Managing Inventory with Proportional Transaction Costs *

Florent Gallien[a], Serge Kassibrakis[a], Semyon Malamud[b,c,d] and Filippo Passerini[a]

[a]Swissquote Bank
[b]Swiss Finance Institute
[c]Ecole Polytechnique Fédérale de Lausanne
[d]Centre for Economic Policy Research

This version: June 2, 2016

Abstract

We solve the problem of optimal inventory management for a CARA market-maker who faces proportional transaction costs and marking to market. Our explicit solution accommodates inventory shocks following an arbitrary compound Poisson process, and allows us to explicitly link the optimal policy to the moment-generating function of the shock distribution. We show that the no-trade region is always wider in the presence of shocks, increases with the order imbalance, and usually decreases with the markups charged by the market-maker. We use our explicit solution to derive several comparative statics results and calibrate our solution to inventory data of Forex clients of a bank. Our findings suggest that optimal accounting for inventory shocks leads to significant utility gains.

Keywords: inventory management, market making, transaction costs

JEL Classification Numbers: G24, G32, G11, D92, C61

*We thank Remy Chicheportiche, Julien Hugonnier, Haobo Jia, and Johannes Muhle-Karbe for helpful comments and remarks. Corresponding author email: semyon.malamud@epfl.ch

1 Introduction

Efficient inventory management is pivotal for market-making activities. Market makers and broker/dealers typically need to hold large inventories of risky assets, and thus face the natural risk/liquidity tradeoff: Holding risky assets exposes their balance sheets to shocks, while offloading these inventories may subject them to significant transaction costs, especially when the underlying securities are illiquid. When the client orders arrive sufficiently frequently, it may be optimal for the dealer to wait until an offsetting client order arrives, instead of directly rebalancing and incurring the liquidity cost. Therefore, the problem is to find an inventory management policy that optimally accounts for the distribution of client orders that the dealer anticipates to receive in the future. The goal of this paper is to study this problem.

We consider an economic agent (a broker/dealer) with constant absolute risk aversion (CARA) preferences from consumption, who faces proportional transaction costs for inventory rebalancing. Clients' orders are assumed to arrive according to a compound Poisson process with an arbitrary distribution of order sizes. The dealer charges clients a proportional cost, equal to his cost in the secondary market, plus a markup. In addition, we assume that inventory positioned are continuously marked-to-market.[1] We show that the optimal inventory policy is always characterized by a no-trade region (a band) around the Merton efficient portfolio, and both the agent's value function and the boundaries of the band can be explicitly characterized through the moment-generating function of the distribution of order sizes. It turns out that the boundaries of the no-trade region crucially depend on whether the inventory shocks are "balanced": If the shocks tend to be more positive (negative) on average,

[1]Marking to market is a common practice for most over-the-counter (OTC) derivatives such as credit default swaps (CDS), as well as for futures contracts and currency forwards. Our model is particularly relevant for currency (Forex) markets that are highly decentralized, with smaller dealers managing inventories of their clients and infrequently rebalancing using larger dealers (typically, major investment banks) as liquidity providers. Trades with liquidity providers often imply significant transaction costs, and hence the optimal inventory-management problem is important.

then the upper (lower) bound of the no-trade region coincides with that in the absence of shocks, while the lower (upper) bound is lower (higher) than that in the absence of shocks. If the shocks are balanced, then the no-trade region for large order-arrival intensity coincides with that for the case without shocks. Interestingly enough, the condition for shocks to be "balanced" is non-trivial and is expressed explicitly in terms of the moment-generating function of the order-size distribution.

One important consequence of our results is that it is always optimal to trade (weakly) less (in the sense that the no-trade region is wider) in the presence of inventory shocks, independent of their distribution. The intuition behind our results is as follows. When jumps are balanced and arrive with high intensity, there is no reason for the agent to change the rebalancing strategy because, on average, jumps cancel each other out, and a positive jump will likely soon be compensated by a negative jump. By contrast, if the shocks are positive on average, there is no reason for the agent to increase or decrease the upper boundary of the band because increasing the boundary would make losses from risk exposure even higher, while decreasing it would not help because positive orders would quickly move the inventory position away from the boundary. At the same time, decreasing the lower boundary does make sense, as the agent can spare some transaction costs, and the new arriving positive orders will quickly correct the order imbalance and bring the inventory closer to the optimal position. Interestingly enough, the effects of imbalance become weaker when the order size distribution has fat tails: Namely, when tails are sufficiently fat, the risk of large orders makes it optimal for the agent to shrink the no-trade region and hold minimal amounts of inventory.

Interestingly, our results imply that there is a threshold level of transaction costs such that, above that threshold, the agent behaves as it would without inventory shocks, while below that threshold the inaction region is wider than in the case without inventory shocks.

This result is counterintuitive: One would expect the inventory effect on the no-trade region to be stronger for large, rather than small, transaction costs.

We also show that, most of the time, the no-trade region is monotone decreasing in the markups the dealer charges to clients: That is, a dealer that charges higher markups holds less inventory. Our findings indicate a potentially new channel linking liquidity (markups) with the willingness of the market-makers to hold inventory. This is particularly interesting given the anecdotal evidence for recent changes in government and corporate bonds liquidity, which has often been attributed to the unwillingness of key markers to hold bond inventories on their balance sheets.

One of the key advantages of our solution is its ability to handle any distribution of order sizes. To test our results, we use proprietary inventory data of an online broker to estimate the actual distribution of order sizes and their arrival intensity, and use our optimal solution to find the optimal rebalancing policy. The flexibility of our solution is important because it allows us to match the distribution non-parametrically. We find that the policy indeed leads to significant utility gains.

We now discuss related literature.

Portfolio management problems with proportional transaction costs have been studied in numerous papers. For example, Constantinides (1986), Eastham and Hastings (1988), Dixit (1989), Fleming et al (1990), Davis and Norman (1990), Shreve et al. (1991), Dixit (1991), Dumas and Luciano (1991), Akian et al. (1996), Balduzzi and Lynch (2000), Liu (2004),[2] Jang et al. (2007), Dai et al. (2011), Lynch and Tan (2011), Soner and Touzi (2013), Shreve and Bichuch (2013), Guasoni et al. (2014), Dumas, Delgado, and Puopolo (2015), and Muhle-Karbe and Kallsen (2015), and Passerini and Vazquez (2015) all show that the optimal policy is characterized by a no-trade region around the frictionless optimal

[2] As in our model, Liu (2004) assumes that the agent maximizes CARA preferences from intermediate consumption. However, in his model there is no marking-to-market, prices follow geometric Brownian motions, and there are no inventory shocks.

portfolio:[3] The agent does not trade when his portfolio position is inside the band, and trades minimally to keep the portfolio inside the band. [4] None of these papers consider the problem of optimal inventory management in the presence of inventory shocks. Furthermore, most of them either assume that transaction costs are infinitesimally small, or derive results that are purely numerical. By contrast, in our paper we characterize the optimal policy for any magnitude of transaction costs and derive analytic properties of the optimal policies explicitly in terms of the characteristic function of the order size distribution. This approach allows us to establish explicit comparative statics results.

Our model predicts that dealers have a target inventory level (the Merton portfolio) to which they mean-revert their inventories; immediately rebalancing is suboptimal due to transaction costs. Hasbrouck and Sofianos (1993), Madhavan and Sofianos (1998), Hansch, Naik, and Viswanathan (1998), Reiss and Werner (1998), and Naik and Yadav (2003) find evidence that market makers indeed manage their inventories by mean-reverting their positions towards a target level. Furthermore, Madhavan and Smidt (1993) find evidence that specialist inventories deviate from their target levels for long periods (up to several weeks), in agreement with our basic findings. Hansch, Naik, and Viswanathan (1998) and Reiss and Werner (1998) also find evidence that differences in inventories across dealers affect the intensity of interdealer trading. The latter finding is consistent with the predictions of Ho and Stoll (1983) for a market with competitive dealers. Our model can be viewed as a first step to incorporating an illiquid secondary market into inventory models of asset pricing and dealer trading. While our analysis is partial equilibrium, the techniques developed in our paper can be extended to general equilibrium effects. In particular, our model allows us

[3] In some of these papers (such as, e.g., Balduzzi and Lynch (2000), Jang et al. (2007), Lynch and Tan (2011), Dumas, Delgado and Puopolo (2015)), investment opportunity set moves over time, and returns are predictable, so that the frictionless optimal portfolio is also stochastic. Furthermore, some papers (such as Passerini and Vazquez, 2015) consider trading with both market and limit orders.

[4] Several papers (such as, e.g., Grinold, 2006; Garleanu and Pedersen, 2013; and Collin-Dufresne et al., 2012) investigate the problem with quadratic trading costs. The structure of the solution is quite different in this case, and there is no existing no-trade band.

to link inventory rebalancing policies to liquidity and markups charged by different dealers. As we explain above, our model is particularly well suited to the Forex market, which has a multi-layer structure (a network) with smaller (periphery) dealers absorbing clients' orders, while central dealers need to absorb orders of both their clients and the periphery dealers. It can also be applied to models of dealer networks (for example, for credit default swaps (Atkeson, Eisfeldt, and Weill, 2015), and municipal bonds (Green, Hollified, and Schürhoff, 2007a,b; Li and Schürhoff, 2014), where dealers have to absorb inventory shocks, charge markups to their clients, and face markups charged by more central dealers.

Several papers (Garman, 1976; Amihud and Mendelson, 1980; Ho and Stoll, 1981, 1983) predict that inventories affect the level of bid and ask prices, with market makers adjusting their quotes up (down) to attract client sell (buy) orders when their inventory is too low (too high). Empirical evidence for such quote adjustments is mixed: Some authors have found these adjustments in foreign exchange markets (Lyons, 1995; Cao, Evans, and Lyons, 2006) and option markets (Garleanu, Pedersen, and Poteshman, 2008), while there seems to be evidence that dealers in the futures markets exhibit the opposite behavior, and adjust quotes up when their inventory is high (Manaster and Mann, 1996). Furthermore, Hansch, Naik, and Viswanathan (1998) and Reiss and Werner (1998) find that individual dealer inventories do not affect the magnitude of each dealer's intraday quotes. A quote adjustment might be particularly problematic when there is marking-to-market (as in foreign exchange, futures, or OTC derivatives markets). In this case, some dealers might only adjust the fees they charge to clients. Pushing these fees up in response to inventory shocks might be problematic due to existing agreements, and present the threat of loosing clients to competitors.

Finally, there is also literature that builds on the Kyle (1985) model of strategic trading with competitive market makers. For example, Vayanos (2001) considers the problem of a strategic agent (insider) trading with a competitive fringe of CARA market makers in which market makers are forced to absorb inventory shocks from the strategic trader, but face no

transaction costs. Incorporating our partial equilibrium solution into a general equilibrium model of inventory management with transaction costs is an important topic for future research.

2 The Problem

An agent (a financial intermediary/broker) manages in continuous time an inventory of a risky asset whose price P_t follows an arithmetic Brownian motion

$$dP_t \;=\; \mu\, dt \;+\; \sigma\, dB_t\,.$$

The risk-free rate is fixed and is given by $r > 0$.

The inventory of the risky asset is denoted by q_t. We assume that client orders arrive according to a compound Poisson process N_t with intensity λ, and jump sizes that are drawn from a distribution $d\eta(x)$. A positive (negative) realization of the jump corresponds to a client's sell (buy) order. The agent also has an option to contact a liquidity provider (or a secondary market) and trade an amount dM_t. As a result, the agents' inventory follows

$$dq_t = dN_t + dM_t.$$

A key assumption in our model is that the asset value is continuously marked-to-market so that the value of the contract is always equal to zero. This is indeed the case for futures contracts, most currency forwards and currency derivatives, as well as many OTC derivative contracts (such as, e.g., CDS). Since the asset value is always zero, neither the clients (when contacting the market maker) nor the market maker himself (when contacting the liquidity provider) pay anything except for the brokerage fee. In agreement with real-world practice, we assume that the fees are proportional: The agent charges his clients a fee that

is proportional to the order flow, given by $\alpha|dN_{kt}|$. At the same time, the liquidity provider also charges the agent a proportional fee, given by $\alpha_l|dM_{kt}|$.[5] The difference between the fees, $\alpha - \alpha_l > 0$, is the mark-up changed by the market maker to his clients. This markup is an important component of market maker's profits.

The total cash flows dC_t of the agent can, therefore, be decomposed as $dC_t = dL_t + d\nu_t$, where the first part is the result of a directional exposure of the existing inventory due to marking-to-market,

$$dL_t = q_t dP_t \,,$$

whereas the second part is composed of the net markup revenues

$$d\nu_t = \alpha|dN_t| \;-\; \alpha_l|dM_t| \,.$$

We assume that the agent maximizes constant absolute risk aversion (CARA) utility from intermediate consumption[6]

$$V \;=\; \max_{\{c_t, dL_t\}} E\left[\int_0^\infty -e^{-\beta t} e^{-\gamma c_t} dt\right]$$

subject to the budjet constraint

$$dW_t \;=\; (rW_t - c_t)dt \;+\; dL_t \;+\; d\nu_t$$

on his wealth W_t.

[5] We interpret the price P_t as the mid-point of the bid-ask interval. In this case, the agent is facing the bid-ask prices $P_t - \alpha_l,\ P_t + \alpha_l$ and he offers the bid-ask prices of $P_t - \alpha, P_t + \alpha$ to his clients.

[6] This assumption is standard in the literature. See, e.g., Vayanos (1999). In the context of a broker/dealer, one can interpret consumption as dividends, or salary/bonus payments.

3 HJB equation

Since the price of the asset follows an arithmetic Brownian motion and inventory shocks are i.i.d., the agent's value function V depends only on the agent's wealth w and the current inventory level q : $V = V(w, q)$. To solve the agent's optimization problem, we follow the standard dynamic programming approach and derive the Hamilton-Jacobi-Bellman equation for the value function. As is common in models with proportional transaction costs, we can decompose the state space into the *no-trade region* and the *trading region:* In the no-trade region, the agent only saves and consumes; in the trading region, however, he immediately contacts the liquidity provider and adjusts his inventory position to end up in the no-trade region.

The HJB equation in the no-trade region only involves optimization over the consumption rate and can be written as

$$\begin{aligned} \sup_c \Bigg\{ & -e^{-\gamma c} + (rw - c + q\mu)V_w + 0.5\sigma^2 q^2 V_{ww} - \beta V \\ & + \lambda \int (V(w + \alpha|x|, q + x) - V(w, q)) d\eta(x) \Bigg\} \;=\; 0\,. \end{aligned} \tag{1}$$

Here, $-e^{-\gamma c}$ is the flow of marginal utility, $(rw - c + q\mu)V_w$ is the infinitesimal change in V due to the drift $rw - c + q\mu$ in the agent's wealth, $0.5\sigma^2 q^2 V_{ww}$ is the infinitesimal change in the value function due to volatility $q\sigma$ of the cash flows generated by marking-to-market, $-\beta V$ is a time discount term, and the last term is the expected change in the value function due to the jumps in the agent's inventory. In the trading region, the value function satisfies the condition

$$\max_{\bar{q}} (V(w - \alpha_l|q - \bar{q}|, \bar{q}) - V(w, q)) = 0\,,$$

which simply ensures that it is indeed optimal for the agent to contact the liquidity provider,

costing him $\alpha_l|q-\bar{q}|$ units of his wealth. Define the rescaled transaction costs

$$\varphi \;=\; r\gamma\alpha_l \;,\; \psi \;=\; r\gamma\alpha$$

paid by the agent to the liquidity provider, and by the clients to the agent, respectively. We will solve the HJB equation by making the standard Ansatz

$$V(w,q) \;=\; -e^{-r\gamma w + a(q)}\,.$$

Here, the function $a(q)$ is the (negative of the) certainty equivalent of the agent's value function, and can thus be used to measure utility gains from the use of different policies.

Substituting the Ansatz into the HJB equation, the optimal consumption rate in the no-trade region is given by

$$c \;=\; -\frac{\log r}{\gamma} \;+\; rw \;-\; \frac{a(q)}{\gamma}\,,$$

whereas $a(q)$ satisfies the following integral equation in the no-trade region:

$$-r + r(\log r + a(q)) + r\gamma q\mu - 0.5\sigma^2 q^2 (r\gamma)^2 + \beta + \lambda \int_{\mathbb{R}} (-e^{-\psi|x| + a(q+x) - a(q)} + 1) d\eta(x) \;=\; 0\,. \tag{2}$$

Finally, the optimality condition in the trading region can be rewritten as

$$a(q) \;=\; \min_{\bar{q}} (a(\bar{q}) + \varphi|q-\bar{q}|)\,.$$

In addition, we need to make sure that the staying at a given inventory level in the trading region for an infenitesimal amount of time leads to a marginal utility loss. Formally, this

means that the inequality

$$-r + r(\log r + a(q)) + r\gamma q\mu - 0.5\sigma^2 q^2 (r\gamma)^2 + \beta + \lambda \int_{\mathbb{R}} (-e^{-\psi|x| + a(q+x) - a(q)} + 1) d\eta(x) \;\leq\; 0$$

holds for q in the trading region. Summarizing, we expect that $a(q)$ satisfies the HJB equation

$$\begin{aligned} &\max \Big\{ -r + r(\log r + a(q)) + r\gamma q\mu - 0.5\sigma^2 q^2 (r\gamma)^2 + \beta \\ &+ \lambda \int_{\mathbb{R}} (-e^{-\psi|x| + a(q+x) - a(q)} + 1) d\eta(x) \;,\; a(q) - (a(\bar{q}) + \varphi|q - \bar{q}|) \Big\} = 0 \,. \end{aligned} \tag{3}$$

In general, solving the HJB equation (3) is a non-trivial task. However, as we show below, it is possible to prove that this equation has a unique solution with natural economic properties.

We start our analysis of the HJB equation (3) with the simple observation that, absent transaction costs, the optimal policy is to keep the inventory at the Merton portfolio

$$q^* \;=\; \frac{\mu}{r\gamma\sigma^2} \,.$$

In this case, $a(q)$ is constant and independent of q, and the agent behaves myopically and simply maximizes the instantaneous risk-return tradeoff $r\gamma q\mu - 0.5\sigma^2 q^2 (r\gamma)^2$. The following is true.

Proposition 1 *If $\varphi = 0$ then the optimal policy is to always keep inventory at q^*.*

In the presence of transaction costs, we conjecture (and later verify) that the optimal inventory policy is characterized by a single band $[q_L, q_H]$ around the Merton portfolio. The optimal policy is to keep the inventory within the band, and to always rebalance to the nearest boundary of the band when the inventory is outside of the band. The following is true.

Theorem 2 *The optimal policy is of a single band type: there exist $q_L < q_H$ and a convex function $a(q) \in C^1(\mathbb{R})$ satisfying (3) and such that*

- *the value function is given by $V(w,q) = -e^{-r\gamma w + a(q)}$*
- *we have*

$$a(q) = a(q_H) + \varphi(q - q_H),\ q > q_H;\ a(q) = a(q_L) + \varphi(q_L - q),\ q < q_L$$

- *the function $a(q)$ satisfies (3) for $q \in [q_L, q_H]$ together with the boundary conditions $-a'(q_L) = a'(q_H) = \varphi$.*

The quantity $|a'(q)|$ measures the marginal disutility of inventory holdings. As long as this marginal disutility is below the marginal cost φ of contacting the liquidity provider, the agent does not take any action. Since $a(q)$ is convex, the marginal disutility monotonically increases from $-\varphi$ to φ when q increases from q_L to q_H, and the agent trades every time this marginal disutility hits φ in absolute value. In order to understand the structure of this marginal disutility, let us first consider the simpler case in which the order arrival intensity is equal to zero: $\lambda = 0$. We denote the corresponding disutility function by $a_0(q)$ and the no-trade region by $[q_L(0), q_H(0)]$. In this case, (2) implies $a_0(q)$ is a quadratic function of q : $a_0(q) = \Phi(q)$ for $q \in [q_L(0), q_H(0)]$, where we have defined

$$\Phi(q) = 1 - \log r + \gamma(-q\mu + 0.5\sigma^2 q^2 r\gamma) - \beta r^{-1}\,.$$

This is intuitive: in a CARA-normal setting, the marginal disutility is (up to a constant) of a mean-variance form. In particular, in solving the boundary conditions $-a_0'(q_L) = a_0'(q_H) = \varphi$, we get that the no-trade region is

$$q_{L,H}(0) = q_* \pm \frac{\varphi}{\gamma^2 r\sigma^2} = q_* \pm \frac{\alpha_l}{\gamma\sigma^2}\,.$$

Thus, the no-trade region is symmetric around the Merton portfolio q_*, and the width of the region is inversely proportional to the agent's risk aversion and to the volatility σ^2. This is consistent with the existing results: As was first noted by Constantinides (1986), and further confirmed by most existing papers in the literature, with a constant investment opportunity set lower, risky asset volatility always leads to a wider no-trade region because a less volatile risky asset return means a smaller utility loss from maintaining sub-optimal risk exposure. Interestingly, the width of the no-trade region is proportional to the transaction cost α_l and not to $\alpha_l^{1/3}$, as is common in the existing models with proportional transaction costs (see, e.g., Muhle-Karbe and Kallsen, 2015). As Muhle-Karbe and Kallsen (2015) explain, the width of the no-trade region is driven by the realized quadratic variation of the frictionless (Merton) portfolio that the agent is trying to track—the more volatile the target portfolio is, the more costly it is for the agent to closely track this portfolio. In our model, the fact that the price follows an arithmetic Brownian motion, combined with the CARA preferences, implies that the target portfolio is constant, and hence has a zero quadratic variation. Thus, tracking this portfolio is "easy" and it is optimal for the agent to choose a narrow no-trade zone.

Differentiating (2), we get the identity

$$a'(q) \;=\; \Phi'(q) \;+\; \frac{\lambda}{r}\int_{\mathbb{R}}(a'(q+x)-a'(q))e^{-\psi|x|+a(q+x)-a(q)}d\eta(x)\,. \tag{4}$$

This identity shows that the marginal disutility is given by the sum of two terms: the marginal disutility $\Phi'(q)$ from holding the inventory level q, and the expected average change in the marginal disutility due to potential future jumps in the inventory level. At the inventory level q_H, we have $a'(q_H) = \varphi \geq a'(q_H + x)$ for all $x \in \mathbb{R}$: Indeed, when the marginal disutility is already at its highest, a jump in the inventory can only push this marginal value down. As a result, we get the inequality $\varphi = a'(q_H) \leq \Phi'(q_H)$, and the same argument implies that $-\varphi = a'(q_L) \leq \Phi'(q_L)$. As we explain above, in the case of no

inventory shocks, the no-trade boundaries satisfy $-\Phi'(q_L(0)) = \Phi'(q_H(0)) = \varphi$. Since Φ is convex, the inequality $\Phi'(q_H(0)) = \varphi = a'(q_H) \leq \Phi'(q_H)$ implies that $q_H \geq q_H(0)$, and a similar inequality holds for q_L. As a result, the no-trade region is always wider in the presence of inventory shocks. We formalize this important observation in the following result.

Proposition 3 *The presence of inventory shocks always widens the no-trade region. That is, we always have* $q_H \geq q_H(0) > q_L(0) \geq q_L$.

At first glance, the result of Proposition 3 is surprising. Indeed, suppose for example that inventory shocks are positive most of the time, so that the agent anticipates his inventory will grow in the near future. Intuitively, one might expect the agent to shift the entire no-trade region to the left because it is less risky to hold large negative inventory if we expect that this inventory will soon jump up (closer to zero) with a high probability. While this intuition is correct for the lower limit y_L, it is not correct for the upper limit y_H, which can only move to the right. This is so because it does not make sense for the agent to rebalance down when inventory is high if he anticipates that it will soon jump up again. In fact, the following is true.

Proposition 4 *If order size is always positive, then* $q_H = q_H(0)$. *If order size is always negative, then* $q_L = q_L(0)$.

We complete this section with an interesting equivalence result that highlights the role of the markups charged by the agent to his clients for optimal trading behaviour. Since the markups are only charged upon an order arrival, they only appear in the HJB equation through the term responsible for inventory jumps. Equation (2) implies that the markups effectively act as an equivalent measure change of distribution of the jumps, whereby the measure $d\eta(x)$ is replaced by $e^{-\psi|x|}d\eta(x)$. Of course, the latter is not a probability measure and needs to be normalized by $\int e^{-\psi|x|}d\eta(x) < 1$. We summarize this observation in the following proposition.

Proposition 5 *Let $\psi_1 > \psi_2$. The behavior of an agent with markup ψ_1 coincides with that of an agent with marup ψ_2 facing a distribution of inventory shocks given by*

$$\left(\int e^{-(\psi_1-\psi_2)|x|}d\eta(x)\right)^{-1} e^{-(\psi_1-\psi_2)|x|}d\eta(x)$$

and the order arrival intensity $\lambda\int e^{-(\psi_1-\psi_2)|x|}d\eta(x) < \lambda$.

Proposition 5 shows that an agent charging higher markups to his clients behaves in exactly the same way as an agent with a lower markup, lower order arrival intensity, and "thinner" tails of the order size distribution. Intuitively, this suggests that the no-trade region should decrease in ψ because large ψ makes the effect of inventory jumps negligible. As we show below, this is indeed always the case when λ is small, but it fails to be the case for large λ.

4 Small Order Arrival Intensity

In this section, we study the case when the order arrival intensity is small. Since the case of zero intensity has been explicitly solved above, we can use the Taylor approximation to compute the approximate no-trade region when λ is small. The following is true.

Proposition 6 *When λ is small, the no-trade region is given by*

$$\begin{aligned} q_L &= q_L(0) + \lambda\, q_L(1) + O(\lambda^2) \\ q_H &= q_H(0) + \lambda\, q_H(1) + O(\lambda^2) \end{aligned}$$

where

$$
\begin{aligned}
q_L(1) &= -\frac{\varphi}{\rho r} e^{\Phi(q_H(0))-\Phi(q_L(0))+\varphi(q_L(0)-q_H(0))} \int_{q_H(0)-q_L(0)}^{\infty} e^{-\psi x+\varphi x} d\eta(x) \\
&\quad -\frac{1}{r} \int_0^{q_H(0)-q_L(0)} e^{-\psi x+x(-\gamma\mu+\rho(x+2q_L(0)))} x \; d\eta(x) \\
q_H(1) &= \frac{\varphi}{\rho r} e^{\Phi(q_L(0))-\Phi(q_H(0))+\varphi(q_L(0)-q_H(0))} \int_{-\infty}^{q_L(0)-q_H(0)} e^{\psi x-\varphi x} d\eta(x) \\
&\quad -\frac{1}{r} \int_{q_L(0)-q_H(0)}^{0} e^{\psi x+x(-\gamma\mu+\rho(x+2q_H(0)))} x \; d\eta(x)
\end{aligned}
$$

and where we have defined $\rho = 0.5\sigma^2\gamma^2 r$.

The following corollary is a direct consequence of Proposition 6.

Corollary 7 *For small* λ*, the no-trade region is*

- *increasing in* λ
- *decreasing in* ψ

The monotonicity in λ perfectly agrees with the results of Proposition 3: inventory shocks always widen the no-trade region. Monotonicity in ψ also agrees with the above intuition: The no-trade region decreases in ψ because large ψ makes the effect of inventory jumps negligible. As we show below (Corollary 12), this monotonicity fails to hold when λ is large.

We now discuss the dependence of the no-trade region on the distribution of order sizes. Suppose first, that order-size distribution is concentrated on the positive side. We would expect that the lower boundary of the no-trade region would be lower because it would be optimal for the agent to reduce trading costs through widening the no-trade region, and then wait until a positive order arrives and corrects the sub-optimal inventory position. Recall that a first-order stochastic dominance (FOSD) shift in a distribution means that some of

the mass of the distribution is shifted to the right. The following proposition shows that this intuition only holds true for small order sizes.

Proposition 8 *Suppose that λ is sufficiently small. Then,*

- *If the order size is concentrated on a small neighbourhood of zero, then a small change in η in the FOSD sense shifts the no-trade region to the left;*
- *If the order size distribution is supported on $\mathbb{R} \setminus [-L, L]$ for a sufficiently large L, then a small change in η in the FOSD sense shifts the no-trade region to the right.*

Why does the behaviour of the no-trade region change for large orders? Interestingly, this effect is driven by the markups charged by the agent, and the strength of the effect is determined by the size of the markup, as measured by $\psi - \varphi > 0$. When order sizes are large, the agent knows for sure that, upon order arrival, he will find himsel deep inside the trading region. Thus, the major contribution to the utility from an order of size x will be given by $e^{-(\psi-\varphi)x}$. Suppose for simplicity that only positive orders arrive, and these orders are of a very large size. Then, (4) implies that, when λ is sufficiently small,

$$\begin{aligned} a'(q) &= \Phi'(q) + \frac{\lambda}{r} \int_{\mathbb{R}} 2\varphi e^{-\psi + a_0(q+x) - a_0(q)} d\eta(x) \\ &= \varphi + \frac{\lambda e^{a_0(q_H(0)) - a_0(q) + \varphi(q_H(0) - q)}}{r} \int_{\mathbb{R}} 2\varphi e^{-(\psi-\varphi)x} d\eta(x) \end{aligned}$$

with $q = q_L(0)$. When jump size x is larger, the log utility loss $-(\psi - \varphi)x$ is larger, and hence so is the expected marginal disutility $a'(q)$. Therefore, the solution q_L to $a'(q_L) = -\varphi$ gets larger when jump size increases.

5 Large Order Arrival Intensity

While the case of small-order arrival intensity provides interesting insights, in reality, order arrival intensities are quite high for many broker/dealers. In this section, we show that,

surprisingly, when λ is large, the integral equation (2) can be reduced to a simple first-order ordinary differential equation (ODE) that depends explicitly on the moment-generating function of the order-size distribution. This allows us to derive various useful properties of the solution; in particular, we show that the behavior of the optimal no-trade region may be very different from that of the case of a small-order arrival intensity.

The key insight comes from the following observation. When λ is very large, the total inventory changes at the rate λ, hence it makes sense to work directly with the *rescaled inventory* $y = q/\lambda$. We will also define

$$\bar{\rho} \;\equiv\; \lambda\rho\,,$$

where, as above, $\rho = 0.5\sigma^2\gamma^2 r$ is the coefficient of $\Phi(q)$ corresponding to q^2.[7] Let also $A(y) = a(\lambda y)/\lambda$. Substituting these expressions into (2), we arrive at the following equation.

$$-r + r(\log r + \lambda A(y)) + r\gamma\lambda y\mu - r\bar{\rho}\lambda y^2 + \beta + \lambda\int_{\mathbb{R}}(-e^{-\psi|x|+\lambda(A(y+x/\lambda)-A(y))}+1)d\eta(x) \;=\; 0\,. \tag{10}$$

With respect to the rescaled inventory, the contribution of a jump is marginal. At the same time, the disutility function itself becomes large because the agent risks accumulating a very large inventory position. These two effects offset each other as λ becomes larger and we get

$$\lambda(A(y+x/\lambda) - A(y)) \;=\; x\frac{A(y+x/\lambda)-A(y)}{x/\lambda} \approx\; A'(y)x\,.$$

As a result, dividing (10) by λ and sending $\lambda \to \infty$, we (formally) arrive at the differential

[7] Note that this means that we are assuming that quantity of risk $\gamma\sigma$ is of the order of $1/\lambda^{1/2}$ (or smaller).

equation

$$r(A(y) - \bar{\Phi}(y)) \; - \; \Psi(A'(y)) \; = \; 0, \tag{11}$$

where

$$\Psi(z) \; \equiv \; \int_{\mathbb{R}} e^{-\psi|x|+xz} d\eta(x)$$

is the moment-generating function of the *markup-adjusted order size distribution* $e^{-\psi|x|}d\eta(x)$, and

$$\bar{\Phi}(y) \; \equiv \; \bar{\rho}y^2 - \gamma\mu y \; - 1/r \, .$$

We also define

$$y_{H,L}(0) \; \equiv \; \frac{\gamma\mu \pm \varphi}{2\bar{\rho}} \; = \; \frac{q_{H,L}(0)}{\lambda} \, .$$

Furthermore, we expect that for large λ, the boundaries of the no-trade region are also roughly proportional to λ : that is, there exist $\bar{y}_L, \bar{y}_H$ such that $q_{H,L} \approx \lambda \bar{y}_{H,L}$ for sufficiently large values of λ. Furthermore, $a'(\lambda y) = A'(y)$, and hence we expect that $A(y)$ satisfies the boundary conditions:

$$-A'(y_L) = A'(y_H) = \varphi. \tag{12}$$

As will we now show, this intuition is indeed correct. Most importantly, the thresholds $\bar{y}_i, \; i = H, L$ can be characterized explicitly. In order to state the next result, we will need the following auxiliary lemma.

Lemma 9 *The function Ψ is convex and satisfies $\Psi'(0) > 0$ if and only if $\int e^{-\psi|x|} x d\eta(x) > 0$. Furthermore, Ψ is monotone increasing (decreasing) if jumps are always positive (negative).*

The key insight from Lemma 9 is that the derivative of Ψ is monotone increasing and tends to be positive (negative) when there is an order imbalance towards positive (negative) orders. This insight will be crucial to the subsequent analysis. Denote by $A_\pm(x;b)$ the unique smooth solution to

$$\Psi(A'_\pm(x;b)) \;=\; r(A_\pm(x;b) - \bar{\Phi}(x)),\;\; A'_\pm(b;b) = \pm\varphi. \tag{13}$$

Let also $\tilde{y}$ be the unique solution to $A_+(\tilde{y}; y_H(0)) = A_-(\tilde{y}; y_L(0))$.

We can now state the main result of this section.

Theorem 10 *In the limit as $\lambda \to \infty$ the function $a(\lambda y)/\lambda$ and the quantities $q_i/\lambda,\ i = H, L$ converge to finite limits $A(y)$ and $\bar{y}_H,\ \bar{y}_L$ respectively. Furthermore,*

(1) If $\Psi'(\varphi) \geq 0 \geq \Psi'(-\varphi)$ then $\bar{y}_i = y_i(0),\ i = H, L$; and $A(y) = \mathbf{1}_{y \leq \tilde{y}} A_-(y; y_L(0) + \mathbf{1}_{y \leq \tilde{y}} A_+(y; y_H(0))$.

(2) If $\Psi'(\varphi) > \Psi'(-\varphi) \geq 0$, then $\bar{y}_H = y_H(0)$, while $\bar{y}_L$ is the unique solution to $A'_+(\bar{y}_L; y_H(0)) = -\varphi$; and $A(y) = A_+(y; y_H(0))$.

(3) If $0 \geq \Psi'(\varphi) > \Psi'(-\varphi)$, then $\bar{y}_L = y_L(0)$, while $\bar{y}_H$ be the unique solution[8] to $A'_-(\bar{y}_H; y_L(0)) = \varphi$; and $A(y) = A_-(y; y_L(0))$.

As we already know from Proposition 4, when order sizes are non-balanced, one of the thresholds will be closer (or even coincide) with the λ-zero threshold, corresponding to the direction to which the order distribution is tilted. Theorem 10 allows us to quantify the meaning of "imbalance" in a fully explicit manner, in terms of the moment-generating

[8] See Figure 1 for an illustration of the construction of the optimal threshold $\bar{y}_H$.

function Ψ. In particular, if the order size distribution is balanced, in the sense that both positive and negative orders arrive with comparable probabilities, then the behaviour of the agent coincides with that for the case of $\lambda = 0$: When λ is high, inventory position moves up and down so frequently that it does not make sense for the agent to adjust the no-trade region to account for future orders. At the same time, when most of the arriving orders are positive, it makes sense for the agent to tilt the inaction region to the left. According to item (2) of Theorem 10, the size of this tilt is determined by the difference between $y_L(0)$ and $\bar{y}_L$. Figure 2 illustrates the dependence of the optimal no-trade region on the expected order size imbalance: As one can see, there is a threshold level of imbalance, such that the no-trade region strictly monotonically increases in the size of imbalance when the latter is above the threshold level.

Include Figure 2 somewhere here

The flexibility of Theorem 10 allows to apply it to an arbitrary distribution. In particular, we can investigate the potential effects of fat tails on the size of the no-trade region. To address this question, we assume that the order size distribution is given by the student-t distribution with a given number ν of degrees of freedom. Importantly, when $\nu < 2$, this distribution does not even have a finite second moment, while for $\nu < 1$ is does not have a finite first moment. Figure 3 shows the dependence of the no-trade region on the degree of fatness of the tail, ν. As $\nu \to \infty$, the distribution converges to a Gaussian one, and the no-trade region converges to that in Figure 1. However, as tails get fatter, the no-trade regions gets narrower and eventually converges to the naive region when $\nu \downarrow 0$. The intuition behind this result is clear: fat tails make holding inventory extremely risky, and the marginal effect of imbalance vanishes. Hence, the agent finds it optimal to hold minimal amounts of inventory and just manage it in a naive way, only accounting for the instantaneous risk-return tradeoff.

Include Figure 3 somewhere here

Our results above indicate that accounting for order-size imbalance may have a large impact on the optimal policies. The next natural question is about the magnitude of the utility gains from using the optimal policy, relative to the naive policy corresponding to $\lambda = 0$. Suppose, for example, that the order size imbalance is negative so that $0 > \Psi'(\varphi)$. In complete analogy to Proposition 4, it is possible to show that the certainty equivalent $a_{naive}(q)$ of the value function for an agent following the naive policy of always keeping the inventory inside the $[q_L(0), q_H(0)]$ region satisfies $\liminf_{\lambda\to\infty} a(\lambda y)/\lambda \geq A_{naive}(y)$ as $\lambda \to \infty$ where $A_{naive}(y) = A(y),\ y \in [y_L(0), y_H(0)]$, while $A_{naive}(y) = A_{naive}(y_H(0)) + \varphi(y - y_H(0))$ for $y \in [y_H(0), \bar{y}_H]$.[9] The utility gains from using the optimal policy are given by the difference between $A(y)$ and $A_{naive}(y)$. As we can see from Figure 4, these utility gains can be quite large for large inventory levels, whereby the utility losses due to a larger risk exposure are more than compensated for by the gains from less frequent rebalancing.

Include Figure 4 somewhere here

The size of the gap between $\bar{y}_L$ and $y_L(0)$ is determined by the optimal tradeoff between the effective cost φ of rebalancing and the risk of accumulating a large exposure to the risk/return tradeoff, captured by the quadratic function $\bar{\Phi}$. When φ is small, what matters is the sign of $\Psi'(0) = \int_{\mathbb{R}} e^{-\psi|x|} x d\eta(x)$. This object is the expected order size, calculated with respect to the markup-adjusted distribution $e^{-\psi|x|}d\eta(x)$, and the decision of whether to tilt the no-trade region is determined by the sign of this expectation. However, when rebalancing costs are high, trading-cost considerations dominate the inventory management considerations, and it is optimal for the agent to ignore the expectations of future shocks. The following is true.

Corollary 11 *Let φ^* be the solution to $\Psi'(-\varphi^*) = 0$ if $\Psi'(0) = \int_{\mathbb{R}} e^{-\psi|x|} x d\eta(x) > 0$ and be*

[9] This is a direct consequence of the comparison theorem for ODEs because in the limit as $\lambda \to \infty$ the function $a(\lambda y)/\lambda$ converges to a solution $\tilde{A}$ to the ODE (13) with $\tilde{A}'(y_L(0)) \geq -\varphi$.

the solution to $\Psi'(\varphi^*) = 0$ *if* $\Psi'(0) < 0$. *Then, if* $\varphi > \varphi^*$ *then the inaction region coincides with that for the case without inventory shocks.*

The explicit characterization of the disutility function $A(y)$ provided in Theorem 10 allows us to derive some intuitive comparative statics results. We start with the dependence on φ.

Proposition 12 $\bar{y}_H, -\bar{y}_L$ *are monotone increasing in* φ.

In the limit when $\varphi \to 0$, *we have:*

- *If* $\Psi'(0) > 0$, *then*

$$\bar{y}_L \approx -2\varphi^{1/2}\left(\frac{r\bar{\rho}}{\Psi'(0)}\right)^{-1/2} + \frac{\gamma\mu}{2\bar{\rho}} + O(\varphi)$$

- *If* $0 > \Psi'(0)$, *then*

$$\bar{y}_H \approx 2\varphi^{1/2}\left(\frac{r\bar{\rho}}{-\Psi'(0)}\right)^{-1/2} + \frac{\gamma\mu}{2\bar{\rho}} + O(\varphi)$$

Not surprisingly, the no-trade region is monotone, increasing in φ. However, Proposition 12 illustrates the tradeoff between rebalancing costs φ, risk exposure $\bar{\rho}$, and order-size imbalance measure $\Psi'(0)$. In addition, the inaction region is asymmetric, with one side being concave in the size of the transaction cost, while the other side is linear in φ. Figure 5 illustrates the concave shape of $\bar{y}_H(\varphi)$.

Include Figure 5 somewhere here

We complete this section with a discussion of the dependence of the no-trade region on ψ. In contrast to φ, the dependence on the no-trade region on the markup ψ can be very different from that in the case of small λ : While Corollary 7 shows that the inaction region is always decreasing in ψ when λ is small, this is no longer the case when λ is large.

Corollary 13 *The following is true:*

(1) If $\Psi'(\varphi) > \Psi'(-\varphi) \geq 0$, *then*

(a) if $\int_{\mathbb{R}} e^{-\psi|x|}|x|xe^{x\varphi}d\eta(x) < 0$ *then* $\bar{y}_L$ *is locally decreasing in* ψ

(b) if $\int_{\mathbb{R}} e^{-\psi|x|}|x|xe^{-x\varphi}d\eta(x) > 0$ *then* $\bar{y}_L$ *is locally increasing in* ψ

(2) If $0 \geq \Psi'(\varphi) > \Psi'(-\varphi)$, *then*

(a) if $\int_{\mathbb{R}} e^{-\psi|x|}|x|xe^{x\varphi}d\eta(x) < 0$ *then* $\bar{y}_H$ *is locally decreasing in* ψ

(b) if $\int_{\mathbb{R}} e^{-\psi|x|}|x|xe^{-x\varphi}d\eta(x) > 0$ *then* $\bar{y}_H$ *is locally increasing in* ψ

Figure 6 illustrates the dependence of the no-trade region on the markup level ψ : for a typical order-size distribution, the condition $0 \geq \Psi'(\varphi)$ (i.e., negative imbalance) implies that the hypothesis of (2)(a) is satisfied, and hence $\bar{y}_H$ should be decreasing in ψ. Figure 6 shows that, in fact, the effect of ψ on the optimal policy can be quite large.

Include Figure 6 somewhere here

The intuition behind Figure 6 is straightforward. As in the case of small jump size (Corollary 7), large ψ reduces the impact of inventory shocks on the agent's utility. Indeed, by Proposition 5, a larger ψ effectively means a lower λ.

It is interesting to discuss the potential link between the findings in Corollary 13 and the post-crisis dynamics of bond market liquidity. As some recent research suggests,[10] bond liquidity (defined either through turnover or through the costs of executing large orders) has dropped, which is associated with the willingness of key markers to hold bond inventories on their balance sheets. Our findings indicate a potentially new endogenous channel linking liquidity (markups) with the willingness of the market-makers to hold inventory.

[10] See, e.g., https://ir.citi.com/jKOkPb5qgnBtgYLUaRw/+EdaufkLLCz6lJsOQQbvUXbm8G/z7Ma2ag==

6 Regime Shifts

In the previous sections, we assume that the order-size distribution does not change over time. In reality, periods with negative order imbalance may be followed by periods with positive order imbalance, and the agent may be able to time these periods and adjust the inventory management policy accordingly.

Suppose that the model parameters $(\mu, \sigma, \lambda, \eta)$ follow a Markov chain with K states $k = 1, \cdots, K$. Denote the transition probabilities of this chain by $\pi_{k_1,k_2},\ k_1, k_2 = 1, \cdots, K$. Let also $V_k(w, q),\ k = 1, \cdots, K$ denote the agent's value function in the state k. Then, standard arguments imply that these value functions satisfy the following system of HJB equations:

$$\sup_c \Bigg\{ -e^{-\gamma c} + (rw - c + q\mu_k)V_{k,w} + 0.5\sigma_k^2 q^2 V_{k,ww} - \beta V_k \\ + \sum_{l=1}^{K} \pi_{k,l}(V_l - V_k) + \lambda \int (V_k(w + \alpha|x|, q + x) - V_k(w, q))d\eta_k(x) \Bigg\} = 0\,.$$

Making the Ansatz

$$V(w, q) = -e^{-r\gamma w + a_k(q)}$$

and substituting it into the HJB equation, we get that the optimal consumption rate in the state k is given by

$$c = -\frac{\log r}{\gamma} + rw - \frac{a_k(q)}{\gamma}\,,$$

whereas the certainty equivalent $a_k(q)$ satisfies the integral equation

$$\begin{aligned} &-r + r(\log r + a_k(q)) + r\gamma q\mu_k - 0.5\sigma_k^2 q^2 (r\gamma)^2 + \beta + \\ &+ \sum_{l=1}^{K} \pi_{k,l}(1 - e^{a_l(q)-a_k(q)}) + \lambda_k \int_{\mathbb{R}} (-e^{-\psi|x|+a_k(q+x)-a_k(q)} + 1) d\eta_k(x) \;=\; 0\,. \end{aligned} \tag{14}$$

In general, this system may be difficult to solve even when $\lambda = 0$. However, for the practically relevant case of a large λ, solving (14) actually reduces to solving the equation separately in each regime if the order-arrival intensity is significantly higher than the intensity of regime shifts. Indeed, in this case, the effect of potential regime shifts is negligible relative to the effect of inventory jumps. We formalize this observation in the following proposition.

Proposition 14 *There is always a unique, convex solution to* (14)*. Furthermore, if* $\min_k \lambda_k$ *is sufficiently large and* $\max_{k,l} \pi_{k,l}\lambda_k^{-1}$ *is sufficiently small, then,* $a_k(q)$ *satisfies* $a(y\lambda_k)/\lambda_k = A_k(y) + O(1/\lambda)$*; and* $q_{k,H,L}/\lambda_k = y_{k,H,L} + O(1/\lambda)$ *where* $A_k(y)$ *solves the system* (11)*,* (12) *with the regime-specific parameters.*

We will use Proposition 14 in the next section to calibrate our model to empirical data.

7 Calibration

In this section, we calibrate our model to data provided by the Swissquote bank.[11] The data consists of the trades of all Swissquote clients for the spot exchange rate USD/NOK for the period Feb 2015 to Aug 2015. Bid-ask spreads are computed as an average across all trades, which gives $\alpha_l = 1.7 * 10^{-3}$ and $\alpha = 2.4 * 10^{-3}$. Annualized spot volatility σ is estimated from historical data, which gives $\sigma = 1.02$. Other relevant parameters are $\gamma = 10^{-10}$, $\mu = 0$, $r = 0.1$.

[11] Swissquote is one of largest Forex dealers in the world.

In order to apply Proposition 14 in this setup, we assume that there are only two regimes: one with a positive imbalance, and one with a negative imbalance. We aggregate trades within each hour, giving us an annualized rate of $\lambda = 24 * 252 = 6048$.[12]

To calibrate this simple two-regime model, we separately pool client order flow for weeks with a positive (respectively, negative) mean order size. Figures 7 and 8 report the empirical densities and cumulative distribution functions in each of the two regimes. As we can see, the two distributions are almost exact mirror images of each other, confirming the intuition that, over a longer horizon, there is no imbalance. At the same time, there is significant imbalance within each regime, and the distribution of order sizes is non-Gaussian, highly skewed, and leptokurtic, implying that using a Gaussian distribution may significantly under-estimate the inventory risk that the agent is facing. In this section, we exploit the ability of our model to deal with arbitrary order-size distributions in a non-parametric fashion; we can use the empirical distribution directly to calculate the corresponding optimal no-trade region.

Figures 9, 10 show the certainty equivalent $A(y)$ together with the corresponding thresholds in the two regimes. As we can see, the utility gains from optimally accounting for the imbalance are highly significant. Figures 9, 10 also show the thresholds $\bar{y}_{H,L,\text{Gaussian}}$ that are calculated from Proposition 14, assuming that order sizes are normally distributed with the same mean and variance as the empirical distribution. As we can see, the impact of non-Gaussianity of the distribution on the optimal policy is quite small. As Figure 3 illustrates, this is an artifact of the structure of the empirical distribution. With enough non-Gaussianity, the differences between the optimal policies based on Gaussian and non-Gaussian distributions with same means and variances can be quite significant.

[12] We remove all hourly trades of size larger than 50000 (in absolute value) to avoid dealing with outliers.

8 Conclusions

Inventory management problems make up an important part of everyday market-making activities and are crucial for market liquidity: For example, the post-crisis drop in liquidity for government and corporate bonds is often attributed to the unwillingness of key markers to hold bond inventories on their balance sheets. In order to efficiently manage inventory, a market-maker should take into account not only the risk-return characteristics of the underlying assets, but also his expectations for future inventory shocks. In this paper, we develop a tractable inventory-management model that allows us to explicitly link the optimal inventory-management policy to the distribution of the size of anticipated inventory shocks. We show that, in the presence of proportional transaction costs, it is optimal for the agent to keep the inventory within a band around the efficient inventory level, and we characterize the boundaries of this band (the no-trade region) explicitly when the order arrival intensity is large and link it to the properties of the moment-generating function of the order-size distribution. We show that, in the presence of inventory shocks, the no-trade region is always wider and tilted towards the order imbalance. The flexibility of our solution allows us to calibrate our model using the actual inventory data of a bank. Our findings suggest that optimally accounting for future inventory shocks leads to significant utility gains.

While our analysis is conducted in a partial equilibrium setting, our technique is very well suited for application in general equilibrium models of over-the-counter markets, such as Duffie, Garleanu, and Pedersen (2007). Using our techniques to develop a general equilibrium model to endogenize spreads, markups, and order-size dynamics is an important direction for future research.

A References

Akian, M., J. L. Menaldi, and A. Sulem, 1996, On an Investment-Consumption Model with Transaction Costs, SIAM Journal of Control and Optimization, 34, 329-364.

Amihud, Y., and H. Mendelson, 1980, Dealership Market: Market Making with Inventory, Journal of Financial Economics 8, 31-53.

Atkeson, G., A. Eisfeldt, and P.-O. Weill, 2015, Entry and Exit in OTC Derivatives Markets, forthcoming in Econometrica.

Bichuch, M., and S. Shreve, 2013, Utility Maximization Trading Two Futures with Transaction Costs, SIAM J. Financial Math 4, 26-85.

Cao, H., M. Evans, and R. Lyons, 2006, Inventory information, Journal of Business 79, 325-363.

Chordia, T., R. Roll, and A. Subrahmanyam, 2001, Market Liquidity and Trading Activity, The Journal of Finance, 56, 501-530.

Comerton-Forde, C., Hendershott, T., Jones, C., Moulton, P., and M. Seasholes, 2008, Time variation in liquidity: The role of market maker inventories and liquidity, Journal of Finance, forthcoming.

Constantinides, G. M. 1986, Capital Market Equilibrium with Transactions Costs, Journal of Political Economy, 94, 842 862.

Dai, M., H. Jin, and H. Liu, 2011, Illiquidity, Position limits, and Optimal Investment for Mutual Funds, Journal of Economic Theory, 146, 1598-1630.

Davis, M. H. A., and A. R. Norman, 1990, Portfolio Selection with Transaction Costs, Mathematics of Operations Research 15, 676 -713.

Dixit, A., 1989, Entry and Exit Decisions under Uncertainty, Journal of Political Economy, 97, 620-638.

Dixit, A., 1991, Analytical Approximations in Models of Hysteresis, Review of Economic Studies, 58, No. 1, 141-151.

Duffie, D., N. Garleanu, and L. H. Pedersen, 2007, Valuation in Over-the-Counter Markets, Review of Financial Studies 20(5), 1865-1900.

Dumas, B. and E. Luciano, 1991, An Exact Solution to a Portfolio Choice Problem under Transactions Costs, The Journal of Finance, 46, 575-595.

Dumas, B., F. Delgado, and G. W. Puopolo, 2015, Hysteresis Bands on Returns, Holding Period and Transaction Costs, Journal of Banking and Finance 57, 86-100.

Garleanu, N., Pedersen, L. and A. Poteshman, 2008, Demand-based option pricing, Review of Financial Studies, forthcoming.

Garleanu, N., and L.H. Pedersen, 2013, Dynamic Trading with Predictable Returns and Transaction costs, The Journal of Finance, 68, 2309-2340.

Garman, M., 1976, Market microstructure, Journal of Financial Economics 3, 257- 275.

Grinold, R., 2006, A Dynamic Model of Portfolio Management, Journal of Investment Management, 4, 5-22.

Guasoni, P., S. Gerhold, J. Muhle-Karbe, and W. Schachermayer, 2014, Transaction Costs, Trading Volume, and the Liquidity Premium, Finance and Stochastics 18(1), 1-37.

Hansch, O., N. Naik, and S. Viswanathan, 1998, Do inventories matter in dealership markets? Evidence from the London Stock Exchange, Journal of Finance 53, 1623- 1655.

Hasbrouck, J., and G. Sofianos, 1993, The trades of market-makers: An analysis of NYSE specialists, Journal of Finance 48, 1565-1594.

Ho, T., and H. R. Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, Journal of Financial Economics, 9(1), 47-73.

Ho,T. and H. R. Stoll, 1983, The dynamics of dealer markets under competition, Journal of Finance 38, 1053-1074.

Green, R.C., B. Hollifield, and N. Schürhoff, 2007a, Dealer Intermediation And Price Behavior In The Aftermarket For New Bond Issues, Journal of Financial Economics 86(3), 643-682.

Green, R.C., B. Hollifield, and N. Schürhoff, 2007b, Financial Intermediation and the Costs of Trading in an Opaque Market, Review of Financial Studies 20(2), 275-314.

Jang, B. G, H. K. Koo, H. Liu, and M. Loewenstein, 2007, Liquidity Premia and Transaction Costs, The Journal of Finance, 42, 2329-2365.

Li, D., and N. Schürhoff, 2014, Dealer Networks, working paper.

Liu, H. 2004, Optimal Consumption and Investment with Transaction Costs and Multiple Risky Assets, The Journal of Finance, 59, 289-338.

Lynch, A. and P. Balduzzi, 2000, Predictability and Transaction costs: the impact on rebalancing rules ad behavior, The Journal of Finance, 55, 2285-2309.

Lynch, A. and S. Tan, 2011, Explaining the Magnitude of Liquidity Premia: the Roles of Return Predictability, Wealth Shocks, and State-Dependent Transaction Costs, The Journal of Finance, 66, 1329-1368.

Lyons, R., 1995, Tests of microstructural hypotheses in the foreign exchange market, Journal of Financial Economics 39, 321-351.

Madhavan, A., and S. Smidt, 1993 , An analysis of daily changes in specialist inventories and quotations, Journal of Finance 48, 1595-1628.

Madhavan, A. and G. Sofianos, 1998, An empirical analysis of NYSE specialist trading, Journal of Financial Economics 48, 189-210.

Manaster, S. and S. Mann, 1996, Life in the pits: Competitive market making and inventory control, Review of Financial Studies 9, 953-975.

Mildenstein, E. and H. Schleef, 1983, The optimal pricing policy of a monopolistic market maker in the equity market, Journal of Finance 38, 218-231.

Naik, N. and P. Yadav, 2003, Do dealer firms manage inventory on a stock-by-stock or a portfolio basis? Journal of Financial Economics 69, 325-353.

Muhle-Karbe, J., and J. Kallsen, 2015, The General Structure of Optimal Investment and Consumption with Small Transaction Costs, forthcoming in Mathematical Finance.

Passerini, F., and S. E. Vazquez, 2015, Optimal Trading with Alpha Predictors, forthcoming in the Journal of Investment Strategies, RISK.

Reiss, P. and I. Werner, 1998, Does risk sharing motivate inter-dealer trading?, Journal of Finance 53, 1657-1704.

Shreve, S.E., H.M. Soner, and G.L. Xu, 1991, Optimal Investment and Consumption with Transaction Costs, Mathematical Finance, 1, 53-84.

Soner, H. M., and N. Touzi, 2013, Homogenization and Asymptotics for Small Transaction Costs, SIAM Journal on Control and Optimization, 51/4, 2893-2921.

Vayanos, D., 2001, Strategic Trading in a Dynamic Noisy Market, Journal of Finance, 56, 131-171.

A Proofs

Proof of Theorem 2. The proof is based on several steps.

Let $X \subset C(\mathbb{R})$ to be the set convex, continuously differentiable functions such that there exists a $Q > 0$ such that $a'(q) = -a'(-q) = \varphi$ for all $q > Q$. Equip this set with a metric

$$\|a - b\|_\infty \;=\; \sup_{\mathbb{R}} |a(x) - b(x)| \,.$$

Importantly, while each function $a, b \in X$ is unbounded, their difference is always constant for sufficiently large x, and hence is bounded, implying that the metric is well defined.

Lemma 15 *The map $\mathcal{I} :\; X_* \to X_*$ defined via*

$$\mathcal{I}(a)(q) \;=\; \inf_{\bar{q}} (a(\bar{q}) + \varphi |q - \bar{q}|)$$

satisfies

$$\|\mathcal{I}(a_1 - a_2)\|_\infty \;\leq\; \|a_1 - a_2\|_\infty$$

for any $a_1, a_2 \in X$.

Proof of Lemma 15. Follows by standard arguments. Q.E.D.

Lemma 16 *Let*

$$\mathcal{Z}(a) \;\equiv\; \log \int e^{-\psi|x| + a(q+x)} d\eta(x)$$

Then, $\mathcal{Z}$ maps X into itself and satisfies

$$\|\mathcal{Z}(a_1) - \mathcal{Z}(a_2)\|_\infty \;\leq\; \|a_1 - a_2\|_\infty$$

and $\mathcal{Z}(a)(q) \leq a(q)$ *for all* $q \in \mathbb{R}$.

Proof. The last claim follows because

$$\log \int e^{-\psi|x|+a(q+x)} d\eta(x) - a(q) \;=\; \int e^{-\psi|x|+a(q+x)-a(q)} d\eta(x) \leq \log \int e^{-(\psi-\varphi)|x|+a(q)-a(q)} d\eta(x) \leq 0\,.$$

Now, we show that $\mathcal{Z}$ maps convex functions into convex. By direct calculation,

$$\frac{d}{dq}\mathcal{Z}(a)(q) \;=\; \frac{\int e^{-\psi|x|+a(q+x)} a'(q+x) d\eta(x)}{\int e^{-\psi|x|+a(q+x)} d\eta(x)}$$

and hence

$$\frac{d^2}{dq^2}\mathcal{Z}(a)(q)$$
$$= \frac{\int e^{-\psi|x|+a(q+x)} (a'(q+x)^2 + a''(q+x)) d\eta(x) \int e^{-\psi|x|+a(q+x)} d\eta(x) - \left(\int e^{-\psi|x|+a(q+x)} a'(q+x) d\eta(x)\right)^2}{\left(\int e^{-\psi|x|+a(q+x)} d\eta(x)\right)^2}$$

and the claim follows from the Cauchy-Schwartz inequality.

Now, by direct calculation, the derivative of $\mathcal{Z}$ is given by

$$(\partial\mathcal{Z})(b(\cdot)) \;=\; \frac{\int e^{-\psi|x|+a(q+x)} b(q+x) d\eta(x)}{\int e^{-\psi|x|+a(q+x)} d\eta(x)}$$

and hence

$$|(\partial\mathcal{Z})(b(\cdot))| \;\leq\; \frac{\int e^{-\psi|x|+a(q+x)} |b(q+x)| d\eta(x)}{\int e^{-\psi|x|+a(q+x)} d\eta(x)} \;\leq\; \sup_x |b(x)|\,.$$

Q.E.D.

Lemma 17 *Let* $G(z,q)$ *be the unique solution to*

$$G \;+\; r^{-1}\lambda \;-\; r^{-1}\lambda e^{-G} e^z \;=\; \Phi(q)\,.$$

Then, the function G is monotone increasing in z and is jointly convex in (z,q), and satisfies

$$\left|\frac{\partial G}{\partial z}\right| \;\leq\; \frac{e^{z-g_*}}{1+e^{z-g_*}}$$

where $g_* = \Phi(q_*) - r^{-1}\lambda$. Furthermore, there exists a $K > 0$ such that $G - \Phi(q) \in [-r^{-1}\lambda, K]$ for all $q \in \mathbb{R}$ whenever $z \leq \Phi(q) + K$.

Proof. The concavity claim follows by the implicit function theorem because the function $rG \;+\; \lambda \;-\; \lambda e^{-G}e^{z} \;- r\Phi(q)$ is jointly concave in (G,z,q). The bound for the derivative follows because $G \;=\; \Phi(q) + r^{-1}\lambda e^{-G}e^{z} - r^{-1}\lambda \geq \Phi(q) - r^{-1}\lambda \geq\; \Phi(q_*) - r^{-1}\lambda$ and hence

$$G_z \;=\; \frac{r^{-1}\lambda e^{-G}e^{z}}{1+r^{-1}\lambda e^{-G}e^{z}} \;\leq\; \frac{e^{z-g_*}}{1+e^{z-g_*}}$$

The last claim follows by direct calculation: we just need to select K large enough. Q.E.D.

Lemma 18 *The map $\mathcal{F}:\;\; a(q) \;\to \mathcal{I}\circ G(\mathcal{Z}(a(\cdot)),q)$ is a contraction, and its iterations converge on compact subsets to a unique fixed point, which is the solution to (1).*

Proof. We need to make sure that the boundaries at which $|a'| = \varphi$ stay bounded when we iterate the map $\mathcal{F}$. This follows because $G - z$ stays bounded from below and hence the identity

$$a' \;=\; \frac{\Phi'(q)}{1+r^{-1}\lambda e^{-G}e^{z}} \;+\; z'(q)\frac{r^{-1}\lambda e^{-G}e^{z}}{1+r^{-1}\lambda e^{-G}e^{z}}$$

guarantees that $|a'$ hits φ for a finite value of q. Q.E.D.

Q.E.D.

Proof of Proposition 6. Standard arguments, based on the implicit function theorem

imply that

$$\begin{aligned} a(q) &= \Phi(q) + \lambda\Phi_1(q) + O(\lambda^2) \\ q_L &= q_L(0) + \lambda q_L(1) + O(\lambda^2) \\ q_H &= q_H(0) + \lambda q_H(1) + O(\lambda^2)\,. \end{aligned}$$

Let

$$\begin{aligned} I(q, q_L, q_H, a(q)) = 1 &- \int_{q_H - q}^{\infty} e^{-\psi x + a(q_H) + \varphi(q + x - q_H) - a(q)} d\eta(x) - \int_{q_L - q}^{q_H - q} e^{-\psi|x| + a(q+x) - a(q)} d\eta(x) \\ &- \int_{-\infty}^{q_L - q} e^{\psi x + a(q_L) + \varphi(q_L - q - x) - a(q)} d\eta(x)\,. \end{aligned}$$

Then, the HJB equation can be written as

$$r(a(q) - \Phi(q)) + \lambda I(q, q_L, q_H, a(q)) = 0$$

and it follows that the term $\Phi_1(q)$ is given by

$$\Phi_1(q) = -\frac{1}{r}\, I(q, q_L(0), q_H(0), \Phi(q))\,.$$

The boundary conditions are given by

$$\begin{aligned} \Phi'(q_L(0) + \lambda q_L(1) + \ldots) + \lambda\Phi_1'(q_L(0) + \lambda q_L(1) + \ldots) &= -\varphi \\ \Phi'(q_H(0) + \lambda q_H(1) + \ldots) + \lambda\Phi_1'(q_H(0) + \lambda q_H(1) + \ldots) &= \varphi\,, \end{aligned}$$

and we arrive at

$$\begin{aligned} q_L(1) &= -\frac{\Phi_1'(q_L(0))}{\Phi''(q_L(0))} = \frac{1}{\sigma^2\gamma^2 r^2}\, \frac{\partial}{\partial q} I(q, q_L(0), q_H(0), \Phi(q))|_{q=q_L(0)} \\ q_H(1) &= -\frac{\Phi_1'(q_H(0))}{\Phi''(q_H(0))} = \frac{1}{\sigma^2\gamma^2 r^2}\, \frac{\partial}{\partial q} I(q, q_L(0), q_H(0), \Phi(q))|_{q=q_H(0)}\,. \end{aligned}$$

Q.E.D.

B Large λ

Proof of Theorem 10. We have

$$a(q) \;=\; \Phi(q) + \frac{\lambda}{r}\int_{\mathbb{R}} (e^{-\psi|x|+a(q+x)-a(q)} - 1) d\eta(x)$$

Since $|a'(x)| \leq \varphi$, we get $|a(q+x) - a(q)| \leq \varphi|x|$ and hence

$$|a(q) - \Phi(q) + \frac{\lambda}{r}| \;\leq\; \frac{\lambda}{r}\int_{\mathbb{R}} (e^{-(\psi-\varphi)|x|} - 1) d\eta(x)\,.$$

Thus, the set of functions $A(y) = a(y\lambda)/\lambda,\ \lambda > 0$ is uniformly bounded and has bounded derivatives $|A'(y)| \leq \varphi$. Hence, by the Arzela-Ascoli theorem, this set is compact and any sequence has a convergent subsequence on any compact interval. The Lebesgue dominated convergence theorem implies that any such convergent subsequence converges to a solution to the ODE

$$\Psi(A'(y)) \;=\; r(A - \bar{\Phi})\,.$$

Let us now show that $y_L,\ y_H$ remain bounded in the limit. Suppose the contrary, and assume without loss of generality, that $y_H \to \infty$. By Proposition 3, the lower bound of the interval on which the ODE holds is below $\bar{y}_L(0)$. Differentiating the ODE, we get the ODE $B' \;=\; \frac{r(B-\bar{\Phi}')}{\Psi'(B(y))}$ for $B = A'$ with $A'(\bar{y}_L(0)) \geq -\varphi$. Then, it suffices to show that B reaches the level of φ for a finite level of y. This is a direct consequence of the properties of the ODE solution.

Thus, we know that for any sequence of λ going to infinity, there is a subsequence along

which the triple $(a(\lambda y)/\lambda, q_H/\lambda, q_L/\lambda)$ converges to a triple function $(A, \bar{y}_H, \bar{y}_L)$ satisfying

$$\Psi(A'(y)) \;=\; r(A(y) - \bar{\Phi}(y)),\;\; A'(\bar{y}_H) = -A'(\bar{y}_L) = \varphi$$

and $\bar{y}_H \geq y_H(0) > y_L(0) \geq \bar{y}_L(0)$. In addition, $|A'(y)| \leq \varphi$ needs to hold for all $y \in [y_L, y_H]$. A necessary condition is that $A''(y) \geq 0$ for $y \to \bar{y}_L, \bar{y}_H$. We have

$$A''(\bar{y}_H) \;=\; \frac{r(A'(\bar{y}_H) - \bar{\Phi}'(\bar{y}_H))}{\Psi'(A'(\bar{y}_H))} \;=\; \frac{r(\varphi - \bar{\Phi}'(\bar{y}_H))}{\Psi'(\varphi)}\,.$$

Assuming that $\Psi'(\varphi) > 0$, we get that we need $\bar{\Phi}'(\bar{y}_H) \leq \varphi$. But this is equivalent to $\bar{y}_H \leq y_H(0)$, which is only possible if $\bar{y}_H = y_H(0)$. Thus, we arrive at the following result.

Lemma 19 *If $\Psi'(\varphi) > 0$, then, $\bar{y}_H = y_H(0)$. If $\Psi'(-\varphi) < 0$ then $\bar{y}_L = y_L(0)$.*

Lemma 20 *The function Ψ is log-convex in the sense that $(\log \Psi)'' \geq 0$. In particular,*

- *if η is supported on $\mathbb{R}_+$ then Ψ is monotone increasing;*
- *if η is supported on $\mathbb{R}_-$ then Ψ is monotone decreasing;*
- *if η is supported on both half-lines, then there is a unique solution z_* to $\Psi'(z_*) = 0$ and Ψ is decreasing for $z < z_*$ and increasing afterwards.*

Let first $\Psi'(\varphi) > 0 > \Psi'(-\varphi)$. Then, by Lemma 19, $\bar{y}_H = y_H(0),\; \bar{y}_L = y_L(0)$. Since A is convex and hence continuous, and hence $A(y) = \mathbf{1}_{y \leq \tilde{y}} A_-(y; y_L(0) + \mathbf{1}_{y \leq \tilde{y}} A_+(y; y_H(0))$.

Let now $\Psi'(\varphi) > \Psi'(-\varphi) > 0$. Define $B(y) \equiv A'(y)$. Then, $B'(y) = \frac{r(B(y) - \bar{\Phi}'(y))}{\Psi'(B(y))}$. At the same time $\bar{\Phi}''(y) \geq 0 = \frac{r(\bar{\Phi}'(y) - \bar{\Phi}'(y))}{\Psi'(\bar{\Phi}'(y))}$. Since $\bar{\Phi}'(y_H(0)) = B(y_H) = \varphi$, the comparison theorem for ODEs implies that $B(y) \geq \bar{\Phi}'(y)$.

Q.E.D.

Proof of Proposition 12. Without loss of generality, we can set $\mu = 0$. This can be achieved by simply shifting the variable y. We only consider the case when $0 \geq \Psi'(\varphi) > \Psi'(-\varphi)$. The case when $\Psi'(\varphi) > \Psi'(-\varphi) \geq 0$ is completely analogous.

Proof of monotonicity in φ. Let $\varphi_1 > \varphi_2$. By continuity, it suffices to consider the case when both φ_1, φ_2 satisfy $0 \geq \Psi'(\varphi) > \Psi'(-\varphi)$. Let A_1 and A_2 be the corresponding solutions. Let also $y_i = -\frac{\varphi_i}{2\bar{\rho}}$ be the corresponding lower boundaries and $\bar{y}_1, \bar{y}_2$ the corresponding upper boundaries. Since the value function is monotone decreasing in φ, we get that $A_1(y) > A_2(y)$ for all y. Since Ψ is monotone decreasing on the range of values taken by the derivatives A_1', A_2'. Since they satisfy the same ODE, we get that

$$\Psi(A_1'(y)) \;=\; r(A_1(y) - \bar{\Phi}(y)) \;\geq r(A_2(y) - \bar{\Phi}(y)) = \Psi(A_2'(y))$$

implies $A_1'(y) \leq A_2'(y)$ for all y. In particular, $\varphi = A_2'(\bar{y}_2) > A_1'(\bar{y}_2)$, which means that $\bar{y}_1 > \bar{y}_2$.

Proof of asymptotic as $\varphi \to 0$.

Differentiating the ODE, we get $A''(y_H(0)) = 0$ whereas

$$A'''(y) \;=\; \frac{r(A'' - \bar{\Phi}'') - \Psi''(A')(A'')^2}{\Psi'(A')}$$

and hence

$$A'''(y_H(0)) \;=\; \frac{-2r\bar{\rho}}{\Psi'(\varphi)}$$

so that

$$A'(y_L) \;\approx\; \varphi + 0.5(y_L - y_H(0))^2 \frac{-2r\bar{\rho}}{\Psi'(\varphi)}$$

Solving $A'(y_L) = -\varphi$ gives

$$y_L \;\approx\; -2\varphi^{1/2}\left(\frac{r\bar{\rho}}{\Psi'(0)}\right)^{-1/2} + y_H(0) + O(\varphi) \;=\; -2\varphi^{1/2}\left(\frac{r\bar{\rho}}{\Psi'(0)}\right)^{-1/2} + \frac{\gamma\mu}{2\bar{\rho}} + O(\varphi)$$

Q.E.D.

Proof of Corollary 13. We only consider the case of $\bar{y}_H$. By the comparison theorem for ODEs, it suffices to show that $\Psi'(z)$ is monotone increasing in ψ for all $z \in (-\varphi, \varphi)$. Indeed, in this case we have

$$A''(y) \;=\; \frac{r(\Phi'(y) - A'(y)}{-\Psi'(A'(y))}$$

is increasing. We have

$$\frac{\partial}{\partial\psi}\Psi'(z) \;=\; -\int_{\mathbb{R}} e^{-\psi|x|}|x|xe^{xz}d\eta(x)$$

and

$$\frac{\partial^2}{\partial z\partial\psi}\Psi'(z) \;=\; -\int_{\mathbb{R}} e^{-\psi|x|}|x|x^2e^{xz}d\eta(x)\,.$$

Thus, $\frac{\partial}{\partial\psi}\Psi'(z)$ is decreasing in z, and therefore it is positive for all z it is for $z = \varphi$ and the claim follows. Q.E.D.

The following scaling rules hold true.

Lemma 21 *Let $\varepsilon > 0$.*

- *If we make a transformation $r \to r\varepsilon,\ \lambda\sigma^2\gamma^2 \to \lambda\sigma^2\gamma^2/\varepsilon^2,\ \gamma\mu \to \gamma\mu/\varepsilon,\ \gamma\alpha \to \gamma\alpha/\varepsilon^2$ and multiplying all inventory shocks by ε, then $\bar{y}_H$ stays invariant.*

- *keep r fixed and change other parameters so that* $\lambda\sigma^2\gamma^2 \to \lambda\sigma^2\gamma^2/\varepsilon^2$, $\gamma\mu \to \gamma\mu/\varepsilon$, $\gamma\alpha \to \gamma\alpha/\varepsilon$ *and multiply all inventory shocks by* ε. *Then,* $\bar{y}_H$ *becomes* $\bar{y}_H\varepsilon$.

Proof. For the first item, it is straightforward to verify that if $A(y)$ solves the original ODE then $A(y)/\varepsilon$ solves the ODE for the new parameter values. For the second case, it is straightforward to verify that $B(y) = A(y/\varepsilon)$ solves the new ODE. Q.E.D.

Lemma 22 *Suppose that* $\Phi'(\varphi) < 0$. *Then,* $\bar{y}_H \ \leq \ \frac{\mu+2\bar{\rho}/\hat{r}}{2\bar{\rho}}$.

Proof. We have

$$B'(y) \ = \ \hat{r}\frac{\bar{\Phi}' - B}{-\Psi'(B(y))} \ \geq \ \hat{r}\frac{\bar{\Phi}' - B}{-\Psi'(-\varphi)}$$

because Ψ' is increasing. Define $\hat{r} \ = \ -r/\Psi'(-\varphi)$. Then, by the comparison theorem for ODEs,

$$B(y) \ \geq \ 2\bar{\rho}y - \mu - 2\bar{\rho}/\hat{r}(1 - e^{-\hat{r}(y-y_L(0))}) \ \geq \ 2\bar{\rho}y - \mu - 2\bar{\rho}/\hat{r}$$

implying that $\bar{y}_H \ \leq \ \frac{\mu+2\bar{\rho}/\hat{r}}{2\bar{\rho}}$. Assuming μ is zero, we get that

$$\bar{y}_H/y_H(0) \ \leq \ \frac{2\bar{\rho}(-\Psi'(-\varphi))}{r\varphi}\,.$$

Q.E.D.

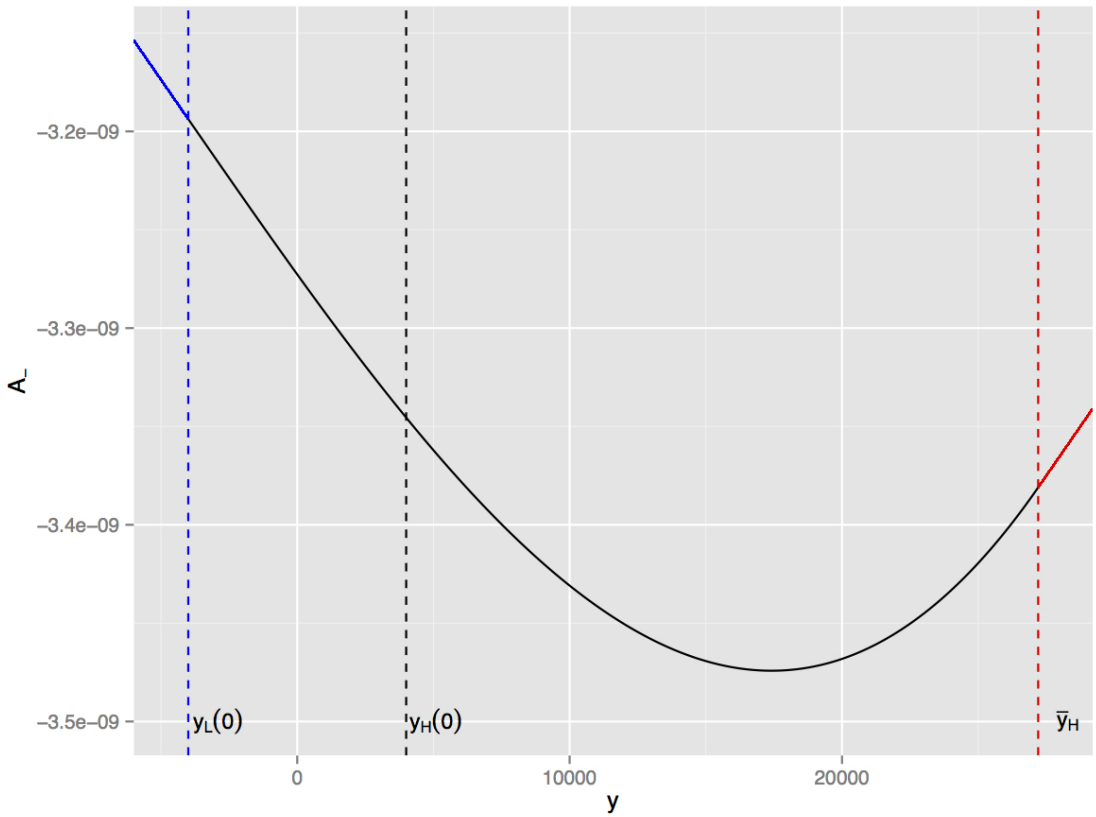

Figure 1: This figure plots A_{-}. The market parameters are set to the following values: $\mu = 0$, $\sigma = 1$, $r = 0.1$, $\alpha_l = 2.0 * 10^{-3}$ and $\alpha = 2.5 * 10^{-3}$. We also set $\lambda = 5 * 10^3$, $\gamma = 10^{-10}$. The order distribution is gaussian with standard deviation $\sigma_I = 20 * 10^3$ and mean $\mu_I = -5 * 10^3$.

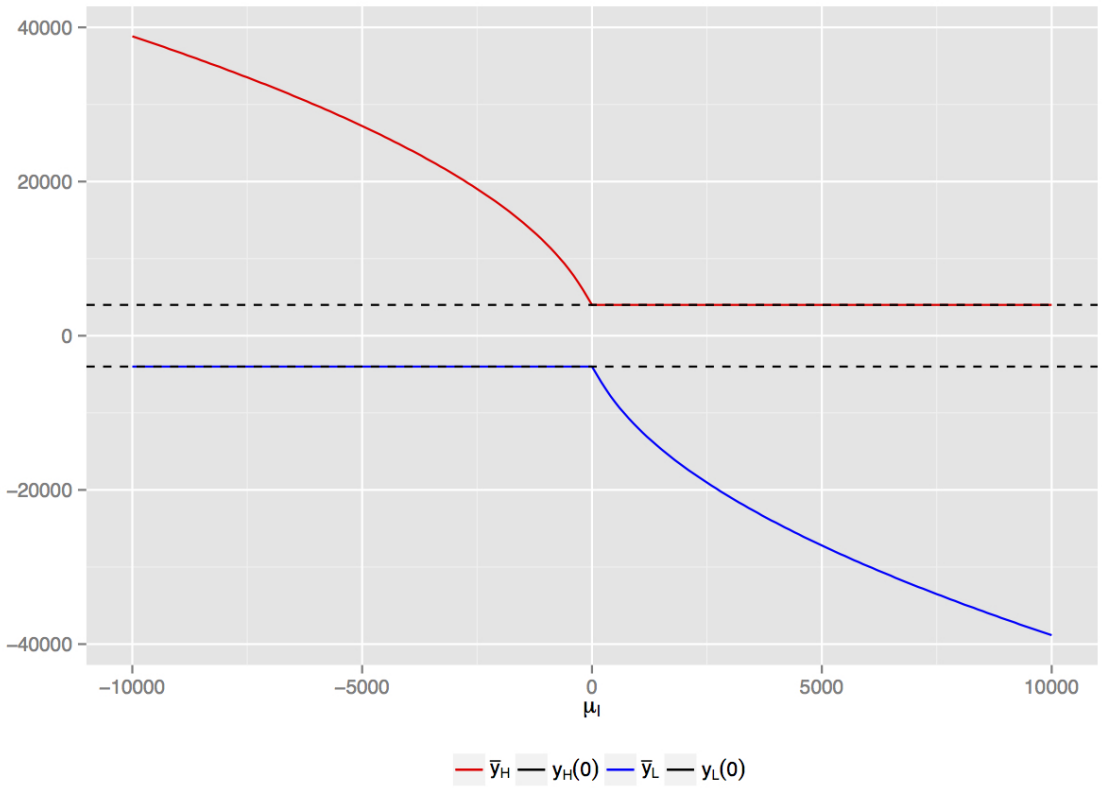

Figure 2: This figure plots the thresholds of the no-trade region (solid lines) as a function of μ_I, the mean of the gaussian client-order distribution. The market parameters are set to the following values: $\mu = 0$, $\sigma = 1$, $r = 0.1$, $\alpha_l = 2.0 * 10^{-3}$ and $\alpha = 2.5 * 10^{-3}$. We also set $\lambda = 5 * 10^3$, $\gamma = 10^{-10}$. The standard deviation of the gaussian order distribution is $\sigma_I = 20 * 10^3$.

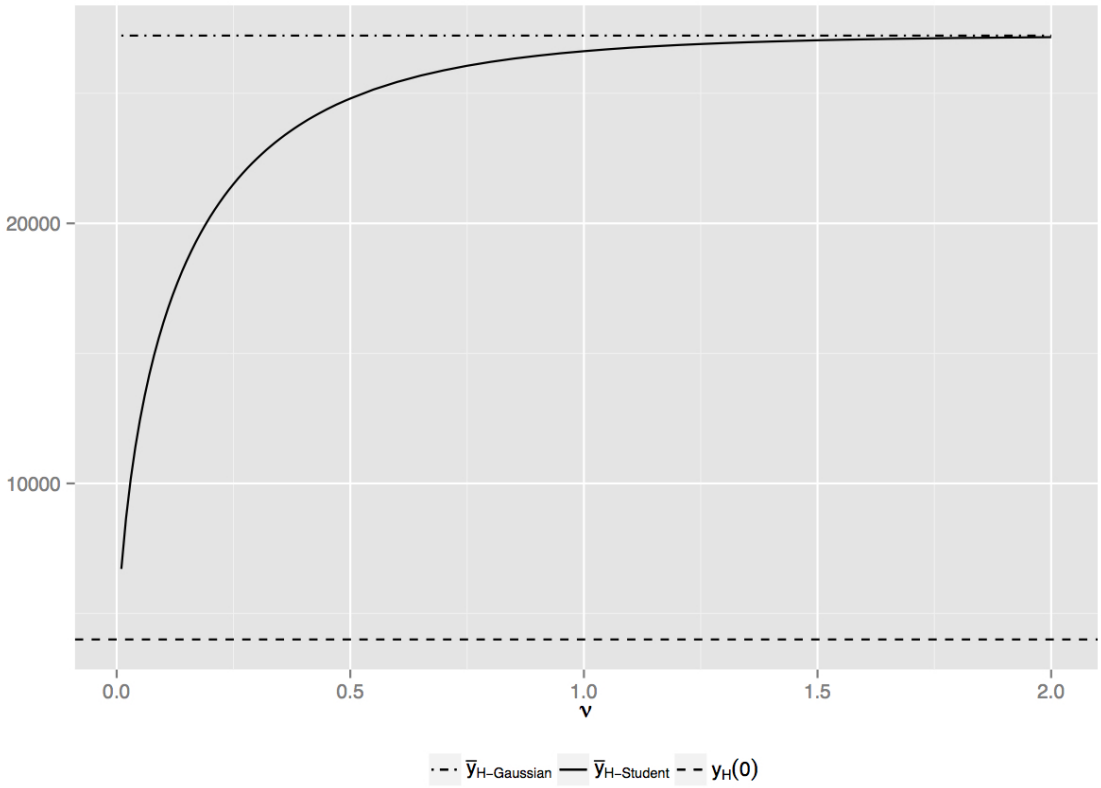

Figure 3: This figure plots the thresholds of the no-trade region (solid lines) as a function of ν, the number of degrees of freedom of the student-t client-order distribution. The market parameters are set to the following values: $\mu = 0$, $\sigma = 1$, $r = 0.1$, $\alpha_l = 2.0 * 10^{-3}$ and $\alpha = 2.5 * 10^{-3}$. We also set $\lambda = 5 * 10^3$, $\gamma = 10^{-10}$. To make a meaningful comparison with Figure 1, we define the order size distribution to be $\eta(x) = \sigma_I^{-1} f((x - \mu_I)/\sigma_I, \nu)$ with $f(x, \nu) = C(\nu)(1 + \nu^{-1}x^2)^{-\frac{\nu+1}{2}}$ is the student$-t$ density with ν degrees of freedom and $C(\nu)$ is a normalization constant. We set $\sigma_I = 20 * 10^3$ and $\mu_I = -5 * 10^3$, as in Figure 1.

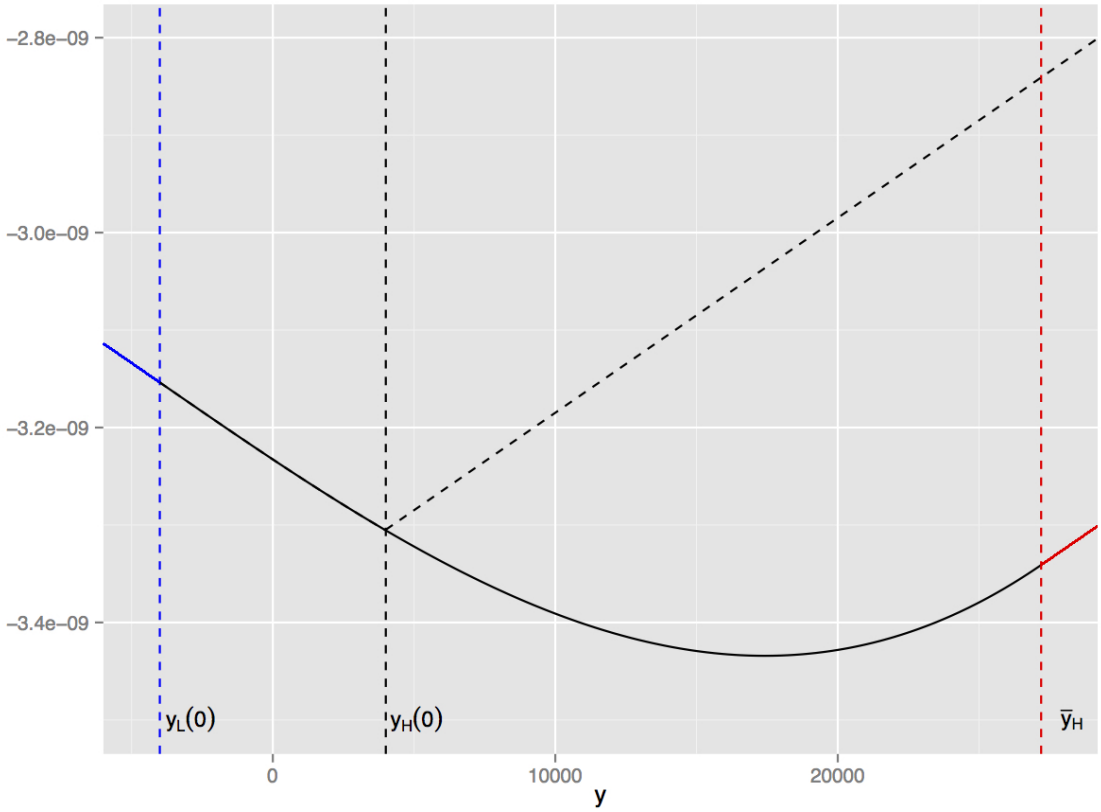

Figure 4: This figure plots A_{-} and A_{naive}. Parameters are set as in figure 1.

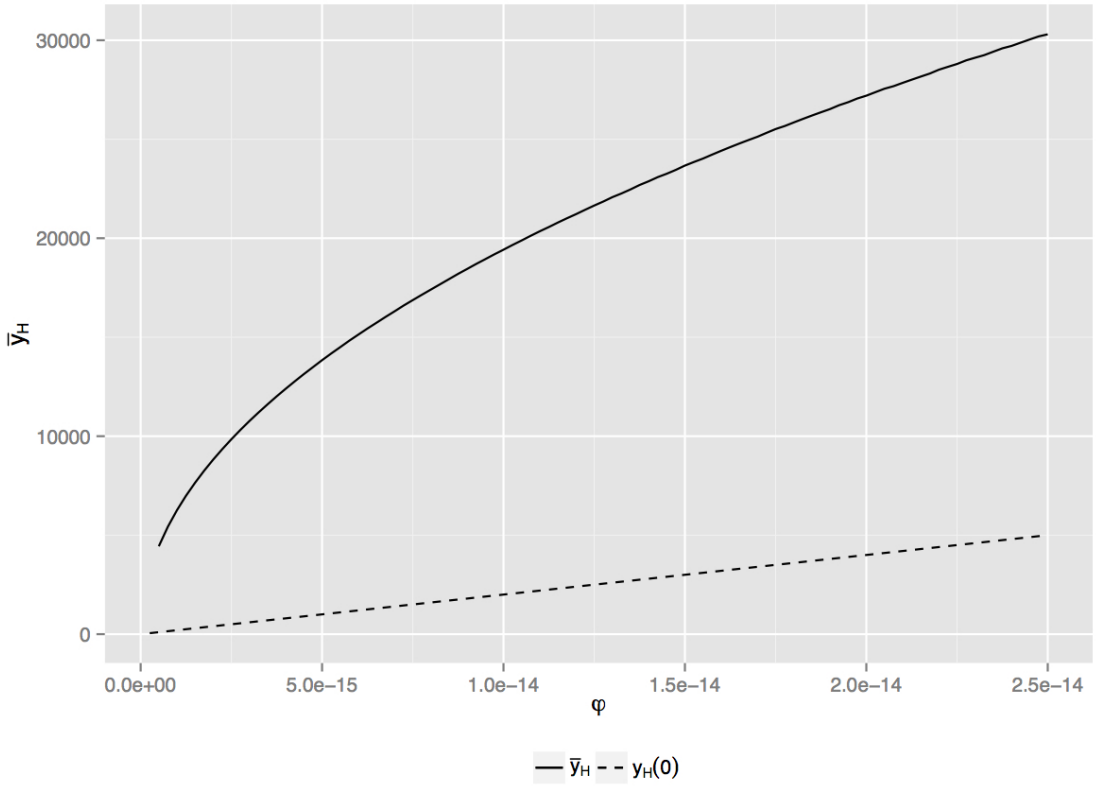

Figure 5: This figure plots $\bar{y}_H$ as a function of φ. The market parameters are set to the following values: $\mu = 0$, $\sigma = 1$, $r = 0.1$ and $\alpha = 2.5 * 10^{-3}$. We also set $\lambda = 5 * 10^3$, $\gamma = 10^{-10}$. The order distribution is gaussian with standard deviation $\sigma_I = 20 * 10^3$ and mean $\mu_I = -5 * 10^3$.

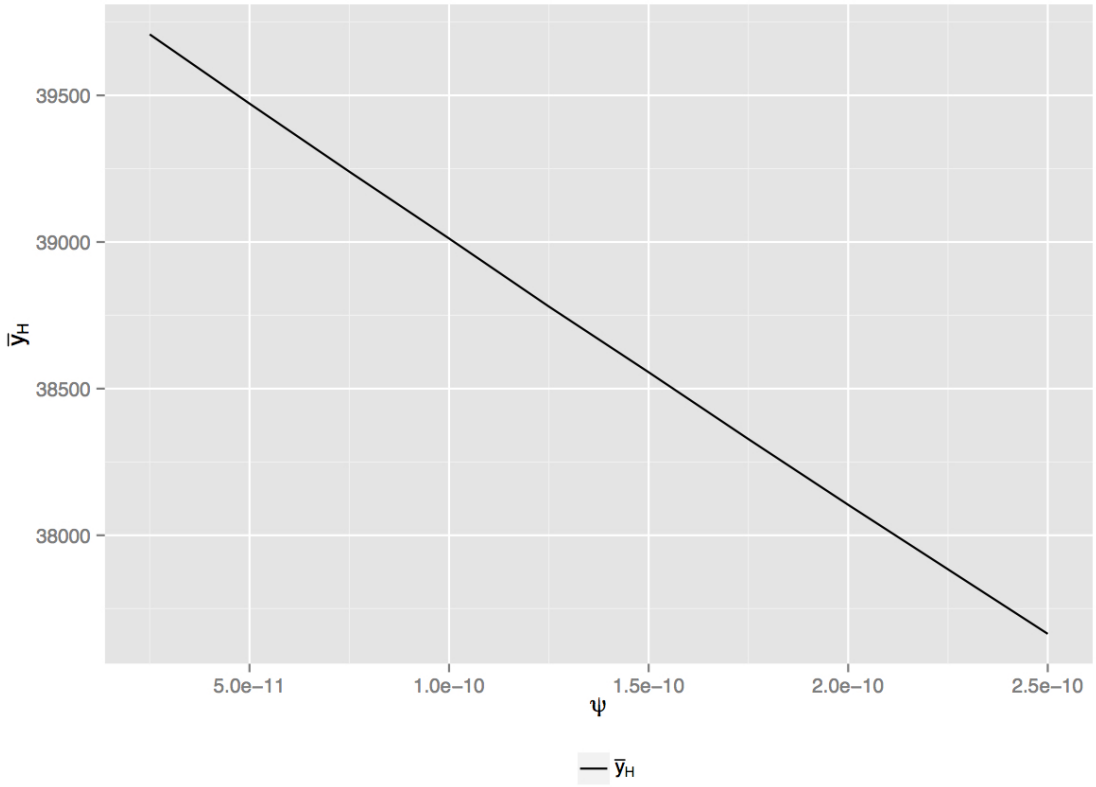

Figure 6: This figure plots $\bar{y}_H$ as a function of ψ. The market parameters are set to the following values: $\mu = 0$, $\sigma = 1$, $r = 0.1$, $\alpha_l = 2.0 * 10^{-3}$. We also set $\lambda = 5 * 10^3$, $\gamma = 10^{-7}$. The order distribution is gaussian with standard deviation $\sigma_I = 30 * 10^7$ and mean $\mu_I = -10 * 10^6$.

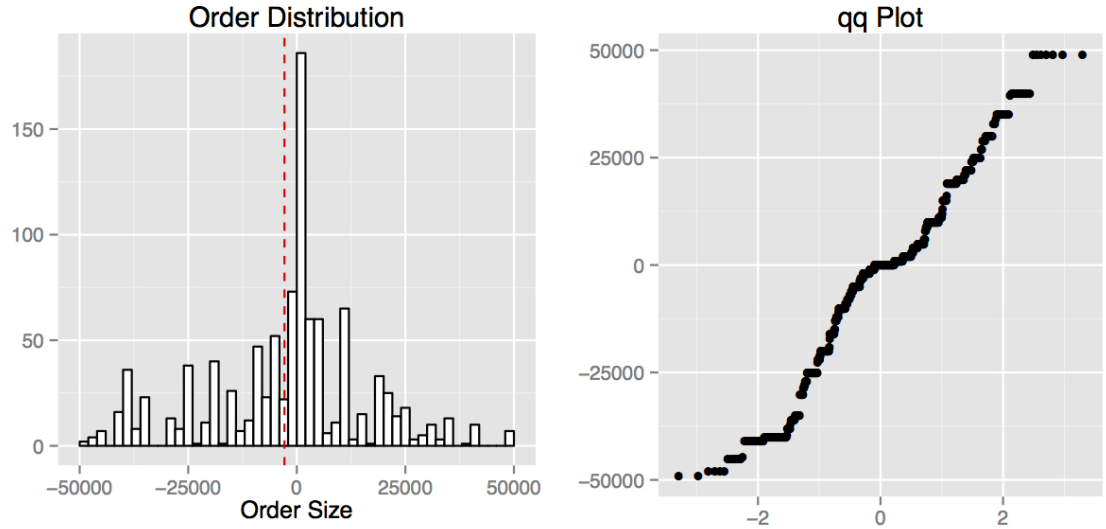

Figure 7: Distribution for client orders with negative imbalance. These are orders that produce a negative aggregated weekly order. The mean is -2843 and the standard deviation is 18466.

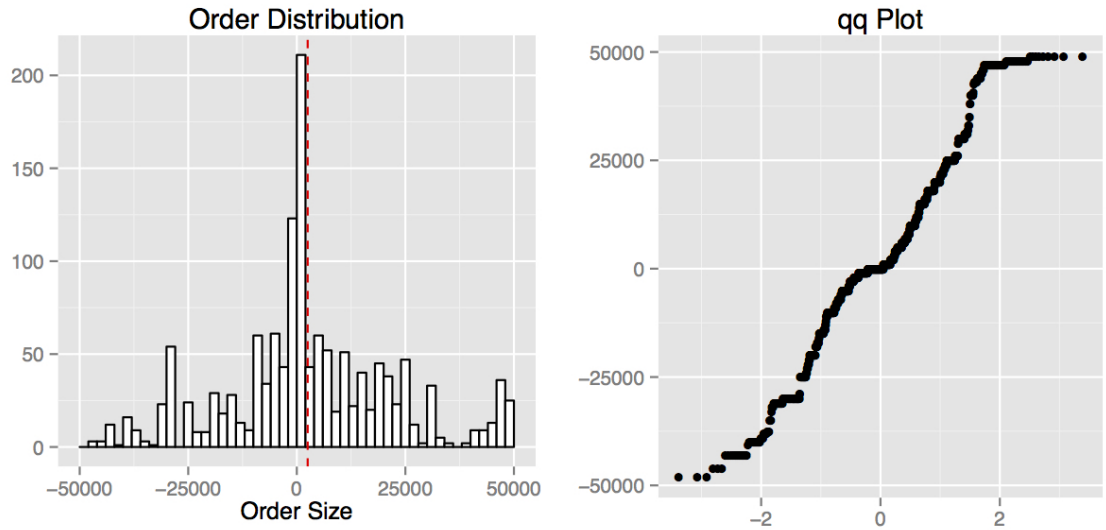

Figure 8: Distribution for client orders with positive imbalance. These are orders that produce a positive aggregated weekly order. The mean is 2505 and the standard deviation is 19734.

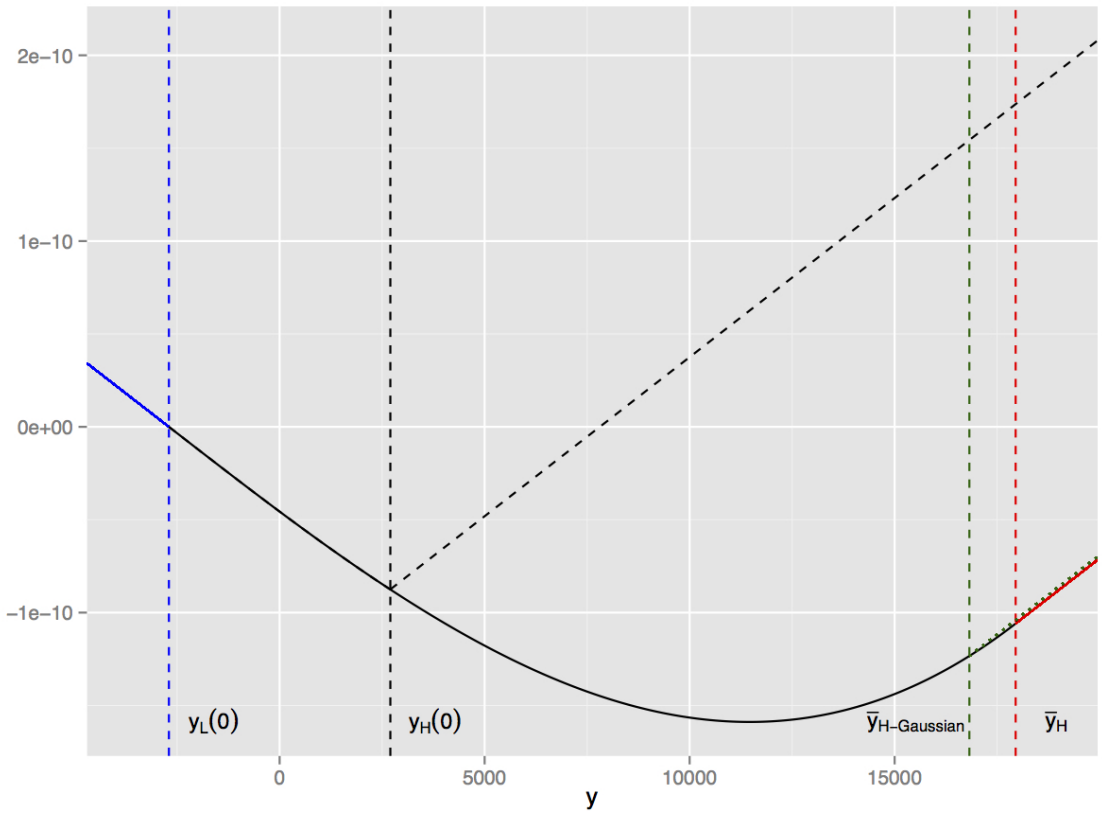

Figure 9: Utility functions for the case of orders with negative imbalance. $y_H(0) = -y_H(0) = 2700$, $\bar{y}_{H-\text{Gaussian}} = 16822$, $\bar{y}_H = 17951$.

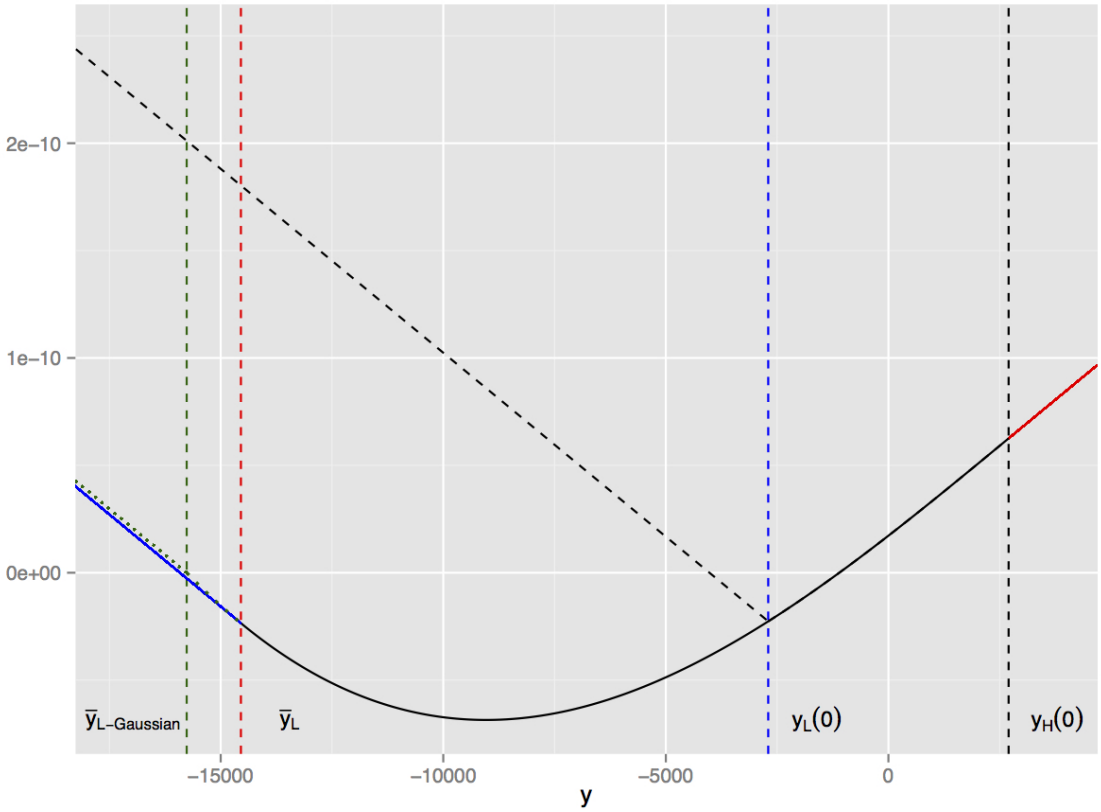

Figure 10: Utility functions for the case of orders with positive imbalance. $y_H(0) = -y_H(0) = 2700$, $\bar{y}_{L-\text{Gaussian}} = -15764$, $\bar{y}_L = -14546$.