<td>0.75***</td>
</tr>
<tr>
<td></td>
<td>(0.332)</td>
<td>(0.241)</td>
<td>(0.115)</td>
<td>(0.188)</td>
</tr>
<tr>
<td>Sep 24-Oct 7 2018</td>
<td>0.19</td>
<td>0.25</td>
<td>0.54***</td>
<td>0.27**</td>
</tr>
<tr>
<td></td>
<td>(0.215)</td>
<td>(0.159)</td>
<td>(0.149)</td>
<td>(0.112)</td>
</tr>
<tr>
<td>Oct 8-Oct 21 2018</td>
<td>0.08</td>
<td>0.08</td>
<td>0.31***</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(0.204)</td>
<td>(0.088)</td>
<td>(0.111)</td>
<td>(0.124)</td>
</tr>
<tr>
<td>Oct 21-Nov 4 2018</td>
<td>0.00</td>
<td>0.09</td>
<td>0.02</td>
<td>0.30**</td>
</tr>
<tr>
<td></td>
<td>(0.230)</td>
<td>(0.110)</td>
<td>(0.072)</td>
<td>(0.147)</td>
</tr>
<tr>
<td>Nov 5 -</td>
<td>0.33</td>
<td>0.08</td>
<td>0.04</td>
<td>0.22**</td>
</tr>
<tr>
<td></td>
<td>(0.293)</td>
<td>(0.103)</td>
<td>(0.125)</td>
<td>(0.087)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.72***</td>
<td>0.84***</td>
<td>0.80***</td>
<td>0.70***</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(0.066)</td>
<td>(0.048)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>N</td>
<td>159</td>
<td>441</td>
<td>520</td>
<td>334</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.087</td>
<td>0.026</td>
<td>0.043</td>
<td>0.064</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses clustered by time.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 3: Volume effect

<table>
<thead>
<tr>
<th>Region</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance sheet</td>
<td>Total</td>
<td>On</td>
<td>Off</td>
<td>Total</td>
<td>On</td>
<td>Off</td>
</tr>
<tr>
<td>Sep 3-Sep 23</td>
<td>745.39*</td>
<td>127.51</td>
<td>617.89**</td>
<td>278.65*</td>
<td>38.02</td>
<td>240.63***</td>
</tr>
<tr>
<td></td>
<td>(439.66)</td>
<td>(263.39)</td>
<td>(254.80)</td>
<td>(152.58)</td>
<td>(120.03)</td>
<td>(88.98)</td>
</tr>
<tr>
<td>Sep 24-Oct 21</td>
<td>1093.74***</td>
<td>598.26**</td>
<td>495.47**</td>
<td>293.54**</td>
<td>208.06*</td>
<td>85.48*</td>
</tr>
<tr>
<td></td>
<td>(410.96)</td>
<td>(269.19)</td>
<td>(212.71)</td>
<td>(130.78)</td>
<td>(117.89)</td>
<td>(48.80)</td>
</tr>
<tr>
<td>Oct 21-</td>
<td>709.12</td>
<td>201.70</td>
<td>507.41**</td>
<td>289.82**</td>
<td>152.14</td>
<td>137.67***</td>
</tr>
<tr>
<td></td>
<td>(430.91)</td>
<td>(262.75)</td>
<td>(254.44)</td>
<td>(142.14)</td>
<td>(125.06)</td>
<td>(50.71)</td>
</tr>
<tr>
<td>2018</td>
<td>-806.28</td>
<td>-859.27***</td>
<td>52.99</td>
<td>12.26</td>
<td>41.17</td>
<td>-28.91</td>
</tr>
<tr>
<td></td>
<td>(555.73)</td>
<td>(309.33)</td>
<td>(334.67)</td>
<td>(187.46)</td>
<td>(172.43)</td>
<td>(55.17)</td>
</tr>
<tr>
<td>Sep 3-Sep 23 2018</td>
<td>169.61</td>
<td>380.27</td>
<td>-210.67</td>
<td>138.25</td>
<td>-183.35</td>
<td>321.60</td>
</tr>
<tr>
<td></td>
<td>(750.64)</td>
<td>(404.16)</td>
<td>(452.59)</td>
<td>(373.05)</td>
<td>(203.04)</td>
<td>(276.39)</td>
</tr>
<tr>
<td>Sep 24-Oct 21 2018</td>
<td>-2575.91***</td>
<td>-1397.81***</td>
<td>-1178.10***</td>
<td>-688.03***</td>
<td>-650.14***</td>
<td>-37.89</td>
</tr>
<tr>
<td></td>
<td>(647.65)</td>
<td>(360.94)</td>
<td>(391.84)</td>
<td>(203.66)</td>
<td>(184.20)</td>
<td>(72.25)</td>
</tr>
<tr>
<td></td>
<td>(647.43)</td>
<td>(355.92)</td>
<td>(406.36)</td>
<td>(214.25)</td>
<td>(189.00)</td>
<td>(76.50)</td>
</tr>
<tr>
<td>Constant</td>
<td>5151.77***</td>
<td>2774.02***</td>
<td>2377.75***</td>
<td>657.19***</td>
<td>562.59***</td>
<td>94.60***</td>
</tr>
<tr>
<td></td>
<td>(322.00)</td>
<td>(213.01)</td>
<td>(158.80)</td>
<td>(109.31)</td>
<td>(101.73)</td>
<td>(26.86)</td>
</tr>
</tbody>
</table>

| N | 278 | 278 | 278 | 278 | 278 | 278 |
| R² | 0.229 | 0.276 | 0.098 | 0.119 | 0.138 | 0.095 |

Robust standard errors in parentheses.

* p < 0.10, ** p < 0.05, *** p < 0.01
11 Appendix 2 - additional material

11.1 Numerical procedure

To be added. Newton's method with method for choosing starting values.

11.2 Exact solutions for time-varying participants

To be added. In line with numerical solutions.
12 Appendix 3 - proofs

12.1 Proof of Proposition 1

Too be added. Follows almost exactly from Sannikov and Skrzypacz (2016) Theorem 1.

12.2 Proof of Proposition 2

Define $G_t$ such that

$$G_t = \int_0^t -b_u(s_u)(X_u^2) + (P(s_u)X_u + p^j(s_u)y_u)(Q^j(s_u)X_u + q^{ii}(s_u)y_u) du + e^{-rt}f(X_t, s_t, \hat{t}_t).$$

$G_t$ has to be a supermartingale for any arbitrary strategy $y_t$ and a martingale for an optimal strategy. Differentiation with respect to $t$ gives

$$dG_t = e^{-rt}(\cdots) + e^{-rt} \sum \frac{\partial f(X_t, s_t, \hat{t}_t)}{\partial X_t} dB_t + (E[f(X_t, s_t', \hat{t}_t')] - f(X_t, s_t, \hat{t}_t))(dN_t - \lambda(s_t)).$$

where $N_t$ is a Poisson process with intensity $\lambda(s_t)$ and $(\cdots)$ is the HJB equation from (3). $G_t$ is a local martingale when the HJB equation holds as the drift is 0. The value function is quadratic and hence it is easy to see that

$$E[\int_0^t e^{-rs} \sum \partial f(X_s, s_s, \hat{t}_s)^2 ds] = e^{-rt} \int_0^t e^{-rs} s ds = c \frac{1 - e^{-rt}(1 + rt)}{r^2} \infty$$

as long as $E[e^{-rt} X^2] \leq \infty$. The transversality condition $E[e^{-rt} f(X_t, s_t, \hat{t}_t)] \to 0$ for any non-explosive solutions since the value function is quadratic.

12.3 Proof of Proposition 3

Proof. Temporary price impact to be added. Follows directly from Sannikov and Skrzypacz (2016).

Suppose a trader announces an additional inventory of $x$. This will eventually be allocated according as $x Z_0(s)$ and total price impact will be $x P(s) Z_0(s)$. The trader will sell $\eta_i(s)x$ and hence the price impact will be $x \frac{P(s)Z_0(s)}{\eta_i(s)}$. \qed
12.4 Proof of Proposition 4

Lemma 1. An equilibrium with homogeneous traders where every trader trades in state $s_1$ and no trader trades in state $s_2$ is unique. It is as the equilibrium where every trader trades all the time with holding cost $b\frac{r+2\lambda}{r+\lambda}$ and discount rate $r\frac{r+2\lambda}{r+\lambda}$.

Proof. First, if there is an equilibrium where all traders trade in state 1 and no traders trade in state 2, equation (3) in state 2 simplifies to

$$
(r + \lambda) A(2) = -\frac{b}{2}1^{ii} + \lambda A(1)
$$

$$
A(2) = -\frac{b}{2(r + \lambda)}1^{ii} + \frac{\lambda}{r + \lambda} A(1)
$$

If we plug this into equation (3) for state 1 and simplify we obtain

$$
r\frac{r+2\lambda}{r+\lambda} A^i(1) + A^i(1)Q(1) + Q^T(1)A^i(1) = \frac{PT(1)Q^i(1) + (Q^i(1))^T P(1)}{2} - \frac{r + 2\lambda b}{r + \lambda} 1^{ii}. \tag{8}
$$

Equation (8) is the equation we would have if every trader would trade all the time, have discount rate $r\frac{r+2\lambda}{r+\lambda}$ and holding cost $b\frac{r+2\lambda}{r+\lambda}$. This system have a unique solution when every trader trades all the time and hence there is also holds when there are two states, but only trading in one of the states. Equation (2) does not contain the discount rate or the holding cost.

The equilibrium of Lemma 1 holds regardless of $\lambda$ if we impose this structure on the trading matrix. If we change $\lambda$ slightly, we will still find this equilibrium. There is one $\lambda$ where we can find an equilibrium in the neighbourhood of the equilibrium of Lemma 1 with non-zero trade by the unrestricted trader in both states after a small change in $\lambda$. To find this $\lambda$, we need to solve equation (2) for the unconstrained traders in state $s_2$. This finishes the proof of proposition 4.
Proof. Equation (2) for the unrestricted traders in state $s_2$ simplifies to

$$2p_uq_u = 2\left(-\frac{b}{2(r + \lambda)} + \frac{\lambda}{r + \lambda}(a_1 - a_{12})\right)q_u$$

$$(1 - \frac{1}{m-1})p_uq_u = 2\frac{\lambda}{r + \lambda}(a_{12} - a_2)q_u$$

$$p_rq_u = 2\frac{\lambda}{r + \lambda}(a_{12} - a_2)q_u$$

where $p_u$ and $p_r$ are the elements of $P(s_2)$ related to the unrestricted traders and $q_u$ is the trading speed for the unrestricted trader in state $s_2$. $a_{ii}$, $a_{ij}$ and $a_{jj}$ are from the solution of the model with no heterogeneity. The conditions hold either if $q_u = 0$ by assumption or if $\lambda = \frac{\gamma N(N-1) m-2}{2N-m} = \bar{\lambda}$.

The rest of the proof is based on $N = 4$ and $m = 3$. To be generalized.

For this proof, I will introduce the following notation: $F(Z; \lambda)$ is the system of polynomial equations that satisfies equations (2) and (3) and $Z$ is the vector of the parameters of the matrices $A$, the price vectors $P$ and the trading matrices $Q$ such that any equilibrium satisfy $F(Z; \lambda) = 0$.

**Lemma 2.** The Jacobian matrix, $J(F; \lambda)$ of $F(Z; \lambda)$ has full rank at an equilibrium.

**Proof.** The proof has three steps. Step one is to assume that there are two equilibria in the neighbourhood of each other, $Z$ and $\hat{Z}$ and find conditions that $\tilde{Z} = Z - \hat{Z}$ needs to satisfy.

Step 1: If both $Z$ and $\hat{Z}$ satisfies (2) and (3). The first order Taylor expansion for $F(Z, \lambda)$ and $F(\hat{Z}, \lambda)$ are

$$F(Z, \lambda) = F(\hat{Z}, \lambda) + J(\hat{Z}, \lambda)(Z - \hat{Z})$$

and

$$F(\hat{Z}, \lambda) = F(Z, \lambda) + J(Z, \lambda)(\hat{Z} - Z).$$
The second order terms are on the order \((Z - \hat{Z})^T(Z - \hat{Z})\) and goes to 0 if \(Z\) is in the neighbourhood of \(\hat{Z}\) and there are no higher order terms. In an equilibrium, this simplifies to

\[
0 = J(\hat{Z}, \lambda)(Z - \hat{Z})
\]  

(9)

and

\[
0 = J(Z, \lambda)(\hat{Z} - Z).
\]  

(10)

The system from (2) and (3) has some properties that makes it possible to simplify this further. First, the system consists of polynomial equations where every element is either independent of \(Z\), \(Z_i\) or \(Z_iZ_j\) where \(Z_i\) and \(Z_j\) are elements of \(Z\) and \(i \neq j\). Every element that is independent of \(Z\) disappears in \(J(Z, \lambda)\) whereas elements that consists of \(Z_i\) enters as a constant in \(J(Z, \lambda)\). We can separate \(J(Z, \lambda)\) such that \(J(Z, \lambda) = J_1(Z, \lambda) + J_2(\lambda)\) where \(J_1(Z, \lambda)\) consists of the elements that enters as \(Z_iZ_j\) and \(J_2(\lambda)\) as the ones that enters as \(Z_i\) and thus as constants in \(J(Z, \lambda)\). We can combine (9) and (10) and get

\[
0 = (J_1(\hat{Z}, \lambda) + J_2(\lambda))(Z - \hat{Z}) - (J_1(Z, \lambda) + J_2(\lambda))(Z - \hat{Z})
\]

\[
0 = (J_1(\hat{Z}, \lambda) - J_1(Z, \lambda))((Z - \hat{Z})
\]

\[
0 = J_1(\hat{Z} - Z, \lambda)(Z - \hat{Z})
\]

\[
0 = J_1(\hat{Z}, \lambda)\hat{Z}.
\]

(11)

The step from the second to the third line use the observation that all elements of \(J_1(Z, \lambda)\) consists of constants time a linear term of \(Z_i\). The solutions to (11) are the same as the solutions to a slightly altered system (2) and (3) where we only keep the terms that are second order.

This system has multiple solutions (38 in the case of \(N = 4\) and \(m = 3\)). By direct computation, I am able that in any equilibrium, all of them turns out to be 0. Mathematica file can be provided. More detailed version to come.
Lemma 3. There are no $\lambda \geq 0$ except $\bar{\lambda}$ where one group of traders do not trade in one of the states except if this is imposed by assumption.

Proof. The next step is to show that there exist an equilibrium with trade by the unrestricted traders in both states if $\lambda \in [0, \bar{\lambda})$. Define $q_{u1}$ and $q_{r1}$ as the trading intensity by the unrestricted and restricted traders in state 1. We have $q_{u1} > 0$, $q_{r1} > 0$ and $q_{u2} > 0$ if $\lambda = 0$ and $q_{u1} > 0$, $q_{r1} > 0$ and $q_{u2} = 0$ if $\lambda = \bar{\lambda}$. Due to the continuity of the solution, if there is an equilibrium for one $\lambda$, there is an equilibrium in the neighbourhood of that equilibrium for values of $\lambda$ sufficiently close by unless $q_{u1} = 0$, $q_{r1} = 0$ or $q_{u2} = 0$. Otherwise the second order condition will fail unless any values for $P < 0$ which is not an equilibrium feature.

First, we know that the only possibility for $q_{u2} = 0$ is $\lambda = \bar{\lambda}$. Next, if $q_{r1} = 0$, this is the same as if the restricted traders will never trade. We can easily solve equation (3) for the restricted traders and obtain $A = -\frac{b}{2(r + \lambda)}$. Plug this into (2) and obtain two equations,

$$2p_{r1}q_{r1} = -2\frac{b}{2(r + \lambda)}q_{r1}$$

and

$$2p_{r1}q_{r1} = 0.$$ 

As a result, we know that there is no $\lambda$ with an equilibrium such that a small change in $\lambda$ changes the equilibrium to one where $q_{r1}$ is non-zero. Next, consider an equilibrium where the unrestricted trader only trades in state 2. This will simplify to the same two conditions as for the the setting when he trades in both states with one exception, the values in $A$ will depend on the number of unrestricted rather than total number of traders and the $\lambda$ required for such an equilibrium is negative. Thus, we know that there is a path of equilibria for $\lambda \in [0, \bar{\lambda})$ such that the second order conditions are satisfied. \qed
**Lemma 4.** Suppose there exist an equilibrium $Z$ that solves (2)-(3) for some $\lambda$, then there is an equilibrium $\hat{Z}$ in the neighbourhood of $Z$ that solves (2)-(3) for a $\hat{\lambda}$ in the neighbourhood of $\lambda$.

**Proof.** If $\hat{\lambda} = \lambda + \varepsilon$ for a small $\varepsilon$ such that $\varepsilon^2 \to 0$ we can rewrite $F(Z; \lambda)$ as

$$F(\hat{Z}; \hat{\lambda}) = F(Z; \lambda) + \tilde{F}(Z; \lambda)\varepsilon + J(F; \lambda)\tilde{Z}\varepsilon$$

where $\tilde{F}(Z; \lambda)$ is a vector of the values where $\lambda$ enters and $\tilde{Z} = \frac{\hat{Z} - Z}{\varepsilon}$. If both $(\hat{Z}, \hat{\lambda})$ and $(Z, \lambda)$ are equilibria it simplifies to

$$0 = 0 + \varepsilon(\tilde{F}(Z; \lambda) + J(F; \lambda)\tilde{Z})$$

which has a unique solution since $J(F; \lambda)$ is invertible. As a result, there exist a $\hat{Z}$ in the neighbourhood of $Z$ that solves $F(\hat{Z}; \hat{\lambda}) = 0$. \hfill $\square$

**Lemma 5.** There can not be a solution to the maximization problem with a solution $Z$ such that $Z_i \to \infty$ where $Z_i$ is the $i$th element of $Z$.

**Proof.** There are three types of elements in $Z$. The ones from $A$, $P$ and $Q$. I will treat each of them separately.

The elements from $A$ determine the total utility. Total utility is the total inventory costs and the price of total order flow. The first is bounded above by 0 and the second is zero sum. Hence, total utility is bounded above by 0. Every trader can deviate and announce inventories such that they do not trade. This puts a lower bound on their utility. As a result, all the elements from $A$ has to be finite in an equilibrium to ensure that utility is bounded.

Next consider the elements from $P$. The $i$th row of (2) satisfies

$$2p_i(s)q^{ii}(s) = 2A^i(s)Q^i(s) = 2A^i(s)\hat{Q}^i(s)q^{ii}(s)$$

which simplifies to

$$2p_i(s) = 2A^i(s)Q^i(s) = 2A^i(s)\hat{Q}^i(s).$$
The values of $A'(s)$ are finite. If the values of $Q'(s)$ are not finite, this means that trader $i$ will either buy infinite amounts from one group and sell the same amount to the other group. This will result in at least one positive eigenvalue of $Q(s)$.

The last part is to consider what happens if $q^{ii}(s)$ is not finite. I will do this in two steps. First, suppose one of the $q^{ii}(s) → ∞$ then $p_i(s) → 0$. To show this, first consider the $i$th element of equation (2). This has $p_i(s)q^{ii}(s) = (A^i(s)Q^i(s))(i)$ where $(A^i(s)Q^i(s))(i)$ is the $i$th element of $(A^i(s)Q^i(s))$. Next consider element $ii$ in (3). This element takes the form $\bar{A} + 2(A^i(s)Q^i(s))(i) - p_i(s)q^{ii}(s) = 0$ where $\bar{A}$ is something finite. If both equations hold when $q^{ii}(s) → ∞$, we need $p_i(s) → 0$.

- Suppose one trader is restricted from trading and the other one trades infinitely fast. The price is linear in the inventory of the restricted traders. If the restricted trades total inventory is 0, then price is 0. This requires that the marginal cost of holding inventory is 0 for the unrestricted traders which is not consistent with an equilibrium.

- Suppose both groups of traders trade infinitely fast. Then the price is zero and this is inconsistent with an equilibrium as above.

- Suppose one group trades infinitely fast within group and with finite speed with the other group and at some time, the inventory reaches 0 for the group that trades with finite speed. As a result, price goes to 0. Traders who trade infinitely fast can deviate and obtain an inventory of zero at no cost.

As a result, there are no equilibria where any elements of $Z → ±∞$.

Lemma 6. An equilibrium where all traders trade when they are unrestricted is locally unique.

Proof. Suppose there are two solutions, $Z$ and $\hat{Z}$ such that $Z - \hat{Z} = \epsilon\bar{Z}$ for some
small $\varepsilon$. Then we have the equation

$$F(\hat{Z}; \lambda) = F(Z; \lambda) + \varepsilon J(Z; \lambda) \hat{Z}.$$ 

This simplifies to

$$0 = 0 + \varepsilon J(Z; \lambda) \hat{Z}$$

and the unique solution is $\hat{Z} = 0$ since $J(Z; \lambda)$ has full rank. \hfill $\Box$

With the previous lemmas, we can prove proposition 4.

Proof. Suppose there is exist an interval $\lambda \in (\lambda_1, \lambda_2)$. Due to lemma 4 and 5 we know that the if $\lambda_1$ or $\lambda_2$ are finite and larger than 0, one of the second order conditions must hold with equality for $\lambda_1$ and $\lambda_2$ unless $\lambda_1 = 0$ or $\lambda_2 \rightarrow \infty$. This leaves three potential intervals, $[0, \bar{\lambda})$, $[\bar{\lambda}, \infty)$ and $[0, \infty)$. The second condition fails to hold for values of $\lambda$ just above $\bar{\lambda}$ and hence, the only possible interval is $[0, \bar{\lambda})$.

Lemma 4 and 5 shows that if there exist an equilibrium for one $\lambda$, then there exist one for a slightly different $\lambda$. As a result, any equilibrium will follow a continuous path from the equilibrium for $\lambda = 0$ to the one for $\lambda = \bar{\lambda}$. The path is unique from 6. \hfill $\Box$

12.5 Proof of Proposition 5

Use the notation $F(Z)$ as the system of equations. Define $Z_0$ as the solution

To be added. By direct computation. This and the following proofs with approximate solutions is calculated as follows:

- Conjecture that the parameters are on the form $Q(s) = Q + \tilde{Q}\varepsilon$ where $Q$ is the solution to the system without heterogeneity in Result 1.

- Remove terms of order $\varepsilon^2$. The remaining terms are all of order $\varepsilon$. 

47
12.6 Proof of Proposition 6
To be added. By direct computation.

12.7 Proof of Proposition 7
To be added. One dimensional unique numerical solution.

12.8 Proof of Proposition 8
To be added. By direct computation.

12.9 Proof of Proposition 9
To be added. By direct computation.
Figure 3: The plots show the total discounted cash-flow and inventory costs of each type of trader if the state continues in state $s_1$ for $n = \{2, \cdots, 15\}$. Trader 1 has $b = 0.9$ and $X_0 = 1$. Trader 2 has $b = 1.1$ and $X_0 = -1$. The plots for the front-runners are the aggregate values for all $2n - 2$ front-runners.