Price Pressures[1]

Terrence Hendershott and Albert J. Menkveld

June 29, 2009

(for slides click [here])

(for supplementary material click [here])

[1]Terrence Hendershott, Haas School of Business, University of California at Berkeley, 545 Student Services Bldg #1900, Berkeley, CA 94720, tel +1 510 643 0619, fax +1 510 643 1412, hender@haas.berkeley.edu. Albert J. Menkveld, VU University Amsterdam, FEWEB, De Boelelaan 1105, 1081 HV, Amsterdam, Netherlands, tel +31 20 598 6130, fax +31 20 598 6020, albertjmenkveld@gmail.com. We benefit from in-depth discussions with Bruno Biais, Sabrina Buti, Robert Engle, Jeremy Large, Sunny Li, Virgiliu Midrigan, Stijn van Nieuwerburgh, Matthew Pritsker, Jan van Schuppen, and Pierre-Olivier Weill. The project was inspired by the Ljungqvist and Sargent NYU lecture series based on their book *Recursive Macroeconomic Theory*. Menkveld gratefully acknowledges Robert Engle for sponsoring his NYU visit in 08/09 and VU University Amsterdam for a VU talent grant.

Price Pressures

Abstract

We study price pressures—price deviations from fundamental values due to a risk-averse intermediary supplying liquidity to asynchronously arriving investors with idiosyncratic hedging values. In our model, the intermediary uses price pressure to mean-revert costly inventory by trading off the size of the price pressure against the cost of remaining in a risky inventory state. Price pressure is associated with the social cost of lower realization of investor hedging value because the intermediary's efforts to mean revert inventory substitutes low-hedge-value investors on the side of the market that reduces the risk of the intermediary's position for high-hedge-value investors on the risk-increasing side. Empirically, twelve years of daily New York Stock Exchange (NYSE) intermediary data reveal economically large price pressures and associated social costs. A $100,000 inventory shock causes an average price pressure of 0.28% and the average transitory volatility in daily stock returns (average price pressure) is 0.49% with substantially larger price pressure effects in smaller stocks. The aggregate lost hedging gains are estimated to be $60 billion for all NYSE common stocks for our sample period.

Keywords: liquidity, inventory risk, intermediary JEL: G12, G14, D61

Studies of price pressure often measure the magnitude of transitory price deviations from fundamental value, commonly referred to as price pressure, due to asynchronously arriving investors with idiosyncratic hedging values demanding substantial immediacy/liquidity from intermediaries supplying liquidity.[1] These provide evidence that non-informational trading frictions and associated price pressures can be large.[2] Price pressure is associated with two sources of welfare loss due to inefficient allocation of risk: the intermediaries' cost of bearing risks and the intermediaries' risk reducing behavior lowering the realized hedge values by investors. By examining the liquidity supplying behavior of intermediaries we can calculate the normal costs for representative investors of trading large quantities for non-informational reasons, the average magnitude of the deviation of prices from fundamental values, and the social costs of price pressure. To characterize the magnitude of price pressure in terms of frequency and size and its social costs we use 12 years of a daily measure of liquidity supplied on the New York Stock Exchange (NYSE).

To study the dynamics of price pressure we construct a single-asset theoretical model of liquidity supply with the reduced-form assumption that investors with less than perfectly elastic demands to trade arrive stochastically. As in Grossman and Miller (1988) natural motivations for asynchronous arrivals are idiosyncratic shocks to agents' consumption or investment opportunity sets which lead to liquidity or hedging needs. If agents are impatient and if it is costly to continuously monitor the market for trading opportunities, the market is not always complete as allocative efficiency via perfect risk sharing is not possible. In this case an intermediary will naturally enter to bear the costs of monitoring the market and stand ready to both buy and sell (Townsend (1978)). To focus our results on price pressure the model has no information asymmetry.

In the infinite-horizon recursive model the intermediary dynamically chooses the prices at which she is willing to buy and sell, the bid and ask prices, respectively. When the intermediary is at her desired position in a security the bid and ask prices symmetrically straddle the security's fundamental value. If a seller then arrives the intermediary buys and her position, also referred to as her inventory level, is higher than desired, exposing her to idiosyncratic price risk. To mitigate this risk the intermediary then stochastically mean reverts her inventory back to zero by attempting

[1] Kraus and Stoll (1972) provide some of the first evidence on liquidity demands from block trades causing price pressure. Harris and Gurel (1986) and subsequent papers on additions to the S&P 500 index find evidence for price pressure. Greenwood (2005) extends this examining transitory price effects upon a weighting change to the Nikkei 225. Coval and Stafford (2007) examine price pressure due to mutual fund redemptions. These are all demand side effects.

[2] Theoretical models examine circumstances that lead to substantial price pressure due to shocks to liquidity supply. Brunnermeier and Pedersen (2008) find that when funding liquidity is low liquidity shocks causes prices to deviate more from fundamentals. Acharya, Shin, and Yorulmazer (2008) examine how crises and financing constraint induces price pressure (fire sale prices) and foreign direct investment even when the foreign firms are inefficient owners.

to sell. She adjusts both the bid and ask prices downward to induce buyers to arrive faster than sellers. In doing this the intermediary skews her quotes downward by setting the average of the bid-ask quotes, the midquote price, to be below the fundamental value. The size of the deviation of the midquote price from the fundamental value is our theoretical and empirical measure of price pressure. As emphasized in Stoll (1978) and Grossman and Miller (1988) this is distinct from the measures of the cost of immediacy based on the width of the bid-ask spread.

The main innovation of our theoretical approach is to facilitate calculation of price pressure's social costs. We refer to investors idiosyncratic desire to trade as their hedging values, although the motivations for trade could be more general. The intermediary's use of price pressure to mean revert her inventory results in lower realization of investor hedging value because the price pressure substitutes for low-hedge-value investors on on the side of the market that reduces the risk of the intermediary's position for high-hedge-value investors on the risk-increasing side. Assuming a specific distribution for investors' hedge value enables the model to quantify the magnitude of the unrealized investor hedging gains due to price pressure as a function of observables in the data. We also examine the measures of price pressure and the associated social costs by numerically solving the Ho and Stoll (1981) model.

Before empirically studying price pressure we use the Kalman filter to estimate a state-space model that decomposes stock prices into their fundamental value and noise: the random walk and stationary components. The stationary component represents pricing error around the fundamental value. One important observation from the theoretical model is that deviations from fundamentals due to intermediation are not immediately reversed, but rather last as long as it takes the intermediary to mean revert her inventory to its desired level. To account for this we allow for the pricing errors to follow an autoregressive process. Because hedging of risk associated with the common market-wide factor is relatively inexpensive we focus on idiosyncratic effects. The estimation also allows for delayed adjustment to the common factor. Throughout the paper we conduct our analysis yearly at the stock level and report averages across stocks by market capitalization quintiles. There is persistence in the pricing errors with first-order autocorrelation coefficients typically between 0.3 and 0.5. The daily volatility of the stationary component of prices is roughly 0.75% as compared to the daily volatility of the random walk component of 1.73%.

To determine how much of this pricing error is attributable to intermediary-related price pressure we employ intermediary data from the New York Stock Exchange from 1994-2005. Market makers who act as intermediaries to supply liquidity on the NYSE, called specialists, are required to report their positions in every security every day.[3] We extend the above estimation of pricing errors to incorporate the NYSE intermediaries' inventory positions by allowing the idiosyncratic inventory level to enter directly into the price equation. The coefficient on inventory represents the conditional price pressure. The standard deviation of the idiosyncratic inventory characterizes the frequency of price pressure. Combining the frequency and conditional price pressure yields the average price pressure which we refer to as the price pressure transitory volatility.

The conditional price pressure for the largest-quintile stocks is 0.02 basis points per one thousand dollars of intermediary inventory. For the smallest-quintile stocks one thousand dollars of inventory results in 1.01 basis points of price pressure. The small conditional price pressure in large stocks increases the frequency of price pressure: the standard deviation of inventory is $1.1 million for the largest stocks versus $165 thousand for the smallest stocks. Combining the frequency and conditional size of price pressure produces estimates of the daily price pressure transitory volatility that range from 0.17% for the largest stocks to 1.20% for smallest stocks. Price pressure contributes substantially to daily volatility in stock prices in small stocks as the average ratio of price pressure transitory volatility to permanent volatility is greater than one.

The social costs associated with price pressure have two components of the inefficient allocation of risk: the intermediary's cost of bearing risks and the intermediary's risk reducing behavior lowering the realized hedge values by investors. While the conditional and unconditional price pressure is larger for small stocks, the greater total potential investor hedging gains in large stocks results in higher social costs from price pressure in the large stocks: unrealized hedge gains for large stocks is $9.86 million per year vs. $0.67 million for small stocks. The average stock's social cost from price pressure is $3.60 million per year. Given that there are roughly 1,500 common stocks traded on the NYSE each year, the total costs for our 12-year sample period is $60 billion.

Prior empirical work on intermediary inventories find support for risk management via inventory control, but weak support for inventories causing price pressure.[4] Hasbrouck (1988) uses trade-signed liquidity demand to identify inventory effects. He finds evidence of negative auto-correlation of trade sign consistent with risk management, but finds no evidence of price pressure. Madhavan and Smidt (1991), Madhavan and Smidt (1993), and Hasbrouck and Sofianos (1993) employ NYSE

[3]Recent technology changes which sped up trading on the NYSE have reduced the role of the specialist in intermediation (Hendershott and Moulton (2007)). Now a more diverse group of traders fulfill the role of temporarily carrying inventory to facilitate trading between asynchronously arriving buyers and sellers. The lack of position data for all these participants makes a broad empirical characterization of price pressure more challenging in the current environment.

[4]Indirect evidence consistent with price pressure exists in the literature on return reversals (Lehmann (1990), Campbell, Grossman, and Wang (1993), and others).

intermediary inventory data and find evidence supporting risk management, but not price pressure.[5] Hendershott and Seasholes (2007) use a long time series of NYSE data and find evidence of both inventory control and price pressure. Their price pressure findings are cross-sectional: a portfolio of stocks where the intermediary is long outperforms a portfolio of stocks where the intermediary is short by 45.4 basis points over two weeks. Our findings extend this portfolio approach to determine the price pressure per dollar of inventory, the average impact of price pressure on stocks' volatility, and the social costs of price pressure.

1 A simple model to characterize price pressure dynamics and its relation to social cost

In this section, we characterize the conditional and average price pressure for a security as a result of the friction of nonsynchronous arrivals of investors who demand immediacy. We model the supply of immediacy with a representative risk-averse intermediary who produces price quotes. If the intermediary's position differs from her desired position her optimal policy sets the midquote (average of the bid and ask quote) on that side of the fundamental value that creates an nonzero average net (signed) liquidity demand to mean-reverts her inventory.[6] In this way she skews the midquote downward if she is long relative to her optimal position and upward if she is short. These transitory price effects manifest themselves as price pressures to liquidity demanders who, when trying to trade into a large positions, push quotes further away from fundamental values.

The intermediary facilitating trade among asynchronous investor arrivals generates inefficient allocation of risk in two ways. First, when the intermediary takes on position she bears risk until the final investor arrives. Second, the intermediary's effort to reduce her risk decrease the amount of hedging value the investors can capture. If the intermediary is risk neutral with unlimited capital then there is no no price pressure and no social costs due to inefficienct allocation of risk.

Social cost of the intermediary bearing risk. The first component of the social cost of inefficient allocation is the intermediary bearing idiosyncratic price risk on her inventory position

[5] Ho and Macris (1984) estimate a model similar to Madhavan and Smidt (1991) using American Stock Exchange intermediary inventory data for stock options. Using data on inventories from London Stock Exchange dealers Hansch, Naik, and Viswanathan (1998), Reiss and Werner (1998), and Naik and Yadav (2003) examine intermediary risk management.

[6] One way to rationalize investor heterogeneity (impatient liquidity demander vs. intermediary) as a general equilibrium outcome is that it is costly to monitor the market (e.g., update quotes to prevent picking-off risk) and it is therefore natural to see some agents specialize who pay the market monitoring cost and stand ready to buy and sell (Townsend (1978)).

between counterbalancing liquidity demand arrivals. In a static model, Grossman and Miller (1988) argue that transitory price effects serve to compensate intermediaries who take on idiosyncratic price risk in matching asynchronous liquidity demand. Intermediaries buy at prices below fundamental value if sellers arrive before buyers or sell at prices above fundamental value if sellers arrive before buyers. We refer to these transitory price effects as price pressures.

Grossman and Miller argue that these transitory price effects are the essence of liquidity. They analyze the size of these effects in an industrial organization context thereby fitting early microstructure models of inventory-management (e.g., Stoll (1978)) "into a larger framework that also encompasses the ultimate demanders and suppliers (p.617)." They compare and contrast the price pressure measure to alternative measures and discuss the limitations of two alternative measures: the bid-ask spread and the liquidity ratio:

1. "... (as Stoll (1978) has emphasized) the bid-ask spread only exactly measures the intermediary's return for providing immediacy if she simultaneously crosses (i.e., executes on both sides of) the trade, one at the bid and the other at the ask (p.628)."

2. "Another widely used empirical measure in inter-market comparison is the liquidity ratio, defined as the ratio of average dollar volume of trading to the average price change during some interval... These measures, of course, tell us at best only about past average associations between price changes and volume. They do not answer the critical question of how the sudden arrival of a larger-than-average order would affect price. Nor do they distinguish adequately among the sources of price volatility. A particular market may display high price variability not because it is illiquid but because new fundamental information arrives frequently (p.630)."

Social cost of unrealized hedge value. The second component of the inefficient allocation is that when the intermediary has an inventory position differing from her desired position she skews the quotes to encourage trading by low-hedge-value investors on the side of the market that mean-reverts her inventory, but at the same time the quote skewing discourages trading by higher-hedge-value investors who arrive at the inventory-increasing side of the market. That is, even if transaction rates are unchanged on nonzero inventory positions the *expected* hedge value of the investors on the side of the market that she discourages from trading is higher than the *expected* hedge value of the investors on the other side of the market that she now encourages to trade. We will further develop this argument and illustrate it in the context of a dynamic model inspired by Ho and Stoll (1981).

1.1 Recursive dynamic inventory model setup

We aim to characterize price pressure dynamics and analyze how they relate to the two components of social cost relative to a first best of efficient risk-sharing among liquidity demanders. Ultimately, we will identify the time series properties of price pressures in the data. The structural model allows us to interpret what we find empirically and, in particular, allows for an analysis of social cost. Ho and Stoll (1981) are the first to set up the dynamic program for the risk-averse intermediary and cast it in a finite time horizon where utility is generated only by final wealth. To better fit our analysis of liquidity supplier inventory and prices on the NYSE we take the model, cast it in recursive form with an infinite horizon, and solve for the intermediary's optimal policy and the resulting stationary distribution.[7]

Liquidity demand process. Ho and Stoll (1981) use a reduced form model for the transaction rates (where transaction size is fixed at one unit) of public buyers and sellers that was first proposed by Garman (1976). These transaction rates are linear in the ask and bid price, respectively, where the public buy rate decreases in the ask price and public sell rate increases in the bid price.

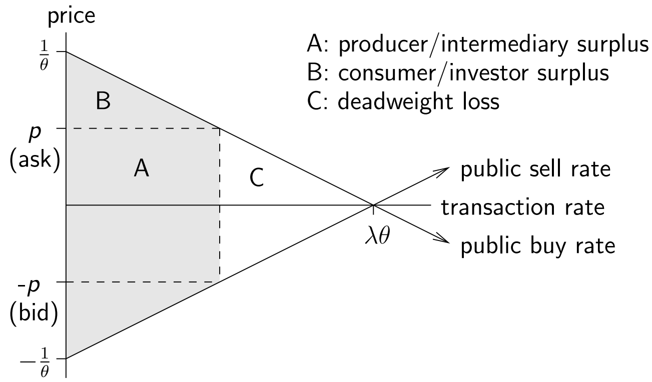
The social cost analysis requires a bit more structure on the liquidity demand process. Rather than separating the two sides of the market, we generalize the investor arrival process by considering agents who arrive at the market and depending on the prices quoted decided whether to buy, sell, or not trade. The investors incur idiosyncratic endowment shocks that can be hedged by an offsetting position in a (single) risky security that trades in the market. This hedge value v is stochastic and zero in expectation as it is a private value relative to a security's common value that we fix at zero throughout for ease of notation.[8] The agent arrives at the market and only trades if v is above the ask (agent buys a unit) or below the bid (agent sells a unit).[9] We parameterize this process as follows. The idiosyncratic shocks happen and agents arrive at the market according to a Poisson process with rate $2\lambda\theta$. Conditional on arrival, the (private) hedge value v is uniformly distributed

[7] Others also study dynamic inventory control. Madhavan and Smidt (1993) further enriches the Ho and Stoll (1981) setting by adding an informed liquidity demander. Amihud and Mendelson (1980) analyze a dynamic inventory model in recursive form where a risk-neutral intermediary maximizes average profit and stays away from an exogenously given maximum inventory constraint as arriving at that state limits her profit opportunity (she can essentially only exit on one side). We work from the Ho and Stoll (1981) model as we focus on the price risk associated with an suboptimal inventory position.

[8] The common value here is the equilibrium price in the absence of the liquidity friction that is modeled. It is the price at which all idiosyncratic shocks wash in the cross-section of investors. We refer to the investors' idiosyncratic desire to trade as their hedging values, although their private motivations for trade could be more general.

[9] If agents have the opportunity to contact the market more than once then v can be interpreted as the hedging benefit from the current trading opportunity until the investor's next contact with the market (see Weill (2007) and Lagos, Rocheteau, and Weill (2009) for more on investors' intertemporal optimization.)

with support $[-\frac{1}{\theta}, \frac{1}{\theta}]$. This specification implies the following graph for the public buy (sell) rate at the ask (bid) price:



where the arrival rate per hedge value is derived as the arrival rate ($2\lambda\theta$) times the hedge value density ($\frac{1}{2}\theta$) and the public buy and sell rates follow by integrating over hedge value rates from the ask to $\frac{1}{\theta}$ and from $\frac{1}{\theta}$ to the bid, respectively. The total hedge value rate (or the "size of the market") is obtained by setting the bid and ask price equal to zero and is thus representated by the size of the triangle defined by $-\frac{1}{\theta}$ and $\frac{1}{\theta}$ on the price axis and $\lambda\theta$ on the transaction rate axis. The size of this area is λ which therefore represents the size of maximum "consumer surplus." The realized hedge value rate depends on the bid and ask price set by the intermediary who "produces" liquidity. If she is risk-neutral and therefore operates on a zero cost rate, she charges the monopoly price $p = \frac{1}{2\theta}$ and, for public buys, captures a producer surplus rate equal to the square defined by A in the graph. The consumer surplus rate on this side of the market is the triangle B, and the deadweight loss due to monopoly power is C on this side of the market.

The second parameter θ captures the dispersion of hedge values and therefore governs the price sensitivity of the net transaction rate (defined as the public buy rate minus the public sell rate) standardized by the investor arrival rate. Lowering of both the bid quote and the ask quote by Δp increases the rate of buys minus sells by $2\Delta p\lambda\theta^2$ which, standardized by the arrival rate of $2\lambda\theta$, equals $\theta\Delta p$. Comparative statics show that this is an important parameter in the intermediary's optimization behavior. θ parallels the price elasticity of demand that a monopolist faces in a standard product market.

The intermediary's dynamic program. We set up the intermediary's optimization in discrete time in order to solve for her optimal policy and obtain a stationary distribution of pricing errors and inventory. This distribution characterizes the series' time series properties, i.e., contemporaneous as well as lagged moments. The Poisson hedge demand arrivals are modeled as discrete probabilities that are linear in prices.[10] The intermediary is risk-averse and maximizes expected utility where her utility is the standard time-separable CRRA utility. She is a monopolist liquidity supplier who produces liquidity supply by issuing firm price quotes that are take-it-or-leave-it prices that liquidity demanders see when arriving at the market. She can therefore not price discriminate and capture the full consumer surplus.

Let $V(.)$ be the maximum utility that the intermediary can achieve when starting off with an inventory of i_0 shares and a wealth of w_0:

$$V(i_0, w_0) = \max_{\{c_t, a_t \leq \frac{1}{\theta}, b_t \geq -\frac{1}{\theta}\}_{t=0}^{\infty}} E[\Sigma_{t=0}^{\infty} \beta^t u(c_t) | i_0, w_0] \tag{1}$$

subject to the budget constraint

$$w_{t+1} = R(w_t - c_t) + i_t \Delta m_{t+1} + a_t \max(-\Delta i_t, 0) - b_t \max(\Delta i_t, 0) \tag{2}$$

and subject to the following transition probabilities across inventory states

$$\begin{aligned} P[i_{t+1} = i_t - 1 | i_t] &= \int_{a_t}^{\frac{1}{\theta}} (2\lambda\theta) \left(\frac{1}{2}\theta\right) dv = \lambda\theta(1 - \theta a_t) \\ P[i_{t+1} = i_t + 1 | i_t] &= \lambda\theta(1 + \theta b_t) \end{aligned} \tag{3}$$

where i_t is the intermediary's inventory position at time t, w_t is her wealth, and c_t is her consumption at time t, β is her discount factor, R is the gross riskfree return, Δm_t is her stochastic dividend in period t+1 (which runs from time t to time t+1) which has zero expectation and a variance equal to σ^2, and b_t (a_t) is the bid (ask) price she sets at time t. The stochastic dividend stream is a modeling device to minimize the accounting in the model; we emphasize that no sources of uncertainty are ignored (see also Ho and Stoll (1981, p.52)).

We exploit the recursive nature of the problem which leads to the following Bellman equation

[10]This is a reasonable approximation as the Poisson process itself is the limiting case of event probabilities that are linear in small discrete time increments.

for the inventory state i (using the law of iterated expectations):

$$\begin{aligned} V(i,w) &= \max_{c_{iw},a_{iw},b_{iw}} u(c_{iw}) + \beta \underset{\Delta m}{E}\Big[V(i+1, R(w-c_{iw}) + i\Delta m - b_{iw})\Big]\lambda\theta(1+\theta b_{iw}) + \\ & \beta \underset{\Delta m}{E}\Big[V(i-1, R(w-c_{iw}) + i\Delta m + a_{iw})\Big]\lambda\theta(1-\theta a_{iw}) + \\ & \beta \underset{\Delta m}{E}\Big[V(i, R(w-c_{iw}) + i\Delta m\Big](1-\lambda\theta(1+\theta b_{iw}) - \lambda\theta(1-\theta a_{iw})) \end{aligned} \tag{4}$$

where Δm is the stochastic dividend in the oncoming period.

With upper bounds on the absolute inventory position and wealth it is straightforward to show that the functional equations defined by (4) define a contraction in the (i,w) space by verifying Blackwell's pair of sufficient conditions (see Ljungqvist and Sargent (2004, p.1012)). This implies the existence of a unique fixed point in the space of bounded continuous functions and therefore guarantees the existence of a unique equilibrium.

The numerical solution illustrates that the intermediary seems to endogenously stay away from large inventory and wealth upper bounds for our calibration where (i) relative risk aversion is larger one and (ii) the reciprocal of the discount rate (β^{-1}) is strictly larger than the riskfree rate (R). The endogenous upper bound on wealth is best understood based on Huggett (1993) which shows that for exogenous nondiversifiable endowment risk the two parameter restrictions guarantee that the agent does not let wealth grow infinitely large. The intuition is that with CRRA utility the agent's willingness to bear absolute risk grows with wealth. The reciprocal of the discount rate being strictly larger than the riskfree rate makes her prefer consuming today over tomorrow. It is therefore perfectly intuitive that at some level of wealth she prefers consuming out of wealth over accumulating more given that risk is less of a consideration on high wealth levels.

The model we propose extends this setting as the exposure to idiosyncratic risk (inventory) is decided upon endogenously. The same intuition applies for why she does not let wealth grow to infinity. In addition, the intermediary does not visit extreme inventory states as her earning potential does not change with inventory whereas her cost does given the stochastic dividend exposure ($i\Delta m$). In the numerical solution we set the bounds wide enough so that the intermediary decides never to visit the upper bound states. We return to this issue when discussing the numerical solution.

1.2 Price pressure and social costs: a model-based interpretation

The structure of the dynamic program allows for a calculation of how price pressures relate to the two components of social cost due to inefficient allocation of risk raised earlier. We focus on the price pressure's effect on the ask quotes as the effect on the bid quotes is symmetric.

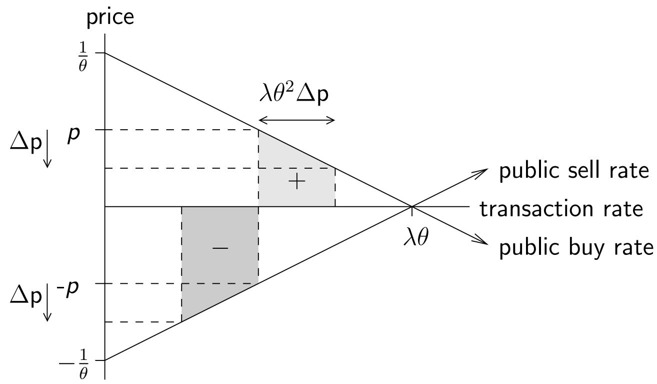
Price pressure as a measure of the intermediary's cost of risk. The first order condition for the ask price in the dynamic program optimization of equation (4) yields

$$\frac{1}{\theta} - a_{iw} = \frac{E[V(i-1, X_{iw} + a_{iw})] - E[V(i, X_{iw})]}{E[V_2(i-1, X_{iw} + a_{iw})]} \equiv -\Delta_i \tilde{V}_{iw} \tag{5}$$

where V_2 is the derivative of the value function with respect to wealth, $X_{iw} = (R(w - c_{iw}) + i\Delta m_{t+1})$, and the expectation is taken with respect to X_{iw} that has the stochastic dividend as its only source of randomness. The right-hand side expression denoted by $-\Delta_i \tilde{V}_{iw}$ is a measure of how a one unit reduction in inventory affects the intermediary's utility. The numerator captures the differential in the value function across inventory states but also across wealth states. The denominator effectively translates the utility difference into wealth units. $\Delta_i \tilde{V}_{iw}$ therefore becomes a measure of the marginal cost of a unit increase in inventory as it represents the amount one has to add to the intermediary's wealth for her to become indifferent between the two.

The ask price difference across inventory states is the change in the conditional price pressure, $a_{i+1,w} - a_{iw}$, and carries the interpretation of the *increase* in marginal cost of adding one inventory unit. If this conditional pressure is constant across inventory state changes (which theory and the numerical solution seem to support) the price pressure linearly increases in the value differential between inventory states: $a_{i+1,w} - a_{iw} = \alpha$ and, therefore, $a_{iw} - a_{0w} = \alpha i$. The average price pressure that we measure empirically as transitory volatility (conditional pressure times the standard deviation of inventory) then becomes a meaningful metric of how costly the idiosyncratic risk is to the intermediary relative to a zero inventory position.

Price pressure as a source of unrealized hedge value. Price pressure not only relates to the social cost of inefficient allocation due to the intermediary bearing risk, it also creates a cost on the liquidity demand side. Even if the size of the bid-ask spread is not changed (so transaction rates are not altered), skewing the quotes so that the midquote does not equal fundamental value (i.e., price is pressured) generates a social cost of unrealized hedge value. The argument is best illustrated graphically:



Here the bid and ask quotes are lowered to mean-revert a positive intermediary inventory position. This makes buys become more likely than sells. The overall transaction rate is not affected as the size of the spread is unchanged. The realized hedge value rate, however, is affected as the hedge value increase due to more buys (light gray area) is smaller than the hedge value lost by fewer sells (dark gray area).

Economically, the expected hedge value of the marginal buyer that the intermediary now accepts is lower than the expected hedge value of marginal seller that she blocks on the other side of the market. This is the source of the loss of hedge value on the demand side as a result of price pressures. If the bid-ask spread is constant across inventory states we can calculate how price pressures affect the realization of investors' hedging gains.

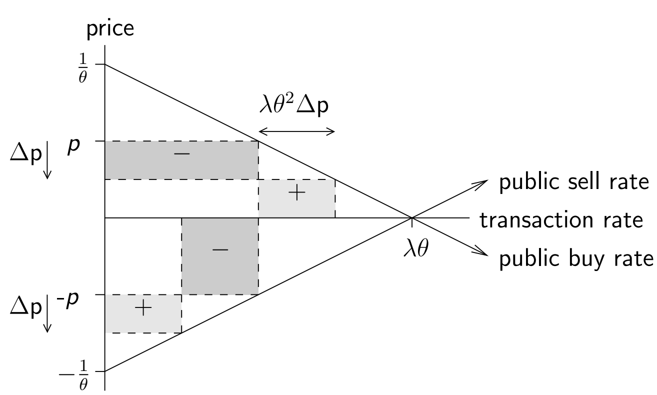
Empirically, the model allows one to retrieve the cost of unrealized hedge value as a function of the time series characteristics of price pressures. In the model, as illustrated by the graph above, the average social cost rate due to unrealized hedge value equals $\lambda\theta^2 E[(\Delta p)^2]$. The empirical equivalent of the right-most factor is the variance of price pressure. The factor by which it needs to be scaled also has an empirical equivalent as, in the model, a price pressure of $-\Delta p$ increases the net transaction rate (buys minus sells) by $2\lambda\theta^2\Delta p$. In the data, a state space model identifies a conditional price pressure parameter α that is the midquote deviation from the fundamental price as a linear function of the intermediary's inventory position, which we denote by I in the empirical analysis. The linearity of Δp in inventory implies that $E[(\Delta p)^2] = \alpha^2 \mathrm{Var}(I)$.

The data allows for estimation of the inventory mean-reversion rate by a first order autoregressive (AR1) process. The sensitivity of the net transaction rate ($\lambda\theta^2$) per price unit is equal to the average amount of inventory that the specialist mean-reverts divided by the conditional price pressure. At inventory state I this expression is $\frac{(1-\rho_1^I)I}{2|\alpha|I}$ where ρ_1^I is the first order autocorrelation in the intermediary's inventory which is a consistent estimator for the AR1 parameter. The I drops out due to linearity of both net transaction volume and price pressure in the inventory state I (motivated by the solution of the economic model). The empirical measure for the social cost rate due to unrealized hedge value thus becomes:

$$soc\_cost\_unrealized\_hedge\_val = \mathrm{Var}(\Delta p)\frac{(1-\rho_1^I)}{2|\alpha|} = \frac{1}{2}|\alpha|\mathrm{Var}(I)(1-\rho_1^I) \tag{6}$$

which is the price pressure variance multiplied by the rate of inventory mean reversion divided by the conditional price pressure; if price pressure is linear in inventory the reduction in realized hedge value is the conditional price pressure times the variance of inventory times one minus the persistence of inventory.

Is the trade-off of the two social costs efficient? The intermediary trades off the two sources of social costs when setting prices. She internalizes the unrealized hedge value loss as less surplus on the demand side implies less surplus for her to capture in the transaction process. But, does the intermediary internalize the unrealized hedge value appropriately so that immediacy production is socially efficient? The following graph illustrates how skewing the quotes affects the part of the surplus she gets:



A comparison with the true social cost depicted in the graph on page 11 shows that her private revenue loss is twice the social cost (essentially due to her inability to price discriminate and internalize the total hedge value effect). As her cost of pressuring price/skewing the quotes is larger than the true social cost, she might not pressure prices enough to obtain the constrained Pareto efficient allocation.

In summary, this subsection establishes an estimate of the social cost due to unrealized hedge value (see equation (6)). It is a first-order approximation based on the assumptions that (i) hedge value rates are constant in an interval around the efficient price, (ii) price pressure is linear in inventory and does not depend on wealth, and (iii) the bid-ask spread is constant across inventory-wealth states. The next section explores whether the last two assumptions are a reasonable characterization of the bid and ask quotes by analyzing the model's numerical solution.

1.3 The numerical solution: the optimal policy and implied stationary distribution

We numerically solve the intermediary's dynamic program which is summarized by equation (1). The calibration follows, to the extent possible, the base case parametrization proposed by Ho and Stoll (1981, p.67). We present the calibration details in Appendix I.

[insert Figure 1]

Figure 1 graphs the numerical solution to the intermediary's dynamic program. Panel A plots the value function which is defined as the maximum discounted utility the intermediary obtains conditional on starting off in a particular inventory-wealth state. Its concavity is generated by the time discounting in the dynamic program (see Ljungqvist and Sargent (2004, §A.2.)). Given the value function and the Bellman equations, it is a straightforward one period optimization to establish the intermediary's optimal control policy (i.e., bid price, ask price, and consumption) in each inventory-wealth state. This control policy in turn determines the system transition laws. If she is started off with sufficient wealth and zero inventory, she eventually ends up in the stationary distribution over the inventory-wealth states which is also depicted in Panel A. The stationary distribution illustrates that she endogenously chooses to stay far from the inventory and wealth upper and lower bounds.

Panel B of Figure 1 illustrates the optimal control policy which consists of the bid price, the ask price, and the consumption in each of the inventory-wealth states. The graphs provide the midquote and the bid-ask spread instead of the bid and ask price separately to better illustrate price pressures. The graph confirms the intuition that the intermediary skews the midquote to generate an transactions that mean-reverts her inventory. She lowers the midquote on long positions and raises it on short positions.

Two observations are particularly useful to motivate the econometric model we propose for the empirical analysis. First, the intermediary's policy seems to support the first order approximation we use in the econometric model which is that price pressure is linear in inventory and does not depend on wealth. The wealth result is particularly useful given that we do not have access to data on the intermediary's wealth level. Second, relative to midquote changes the bid-ask spread appears constant across states.[11] It is only slightly higher than the spread that a risk-neutral intermediary would charge: $2(2\theta)^{-1} = 0.2$.[12] This is a reassuring result as it confirms the Grossman and Miller (1988) argument that the spread captures the cost of processing orders (or monopoly rents in our case), whereas midquote changes capture the dynamic trade-off the intermediary does between staying one more period on a costly suboptimal position or discounting prices to mean-revert that position. This apparent orthogality of monopoly rents that are reflected in the spread and limited risk-bearing capacity that lead to price pressures suggests that competition in the intermediation sector on unchanged overall risk-bearing capacity will drive down the bid-ask spread to marginal cost but will leave price pressures unchanged. That is, the fundamental trade-off of discounting price vs. bearing idiosyncratic risk for an additional period remains. We note that this is our intuition and to formally derive this result is beyond the scope of the current paper.

The consumption graph illustrates that the intermediary self-insures against adverse dividend shocks by maintaining a wealth buffer. She consumes less than expected earnings on low wealth levels and more than expected earnings on high wealth levels. This is best illustrated by the zero inventory policies as her expected earnings of a potential current period transaction equal a constant 2*(1-0.5*0.23)*10=0.98 as prices are unchanged across the wealth dimension. She consumes less than these expected earnings on low wealth levels and thus self-insures against adverse future dividend shocks through savings. She consumes more than expected earnings and thus eats out of her large wealth buffer on high wealth levels. This consumption pattern thus makes wealth mean-reverting. We further notice that she reduces her consumption when on a larger nonzero inventory which reflects the higher need for a wealth buffer given that she is in a more risky state.

Comparative statics. The supplementary material presents comparative statics on the conditional price pressure (the slope of the midquote inventory curve), the average price pressure (transitory volatility), and the duration of price pressures. It compares the base-case numerical solution to the ones where friction is increased either through higher fundamental volatility (σ), a lower total hedge value rate (λ), or more dispersed hedge values (θ). One result that one could

[11] Zabel (1981) and Mildenstein and Schleef (1983) find that the spread is independent of inventory. Ho and Stoll (1981) find inventory has a very small effect on the spread.

[12] It is not surprising that the price of immediacy, i.e., the spread, is slightly higher since the "marginal cost of production" is no longer zero but becomes positive when the intermediary is risk-averse.

not get from a static model is that higher fundamental volatility or a lower total hedge value rate increase conditional price pressure (α) more than transitory volatility ($\alpha\sigma(i)$).[13] The reason is that the intermediary endogenously decides to not let inventory grow large, i.e., the standard deviation of inventory declines. This is a useful result to have when comparing price pressure results across small- and large-cap stocks.

2 An empirical search for conditional and average price pressure and inventory dynamics

In this section we analyze a 12-year balanced panel of 697 NYSE stocks that has end-of-day observations on the midquote (i.e., the average of the bid and ask price) and the NYSE specialist inventory position. We aim to identify the conditional and the average size of price pressure for these stocks both in the cross-section and through time. The structural model suggests an interpretation of the size of the average pressure as a measure of the social cost due to the intermediary bearing risk. In addition, we characterize the inventory dynamics with a first order autoregressive process that, again contingent on the structural model, allows us to calculate the size of the social cost of price pressure due to unrealized hedge value (see equation (6)). Before any estimations, we discuss the data and provide some summary statistics.

2.1 Data and summary statistics

We use CRSP, the NYSE's Trade and Quotes (TAQ), and a proprietary NYSE dataset with specialist positions to prepare data on the end-of-day midquote and NYSE specialist position along with other variables from 1994 through 2005. We construct a balanced panel to make results comparable through time (which controls for stock fixed effects). We start with the sample of all NYSE common stocks that can be matched across TAQ and CRSP and retain the stocks that are present throughout the whole sample period. We then use the stock-split and dividend information from CRSP to remove these effects from the midquote prices in TAQ. Stocks with an average share price of less than \$5 are removed from the sample, as are stocks with an average share price of more than \$1,000. The resulting sample comprises 697 common stocks. Stocks are sorted into quintiles based on market capitalization. Quintile 1 refers to the large-cap stocks and quintile 5 corresponds

[13] Strictly speaking, α is the parameter in the econometric model that captures the linear relationship between midquote and the inventory. Its equivalent in the economic model is the population regression of the midquote on inventory using the stationary distribution.

to the small-cap stocks. We further convert the NYSE specialist position which is in shares in the original database into dollars in order to facilitate comparisons across stocks. We multiply the position with the sample average price so as not to introduce daily price changes in the inventory variable. This would contaminate it as explanatory variable for the transitory price effect in the econometric model.

[insert Table 1]

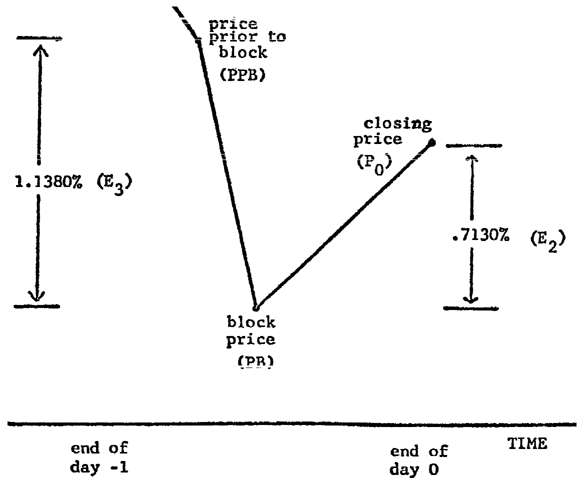
Summary statistics. Table 1 presents the mean of various trading variables by size quintile. It also presents the within variation which is defined as the standard deviation of the data series after removing stock fixed effects. It thus gives a sense of the variable's variability through time. The statistics lead to a couple of observations. First, the average position of the specialist is positive and economically significant. For the large-cap stocks in Q1, for example, she maintains an almost half a million dollar average inventory position. This position is undoubtedly driven by a cost asymmetry across being in a long and a short position in the course of the trading process (something that is beyond the scope of this study).[14] The inventory position for the small-cap stocks in Q5 is $77,900 and thus is considerably smaller than the average position for the large-caps. Second, the within standard deviation in specialist inventory is $1.4 million, which is substantial relative to her average position. It suggests that the specialist is an active intermediary in matching buyers and sellers through time. We will disaggregate this variation by year and by quintile in a later table. Third, the market capitalization is $34 billion for Q1 stocks and declines to $290 million for Q5 stocks. The effective half-spread, which is defined as the distance between a the transaction price and the prevailing midquote, is 8 basis points for Q1 stocks and 46 basis points for Q5 stocks. These numbers show that there is considerable heterogeneity across stocks.

2.2 State space model to distinguish price pressures from efficient price innovations

The challenge to identify price pressures in real markets is that net order imbalance (the dual of the intermediary's inventory change) might convey information as well as cause pressure which makes prices "overshoot." This well-known pattern has been documented in various event studies. Kraus and Stoll (1972), for example, show that prices overshoot in the event of a block trade, i.e.,

[14] We interpret the zero position of our model as the deviation from the long-term optimal position of the specialist.

a transaction where the intiating party wants to trade an extraordinary amount of shares. They document the following pattern for block sales:



source: Kraus and Stoll (1972, p.575)

To identify the pressure effect in the presence of an information effect, we use the state space approach of Menkveld, Koopman, and Lucas (2007) which models an observed, high-frequency price series as the sum of two unobserved series: a nonstationary efficient price series ("information") and a stationary series that captures transitory price effects ("pressures"). We use log prices throughout the paper and remove a required return by subtracting a linear trend with a slope equal to the riskfree rate plus beta times a market risk premium of 6%. In its simplest form the model structure for the detrended log price is:

$$p_t = m_t + s_t, \tag{7}$$

$$m_t = m_{t-1} + w_t, \tag{8}$$

where p_t is the observed price, m_t is the unobserved efficient price, s_t is the unobserved transitory price effect, and w_t is the innovation in the efficient price. s_t and w_t are mutually uncorrelated and normally distributed. It is immediate from the structure of the model that only draws on w_t affect the security's price permanently; any draw on s_t is temporary as it affects prices only at a single point in time. The model is estimated with maximum likelihood where the likelihood is calculated by means of the Kalman filter.

We prefer a state space approach over other approaches (e.g., GMM or an ARIMA model) for several reasons. First, maximum likelihood estimation is asymptotically unbiased and efficient. Second, the model implies that the differenced series is an invertible MA(1) time series model which implies a infinite lag AR model. This is particularly cumbersome if the price series (or the inventory series that we will use as explanatory variable for s_t) have missing values. The Kalman filter ensures maximum efficiency as it deals in the most efficient way with missing values. That is, it does not lose any information as it considers the likelihood of *all* level series changes even if they necessarily extend over multiple periods in case of missing observations. Any method based on the differenced series does not consider that information. Third, after estimation, the Kalman smoother (essentially a backward recursion after a forward recursion with the Kalman filter) facilitates a series decomposition where at any point in time the efficient price and the transitory deviation are estimated using past and future prices. This allows for an in-sample decomposition of prices (as we will do to illustrate our results). We refer to Durbin and Koopman (2001) for an extensive treatment on the use of state space models to analyze economic times series.

In the remainder of this subsection develops the general state space model to be taken to the data. We first develop the latent efficient price process and then the stationary price deviations that should capture the price pressures that we are after.

Unobserved efficient price process. We use the model to analyze daily midqoute series by stock-year. The efficient price series now is a martingale that consists of two components:

$$m_{it} = m_{i,t-1} + \beta_i \hat{\gamma}_t + w_{it}, \quad w_{it} = \kappa_i \hat{I}_{it}^{idio} + u_{it} \tag{9}$$

where the subscript i indexes stocks, t indexes days, $\hat{\gamma}_t$ is a common factor innovation that is obtained as the residual of an autoregressive time series model applied to the market cap weighted average of standardized midquote returns (we standardize to control for heteroskedasticity), w_{it} is the idiosyncratic innovation, $\hat{I}_{it}^{idio}$ is the idiosyncratic inventory innovation (same procedures are used as described for $\hat{\gamma}_t$) that represents the "surprise" net order imbalance which is potentially informative, and u_{it} is the stock-specific innovation orthogonal to this order imbalance innovation and assumed to be a normally distributed white noise process.[15] The decomposition of efficient price innovation in into a common factor component ($\beta_i \hat{\gamma}_t$) and an idiosyncratic component (w_{it}) is relevant for our purposes as only the latter represents undiversifiable risk for the specialist. The common factor risk is easily hedged through highly liquid index products. It is for the same reason

[15]It is necessary to add (although not the focus of the paper) the idiosyncratic inventory innovation term in the martingale equation to ensure unbiased estimation. The reason for a potential "omitted variable bias" is that the explanatory variable for the efficient price innovation (w_t) correlates with the explanatory variable for price pressure (s_t), i.e., the order imbalance innovation in period t correlates with inventory at time t (which is at the end of period t).

that we remove the common factor from inventory dynamics as the price risk over a market-wide shock to a specialist firm inventory is again easily hedged through index products.

Unobserved transitory price deviations process. We propose the following process for the stationary price deviations:

$$s_{it} = \varphi_i s_{i,t-1} + \alpha_i I_{it}^{idio} + \beta_i^0 \hat{\gamma}_t + \cdots + \beta_i^k \hat{\gamma}_{t-k} + \varepsilon_{it} \tag{10}$$

where the error term ε_{it} is normally distributed and uncorrelated with w_{it}. The general specification nests the two models that we will take to the data. In an exploratory model without the inventory position we set α_i equal to zero. This allows us to test whether the transitory price effects are persistent at the daily frequency. In the main model where we allow these transitory price effects to be explained by inventory we set φ_i equal to zero as the inventory series now captures the persistency. The gamma terms enter the stationary price effect equation to capture a documented lagged adjustment to common factor innovation that is particularly prevalent for relatively inactively traded small-stocks (see, e.g., Campbell, Lo, and MacKinlay (1997)). In the proposed specification the beta coefficient in the efficient price process captures the long-term impact of a common factor shock on the price of the security and any lagged adjustment shows up through negative beta coefficients in the transitory price effects equation.

Observed price process. We close the econometric model with the observation equation:

$$p_{it} = m_{it} + s_{it}. \tag{11}$$

As we intend to analyze price pressures in the cross-section as well as in the time dimension we do all empirical analysis consistently by stock-year. To report the 697*12 stock-year results we aggregate stocks into bins according to their size. The allocation across size quintiles is fixed throughout the sample so as to ensure that bins are comparable across years. Means are calculated for each bin along with the number of t-statistics that are outside of the 0.10 to 0.90 quantile interval. These t-statistics are available as supplementary material on the authors' websites. The tables report the p-value of a meta test statistic that counts the number of significant t-values in a size-year bin (and in the aggregation across bins). This statistic is binomially distributed where the probability of "success" equals the significance level of the t-test performed for each stock-year estimation, i.e., 0.20 in our two-tailed test.[16]

Time series statistics as early evidence on price pressures. Before turning to the model estimates, some time series statistics are useful to test whether the price effects that we are after are born out by the data.

[insert Table 2]

Table 2 reports first and second order autocorrelation of idiosyncratic midquote returns. The effects of contemporaneous and lagged adjustment to the common factor innovation are removed by regressing the midquote return on the common factor innovation up to four lags. The residuals serve as the idiosyncratic returns. Consistent with the individual stocks autocorrelation results in Campbell, Lo, and MacKinlay (1997) the average first order autocorrelation is negative in 64 of the 70 size-year bins. The low p-values indicate that the coefficient estimates are highly significant at conventional significance levels.[17] The negative first order autocorrelation is consistent with transitory price effects as the simple state space model of equations (7) and (8) implies a negative first-order correlation in midquote returns. The table further shows that the second order autocorrelation is also significantly negative which is an early indication that the simple model needs to be extended to allow the price deviations s_t to be positively autocorrelated. Price pressures appear to carry over days rather than being just an ultra-high-frequency intraday phenomenon. This also implies that the unconditional deviations are potentially much larger relative to fundamental volatility than the simple first-order autocorrelations seem to suggest.[18]

Table 2 further reports the standard deviation of specialist inventory, its autocorrelation, and a cross-correlation with subsequent midquote returns as early evidence on the conjectured relation between transitory midquote deviations and inventory. The standard deviation of specialist end-of-day inventory varies is \$1.131 million for the large-cap stocks and monotonically decreases to \$165,000 for the small-cap stocks. It is relatively constant throughout time but tapers off in the last few years in the sample. The cross-sectional variation is undoubtedly due to a higher fundamental

[16] We believe the inference procedure does not suffer from commonality across stocks as the empirical analysis focuses on idiosyncratic effects by removing a common factor in both the price and in the inventory series.

[17] Overall, 3525 t-values are significant and 4836 t-values are insignificant. The sign of the significant t-values is primarily negative (2427 negative vs. 1098 positive) which shows that the negative means are statistically significant. We omit extensive statistical significance discussion in the remainder of the document for brevity.

[18] The AR(1) process for the stationary term s_t implies a first-order autocorrelation in midquote returns of

$$\rho_1 = \frac{-\frac{1-\varphi}{1+\varphi}\sigma_\varepsilon^2}{\sigma_w^2 + \frac{2}{1+\varphi}\sigma_\varepsilon^2} \tag{12}$$

where φ is the autoregressive coefficient in the s_t process. The autocorrelation is less negative if pressure is more persistent, i.e., if φ increases.

volatility and a smaller market (lower hedge value rate) for small-cap stocks. The economic model predicts that the intermediary shies away from frequent and large nonzero inventory position which explains why the inventory volatility is considerably smaller for small-cap stocks.

The inventory volatility is roughly twice the mean inventory across all quintiles. This is evidence of an active intermediation on a day over day level. In other words, inventory management is not a phenomenon that is restricted to the intraday ultra-high frequency level where intermediaries "go home flat." The table further documents a significant first-order autocorrelation in inventories which show that these positions could last for multiple days. Again, there is considerable cross-sectional variation as the average autocorrelation for the large-cap stocks is 0.28 which monotonically increases to an average autocorrelation of 0.72 for small-cap stocks.[19] The specialist seems to trade out of most of an end-of-day position in the course of the next day for the large-cap stocks, whereas it take multiple days for the small-cap stocks. Finally, we calculate the correlation between today's inventory position and tomorrow's midquote return to verify whether the the two sets of results in the table can be reconciled. The last panel in the table shows that today's inventory position correlates significantly with tomorrow's midquote return. The positive signs are consistent with the economic model as the intermediary lowers the midquote on a long position (relative to the long-term average) which elicits an order imbalance (more investor buying than selling) that makes inventory mean-revert and, as a consequence, leads to lower midquote discounts. This creates a positive correlation between today's position and tomorrow's midquote return.

State space model estimates. We estimate two versions that are nested in the state space model defined by equation (9), (10), and (11). The first version does not include the idiosyncratic inventory position as explanatory variable and captures all potential persistence in an AR(1) process on the price deviation, i.e., it sets α_i equal to zero. The second version replaces the latent persistence in price deviations with the idiosyncratic inventory position relative to the long-term mean inventory, i.e., it sets φ_i equal to zero.

[insert Table 3]

Table 3 presents the model estimates without inventory as an explanatory variable for price pressure. It shows that these transitory price effects are large and persistent. In terms of size, the standard deviation of the error term in the stationary price pressure equation (10) is statistically significant and is 58 basis points for the large-cap stocks and 67 basis points for the small-cap

[19] These inventory autocorrelations are lower than the puzzlingly large autocorrelations found in NYSE data from the late 1980s and early 1990s in Madhavan and Smidt (1993) and Hasbrouck and Sofianos (1993).

stocks. These price pressures are persistent as the AR(1) coefficient are significant and range from 0.46 for the large-cap stocks to 0.39 for the small-cap stocks. With these numbers one can calculate the unconditonal price pressure with is reported in Panel E and ranges from 83 basis points for small-cap stocks to 79 basis points for large-cap stocks. The table further reports the size of these pressures relative to the size of the innovation in efficient price through variance ratios. These ratios are 0.40 for the large-cap stocks and 0.31 for the small-cap stocks, respectively. They increase to 0.66 and 0.37, respectively, if only idiosyncratic innovations in the efficient price are considered. These ratio illustrate that these price pressures are a nonnegligible part of daily price volatility.

The estimates further confirm two well-documented empirical facts for the cross-section of stocks. The standard deviation of idiosyncratic daily midquote returns is 142 basis points for the large-cap stocks and increases monotonically to 211 basis points for the small-cap stocks. We further find significant evidence of lagged adjustment to common factor innovations predominantly for small-cap stocks. These results are reported in an extended version of the table that is included in the supplementary material.

[insert Figure 2]

Figure 2 illustrates the model's estimates for twenty trading days in a Q5 stock (PERMNO 635030) starting January 8, 2002. It exploits one attractive feature of the state space approach, which is that conditional on the model's parameter estimates the Kalman smoother generates estimates of the unobserved efficient price and price pressure processes conditional on all observations, i.e., it uses past and future observations to estimate the efficient price m_{it} and the temporary price deviation s_{it} at any date t in the sample. The first graph in Panel A plots the observed midquote along with the efficient price estimate and illustrates that the midquote can be far removed from the efficient price. The second graph further decomposes the efficient price into the common factor random walk, the idiosyncratic innovation random walk, and an integrated required return. Panel B decomposes the stationary price deviation series into a delayed adjustment to the common factor (left) and the nonexplained remainder part (right). It also plots the idiosyncratic inventory as a deviation from its long-term mean on a second y-axis in the graph with the remainder part. The plot provides suggestive evidence that the temporary price deviations might indeed be the price discounts that a representative intermediary applies to mean-revert inventory. We turn to estimation of the model with inventory as explanatory factor to do proper empirical testing.

[insert Table 4]

Table 4 reports the estimates of the state space model with inventory as an explanatory variable for price deviations and their persistence. The key parameter α_i that measures the conditional price pressure has the conjectured sign and is highly significant as 4045 t-values are significantly negative, 4051 are insignificant, and only 265 are significantly positive. Prices are low when the intermediary is on a long position and high when she is on a short position relative to her long-term mean inventory. There is substantial cross-sectional variation in conditional price pressure as α_i is -0.02 for the large-cap stocks and -1.01 for the small-cap stocks. These numbers are economically significant as a $1.131 million (one standard deviation) position change in specialist inventory creates a price pressure of 1131*0.02=17 basis points (cf. effective half-spread of 8 basis points, see Table 1). A similar position change in the small-cap stocks would create a price pressure of 1131*1.01=1142 basis point or 11.42%! The higher conditional price pressures are undoubtedly the result of higher fundamental volatility and a thinner market (lower hedge value rate) in the context of the economic model. The results further show that the price pressure of inventory is of the same order of magnitude as the information in surprise imbalance as captured by the parameter κ_i (available in the supplementary material). The average α_i is -0.28; the average κ_i is -0.35.

The table further shows that the average price pressure varies less in the cross-section than the conditional price pressure. By the logic of the economic model and as evidenced by a lower standard deviation of specialist inventory, the intermediary visits large inventory positions less often on higher fundamental volatility or when the market is thin. The average pressure which is measured as the conditional pressure times the standard deviation of inventory is 17 basis points for the large-cap stocks and 120 basis points for the low-cap stocks. The conditional pressure is roughly 7 times higher for the small-cap stocks relative to the large-cap stocks whereas the conditional pressure is 50 times higher. This is consistent with a dampening effect that dynamic inventory control has when fundamental volatility is increased or the market is thinner. The table also reports the size of these average price pressures relative to permanent volatility. The variance ratio is 0.02 for the large-cap stocks which indicates that these price pressures are small relative to fundamental volatility. This is different for small-cap stocks where the ratio is 1.32: transitory volatility due to price pressures are larger than fundamental volatility. Finally, price pressure appears to capture almost half of the variance of the transitory price deviations in the benchmark model that does not use inventory data (price pressure vs. transitory price deviation ratio is $(49/75)^2 = 0.42$ (cf. Panel E in Table 3 and 4)).

The social cost due to unrealized hedge value is economically significant and varies considerably in the cross-section of stocks. The structural model suggests a proxy for this social cost based on the the conditional price pressure, the variance of inventory, and the first order autocorrelation of inventory (see equation (6)). Panel H of Table 4 presents the cost estimate which is \$9.86 million per year for the largest-cap stocks and monotonically declines to \$0.67 million for the smallest-cap stocks. Cross-sectionally, these results are the exact opposite of the average price pressure results that increased with size. This is undoubtedly caused by the thickness of the market (hedge value rate) as it has an opposite effect on the size of average price pressure and the unrealized hedge value rate.

The aggregate magnitude of the social costs of unrealized hedge value is economically large. The average yearly costs in \$3.60 million per NYSE common stock. Multiplying this by the 697 stocks in the panel is \$2.5 billion per year. On average our balanced panel comprises less than half of the total common stock years. Assuming that social costs in the stocks which are not in the panel is the same as stocks in the panel would double the annual costs to \$5.0 billion.[20] Therefore, the total social costs for our 12-year sample period are estimate to be \$60 billion. As noted earlier this is only one component of the inefficiency allocation of risk. The intermediary's cost of bearing risk is the other component which we are not able to measure.

2.3 Price pressure and NYSE market structure

When interpreting our results it is worth discussing the institutional structure of the specialist intermediary at the NYSE. The NYSE grants the specialists a central position in the trading process and imposes obligations upon the specialists. Panayides (2007) shows that the most significant obligation, the Price Continuity Rule which "requires the specialist to smooth transaction prices by providing extra liquidity as necessary to keep transaction price changes small," is important at the transaction-level horizon. Panayides finds that the rule causes specialists to accumulate inventory to prevent "transaction prices from overshooting beyond their equilibrium levels." This causes inventories to be positively associated with transitory price effects, the opposite of our relation between intermediary inventory and price pressure. Therefore, if the Price Continuity Rule manifests itself at a daily frequency it causes underestimation of price pressures associated

[20]It is possible that the stocks that are not present for the entire sample period are smaller than stocks present throughout. On the other hand, the number of stocks listed on the NYSE is highest in 1999-2001 when the social costs of price pressure are substantially above the full sample average. Given that the goal is to calculate an approximate number we do not quantify either of these two effects.

with inventory.

While the NYSE designates a single intermediary, it is possible for other investors to compete with the specialist by placing limit orders to supply liquidity. Such a possibility is especially important given the NYSE's recent market structure changes (after our sample period) which resulted in a reduced role for the specialist (Hendershott and Moulton (2007)). Additional liquidity suppliers could reduce the width of the bid-ask spread and could also reduce the social costs of intermediaries bearing risk. The most efficient manner to share risk is for the inventory to be immediately and equally shared across all liquidity suppliers, leading to perfectly correlated positions.[21] How would this affect our estimates of price pressure? First, the conditional price pressure per unit of inventory should be adjusted by the specialist's fraction of inventory, e.g., if the specialist carries one half of the total inventory then the conditional price pressure should be multiplied by one half. The average price pressure is unaffected by additional liquidity suppliers because the standard deviation of inventory is adjusted by the reciprocal of the adjustment to the conditional price pressure. The unrealized hedge value is underestimated by the same magnitude as the underestimation of the standard deviation of inventories (see equation (6)).

3 Conclusion

We construct a theoretical model to understand and characterize the effects of price pressure—the deviation of prices from fundamental values due to the inventory risks born by an intermediary providing liquidity to asynchronously arriving investors with idiosyncratic hedging needs. The structure of the model allows for estimation of the social costs of price pressure due to a reduction in investors' realized hedging value. Empirically we estimate price pressure using 12 years of NYSE intermediary data. We find:

1. A \$100,000 inventory shock causes price pressure of 1.01% for the small-capitalization stocks and 0.02% for the large-cap stocks. Price pressure conditional on inventory reduces the variability of the intermediaries' inventory.

2. The daily transitory volatility in stock returns due to price pressure (a measure of average price pressure) is large: 1.20% and 0.17% for small and large stocks, respectively. For small

[21] Lagos, Rocheteau, and Weill (2009) find that competitive intermediaries implements the constrained Pareto optimal solution. Our focus is on deviations from first-best, i.e., the allocation where all idiosyncratic risk is immediately and perfectly shared across investors.

stocks the ratio of transitory volatility due to price pressure to the permanent (random-walk or efficient price) volatility is greater than one.

3. A model-based estimate of price pressure's annual social cost from lower realized hedging gains is \$9.86 million and \$0.67 million for large and small stocks, respectively. The aggregate social cost is estimated to be \$60 billion for NYSE common stocks for our sample period. The yet unmeasured social costs of inefficient allocation from the intermediary costs of bearing risk make the total social costs related to price pressure even higher.

The significant social costs of price pressure suggest that a goal of financial market regulation should be to mitigate price pressure. One way to do this is by increasing capital for intermediation as the greater the risk bearing capacity of the intermediaries the smaller the price pressure. Another approach would be to lower costs for investors to monitor the market. This would lead to investor trading being more responsive to price pressures, reducing the duration of price pressure by allowing intermediaries to mean-revert their inventories more quickly.

Appendix I: Details on the calibration and the solution method

Base case model calibration. We numerically solve the intermediary's dynamic program which is summarized by equation (1). The calibration follows, to the extent possible, the base case parametrization proposed by Ho and Stoll (1981, p.67).

1. The intermediary's coefficient of relative risk aversion (ρ) is 2.
2. The time interval length is one day.
3. The arrival rate function is $g(x) = \lambda_0(1 - \theta x)$ where $\lambda_0 = 1$ and $\theta = 5$ so that each day the representative investor arrives and considers a trade. The transaction size is 10 units.[22]
4. The stochastic dividend risk on one inventory unit equals $\sigma = \sqrt{(}0.50/200) = 0.05$ where one year contains 200 trading days.
5. The intermediary's discount rate β is 0.9 to capture the two week horizon Stoll proposes for the intermediary.[23]
6. The intermediary earns a gross daily riskfree rate of $R = 1 + 0.10/200$ on her savings for the next period.

We then discretize the set of inventory-wealth states (i, w) by choosing a grid with carefully selected bounds for both the inventory and the wealth dimension of the state. We choose upper bounds for inventory and wealth so that the state-space become a finite set. We set inventory bounds equal to -50 and 50 and verify

[22] Ho and Stoll (1981) use the same θ but set $\lambda_0 = 2,000$ as the annual transaction rate. To match their arrival rate we set the transaction size equal to 10 with the interpretation that one successful daily transaction means 2,000/200=10 intraday transactions of one inventory unit.

[23] A 0.9 discount rate implies that roughly two-thirds of the total discounted value of a dividend stream of one falls into the first two weeks.

that ex-post these inventory states indeed appear to be nonbinding constraints.[24] The step size for the inventory dimension is governed by the transaction size 10. We choose a wealth upper bound of 10 where we again verify that the limit does not bind endogenously. We set the wealth lower bound equal to 0 which is an absorbing bankruptcy state from which point on consumption is arbitrarily small.[25] The wealth step size is constrained by the stochastic dividend risk which is implemented as a discrete random variable where the outcome is plus or minus σ with equal probability. As the inventory dimension only contains multiples of 10, the stochastic dividend commands a (maximum) wealth step size of 10*0.05=0.5.[26]

We then set up the discretized version of the Bellman equation which is a trivial extension of the system defined by equation (4). The step size on the wealth dimension necessarily discretizes the bid and ask prices that the intermediary can choose, but we can get as close as one might desire to a continuum of prices by shrinking the wealth step size at the cost of computational speed. After discretizing, we solve the dynamic program numerically by iterating on the Bellman equation where we rely on the contraction property of the recursive mapping to achieve convergence.

Bellman equations for the interior of the inventory-wealth state domain. We arrive at the following discrete state dynamic program

$$\begin{aligned} V(i,w) &= \max_{\tilde{w}_{iw},a_{iw},b_{iw}} u(w - R^{-1}\tilde{w}_{iw}) \\ &\quad +\beta \mathop{E}_{\Delta m}[V(i,\tilde{w}+i\Delta m)](1-g(a_{iw})-g(b_{iw})) \\ &\quad +\beta \mathop{E}_{\Delta m}[V(i+1,\tilde{w}+i\Delta m+b_{iw}\Delta i)+\Delta m]g(b_{iw}) \\ &\quad +\beta \mathop{E}_{\Delta m}[V(i-1,\tilde{w}+i\Delta m-a_{iw}\Delta i)+\Delta m]g(a_{iw}) \\ g(x) &= 1-\theta x \quad \text{(probability of a transaction)} \\ & \quad i\in\{-50,-40,\dots,50\} \quad w\in\{0,0.5,\dots,10\} \end{aligned} \tag{13}$$

where the three expected value terms correspond to no-arrival, seller-arrival, and buyer-arrival, respectively. Equation (13) defines the Bellman equations for the nonboundary inventory-wealth states. But, before turning to the boundary Bellman equations, we need to make sure that we stay on the inventory-wealth grid in the iterations. This motivates a binomial distribution for stochastic dividend

$$P[\Delta m = -\sigma] = P[\Delta m = +\sigma] = 0.5 \tag{14}$$

and restricts the set of admissible controls to

$$a_{iw}, b_{iw} \in \{0, 0.05, 0.10, \dots\} \text{ s.t. } g(a_{iw}), g(b_{iw}), 1-g(a_{iw})-g(b_{iw}) \geq 0 \tag{15}$$

$$\tilde{w}_{iw} \in \{0, 0.5, \dots, \min(10, 0.5\lfloor 2Rw \rfloor)\} \tag{16}$$

The upper bound for end-of-period wealth $\tilde{w}_{iw}$ prevents negative current period consumption.

Bellman equations for the boundary of the inventory-wealth state domain. As for the boundary inventory states, we adjust the Bellman equation defined in equation (13) by reducing the set of admissible controls. For the maximum inventory state, for example, we restrict the bid price to $b_{Iw} = \theta^{-1}$ (which effectively sets the seller arrival rate to zero). As for the boundary wealth states, in the zero wealth state we keep the intermediary on a low enough consumption level relative to her earning power in the nonzero wealth state to make her endogenously choose to stay away from the bankruptcy state. In states close to the maximum wealth state, we redirect her to the maximum wealth state if she enjoys a stochastic dividend

[24] One has to be careful here, as in addition to it being a very risky state, the intermediary also internalizes that she can only exit this state on one side (i.e., less inventory) which limits her earning potential. Ideally, we would like to let the inventory upper bound I tend to infinity to rule out this nonrisk explanation. We "simulate" such limit numerically and verify that the solution does not change by taking larger upper bounds.

[25] CRRA utility satisfies the Inada conditions so that intermediaries will steer clear of this state.

[26] Technically, the intermediary could still end up strictly outside of wealth interval if she gets a large enough draw on the stochastic dividend. In the implementation, we steer her to the lower- or upper bound of wealth in the cases that this happens. This does not affect our solution as the intermediary endogenously chooses to stay away from those states.

that would make her transit to a larger wealth than the maximum wealth. One interpretation is that the government taxes "excessive wealth" away in these high wealth states. We reiterate that these assumptions on the maximum inventory or maximum wealth Bellman equations are inconsequential as we choose the maximum inventory and maximum wealth boundary large enough so that the intermediary endogenously chooses never to get near to these states (see discussion footnote 26).

From this point on we follow Ljungqvist and Sargent (2004, §4.2, p.95), i.e., we iterate on the Bellman equation and the contraction property of the recursive mapping guarantees convergence to the unique solution.

Appendix II: Details on the likelihood optimization

The likelihood of the state space model described by equations (9-11) is optimized in three steps to minimize the probability of finding a local maximum. The optimization is implemented in Ox using standard optimization routines. The Kalman filter routines are from ssfpack which is an add-on package in Ox (see Koopman, Shephard, and Doornik (1999)).

1. An OLS regression of the log price series first difference on contemporaneous and lagged $\hat{\gamma}_t$ yields starting values for β_i and $\beta_i^0, \ldots, \beta_i^k$ (see equations (9,10)). These β estimates are fixed at these values until the final step.

2. The likelihood is calculated using the Kalman filter (see Durbin and Koopman (2001)) and optimized numerically using the quasi-Newton method developed by Broyden, Fletcher, Goldfarb, and Shanno. In the optimization all parameters are free except for the βs and $(\sigma(\varepsilon), \varphi)$ which are fixed at values that are picked from a nine by nine grid. φ ranges from 0 to 0.8 and $\sigma(\varepsilon)$ ranges from 0 to a stock-specific upper bound that is calculated assuming that 80% of a stock's unconditional variance is price pressure. The likelihoods are compared across all 9*9=81 optimizations and the $(\sigma(\varepsilon), \varphi)$ value that yields the highest likelihood is kept as starting value for the final optimization. The rationale for this step is to prevent numerical instability of the quasi-Newton optimization. That is, if all parameters are free on arbitrary starting values the optimization routine often runs off to a persistence parameter φ that approaches one and a price pressure volatility that approaches the the stock's unconditional volatility, i.e., it starts to load the observed price series on two nonstationary series (i.e., the efficient price and the price pressure) and becomes unstable.

   The Kalman filter is initialized with a diffuse distribution for the unobserved efficient price m_0 and the unconditional price pressure distribution for s_0, i.e., $s_0 \sim \mathrm{N}(0, \frac{\sigma^2(\varepsilon)}{1-\varphi^2})$.

3. The likelihood is optimized where all parameters are free and starting values for $(\beta_i, \beta_i^0, \ldots, \beta_i^k, \sigma(\varepsilon), \varphi)$ are equal to those found in steps 1 and 2.

This procedure proves numerically stable as we have strong convergence in the likelihood optimization for all of our stock-year samples, i.e., convergence both in (i) the likelihood elasticity w.r.t. the parameters and (ii) the one-step change in parameter values (they both become arbitrarily small).

References

Acharya, V. V., H.-S. Shin, and T. Yorulmazer (2008). Fire-sale fdi. Manuscript, NYU.

Amihud, Y. and H. Mendelson (1980). Dealership market: Market-making with inventory. *Journal of Financial Economics 8*, 31–53.

Brunnermeier, M. K. and L. H. Pedersen (2008). Market liquidity and funding liquidity. *Review of Financial Studies (forthcoming).*

Campbell, J. Y., S. J. Grossman, and J. Wang (1993). Trading volume and serial correlation in stocks returns. *The Quarterly Journal of Economics 108*, 905–939.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets.* Princeton, New Jersey: Princeton University Press.

Coval, J. and E. Stafford (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics 86*, 479–512.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Models*. Oxford: Oxford University Press.

Garman, M. B. (1976). Market microstructure. *Journal of Financial Economics 3*, 257–275.

Greenwood, R. (2005). Short- and long-term demand curves for stocks: Theory and evidence on the dynamics of arbitrage. *Journal of Financial Economics 75*, 607–649.

Grossman, S. J. and M. H. Miller (1988). Liquidity and market structure. *Journal of Finance 43*, 617–633.

Hansch, O., N. Y. Naik, and S. V. Viswanathan (1998). Do inventories matter in dealership markets? evidence from the london stock exchange. *Journal of Finance 53*, 1623–1656.

Harris, L. and E. Gurel (1986). Price and volume effects associated with changes in the s&p 500 list: New evidence for the existence of price pressures. *Journal of Finance 41*, 815–829.

Hasbrouck, J. (1988). Trades, quotes, inventories, and information. *Journal of Financial Economics 22*, 229–252.

Hasbrouck, J. and G. Sofianos (1993). The trades of market makers: An empirical analysis of nyse specialists. *Journal of Finance 48*, 1565–1593.

Hendershott, T. and P. C. Moulton (2007). Speed and stock market quality: The nyses hybrid. Manuscript, UC Berkeley.

Hendershott, T. and M. S. Seasholes (2007). Market maker inventories and stock prices. *American Economic Review 97*, 210–214.

Ho, T. and R. Macris (1984). Dealer bid-ask quotes and transaction prices: An empirical study of some amex options. *Journal of Finance 39*, 23–45.

Ho, T. and H. R. Stoll (1981). Optimal dealer pricing under transaction cost and return uncertainty. *Journal of Financial Economics 9*, 47–73.

Huggett, M. (1993). The risk-free rate in heterogeneous agent incomplete insurance economies. *Journal of Economic Dynamics and Control 17*, 953–969.

Koopman, S., N. Shephard, and J. Doornik (1999). Statistical algorithms for models in state space using ssfpack 2.2. *Econometrics Journal 2*, 113–166.

Kraus, A. and H. Stoll (1972). Price impacts of block trading on the new york stock exchange. *Journal of Finance 27*, 569–588.

Lagos, R., G. Rocheteau, and P.-O. Weill (2009). Crashes and recoveries in illiquid markets. Manuscript, UCLA.

Lehmann, B. (1990). Fads, martingales, and market efficiency. *The Quarterly Journal of Economics 105*, 1–28.

Ljungqvist, L. and T. J. Sargent (2004). *Recursive Macroeconomic Theory.* Cambridge Massachusetts: The MIT Press.

Madhavan, A. and S. Smidt (1991). A bayesian model of intraday specialist pricing. *Journal of Financial Economics 30*, 99–134.

Madhavan, A. and S. Smidt (1993). An analysis of changes in specialist inventories and quotations. *Journal of Finance 48*, 1595–1628.

Menkveld, A. J., S. J. Koopman, and A. Lucas (2007). Modelling round-the-clock price discovery for cross-listed stocks using state space methods. *Journal of Business & Economic Statistics 25*, 213–225.

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. Y. and P. K. Yadav (2003). Do dealer firms manage inventory on a stock-by-stock or a portfolio basis? *Journal of Financial Economics 69*, 325–353.

Panayides, M. (2007). Affirmative obligations and market making with inventory. *Journal of Financial Economics 86*, 513–542.

Reiss, P. and I. Werner (1998). Does risk sharing motivate interdealer trading? *Journal of Finance 53*, 1657–1704.

Stoll, H. R. (1978). The supply of dealer services in securities markets. *Journal of Finance 33*, 1133–1151.

Townsend, R. M. (1978). Intermediation with costly bilateral exchange. *Review of Economic Studies 45*, 417–425.

Weill, P.-O. (2007). Leaning against the wind. *Review of Economic Studies 74*, 1329–1354.

Zabel, E. (1981). Competitive price adjustment without market clearing. *Econometrica 49*, 1201–1221.

Table 1: Summary statistics

This table presents summary statistics on the dataset, which combines CRSP, NYSE's Trade and Quotes (TAQ) and a proprietary NYSE dataset. It is a balanced panel that contains daily obersvations on 697 NYSE common stocks from January 1994 through December 2005. Stocks are sorted into quintiles based on market capitalization, where quintile 1 contains large-cap stocks.

variable	description (units)	source	mean Q1	mean Q2	mean Q3	mean Q4	mean Q5	st. dev. within[a]
$midquote_{it}$	closing midquote, div/split adjusted[b] ($)	NYSE	53.76	44.52	36.65	28.58	19.21	22.20
$invent\_shares_{it}$	specialist inventory at the close (1,000 shares)	NYSE	8.19	5.55	4.33	3.23	5.39	34.19
$invent\_dollar_{it}$	specialist inventory at the close[b] ($1,000)	NYSE/CRSP	412.65	168.95	129.48	75.44	77.90	1,383.43
$shares\_outst_{it}$	shares outstanding (million)	CRSP	729.92	157.74	70.08	36.26	18.73	283.75
$market\_cap_{it}$	shares outstanding times price ($billion)	CRSP	34.29	5.34	2.06	0.88	0.29	11.57
$espread_{it}$	share-volume-weighted effective half spread (bps)	TAQ	8.41	12.46	16.50	24.60	46.12	24.20
$dollar\_volume_{it}$	average daily volume ($million)	TAQ	88.21	23.44	10.13	3.63	0.99	42.31
$specialist\_particip_{it}$	specialist participation rate (%)	NYSE	12.31	12.73	14.10	16.58	20.87	8.30
#observations: 697*3,018 (stock*day)								

[a]: Based on the deviations from time means, i.e., $x^*_{it} = x_{it} - \overline{x}_i$.

[b]: We adjust all price series to account for stock splits and dividends.

Table 2: Specialist inventory mean reversion estimates by year and size quintile

This table implements an estimation strategy to capture dynamics between log price changes and specialist inventory. We explicitly recognize that both series contains a common factor γ_t, which we estimate as the cross-sectional mean each day of the standardized series. We refer to this estimate as $\hat{\gamma}_t$. As an example we describe the procedure for specialist inventory. For each stock, each year we multiply specialist inventory by the average price (to get a dollar amount to facilitate comparison across stocks) and then regress it on $\hat{\gamma}_t$. We save the residuals as the idiosyncratic component of specialist inventory in dollar, i.e., I_{it}^{idio}. We perform the following regressions by size quintile (Q1 contains the largest stocks) and by year:

$$y_{it} = \alpha_i + \beta_i x_{it} + \varepsilon_{it}$$

The table reports p-values in brackets. These p-values are based on a test statistic that counts the number of significant t-values across all stock-year ML estimates in the bin. The test statistic is binomially distributed under the null (we use the 0.1 and 0.9 quantiles in the t-test).

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	all
Panel A: Autocorrelation 1st lag coef log price change ($x_{it} = y_{i,t-1}$)													
Q1	−0.03	−0.03	−0.05	−0.08	−0.02	−0.00	−0.01	−0.01	−0.05	−0.00	0.00	−0.02	−0.03
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	(0.003)	(0.001)	(0.000)	(0.005)	(0.005)	(0.026)	(0.000)
Q2	−0.02	−0.02	−0.03	−0.07	−0.01	−0.02	−0.04	−0.00	−0.06	−0.02	−0.02	−0.03	−0.03
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q3	−0.00	0.00	−0.02	−0.06	−0.01	−0.04	−0.06	−0.02	−0.06	−0.04	−0.03	−0.03	−0.03
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Q4	−0.01	−0.02	−0.03	−0.05	−0.00	−0.03	−0.07	−0.04	−0.08	−0.06	−0.08	−0.04	−0.04
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q5	−0.05	−0.04	−0.02	0.00	0.04	0.01	−0.03	0.00	−0.02	−0.01	−0.04	−0.02	−0.02
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
all	−0.02	−0.02	−0.03	−0.05	−0.00	−0.02	−0.04	−0.01	−0.05	−0.03	−0.03	−0.03	−0.03
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel B: Autocorrelation 2nd lag coef log price change ($x_{it} = y_{i,t-2}$)													
Q1	−0.03	−0.05	−0.04	−0.02	−0.03	−0.02	−0.06	−0.05	−0.01	−0.01	−0.02	−0.01	−0.03
	(0.001)	(0.000)	(0.000)	(0.094)	(0.135)	(0.001)	(0.000)	(0.000)	(0.026)	(0.249)	(0.015)	(0.135)	(0.000)
Q2	−0.03	−0.03	−0.03	−0.01	−0.03	−0.01	−0.04	−0.02	−0.01	−0.02	−0.00	−0.02	−0.02
	(0.001)	(0.003)	(0.009)	(0.320)	(0.001)	(0.249)	(0.001)	(0.000)	(0.015)	(0.001)	(0.009)	(0.649)	(0.000)
Q3	−0.02	−0.02	−0.02	−0.01	−0.01	−0.01	−0.04	−0.02	−0.00	−0.00	−0.01	−0.01	−0.02
	(0.584)	(0.070)	(0.045)	(0.664)	(0.584)	(0.029)	(0.000)	(0.017)	(0.010)	(0.029)	(0.029)	(0.001)	(0.000)
Q4	−0.01	−0.02	−0.02	−0.03	−0.02	−0.00	−0.03	−0.02	0.00	−0.01	0.01	−0.00	−0.01
	(0.001)	(0.249)	(0.041)	(0.041)	(0.320)	(0.063)	(0.009)	(0.094)	(0.003)	(0.003)	(0.003)	(0.001)	(0.000)
Q5	−0.01	−0.01	−0.01	−0.00	−0.00	0.00	−0.01	−0.00	−0.01	−0.01	0.00	0.00	−0.01
	(0.029)	(0.000)	(0.017)	(0.017)	(0.000)	(0.010)	(0.500)	(0.000)	(0.017)	(0.009)	(0.087)	(0.070)	(0.000)
all	−0.02	−0.02	−0.02	−0.01	−0.02	−0.01	−0.04	−0.02	−0.00	−0.01	−0.00	−0.01	−0.02
	(0.000)	(0.000)	(0.000)	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel C: Standard deviation of idiosyncratic component specialist inventory I_{it}^{idio}													
Q1	691	968	813	964	1126	1336	1344	1489	1472	1122	1119	1128	1131
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q2	472	510	488	524	530	695	819	647	448	391	441	400	530
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q3	374	429	383	372	430	452	668	437	293	242	266	271	385
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q4	226	254	255	261	291	315	320	333	229	163	145	147	245
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q5	167	159	167	234	204	223	210	186	129	108	95	95	165
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
all	386	464	421	471	516	604	672	619	514	405	413	408	491
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel D: AR coef estimates idiosyncratic component specialist inventory I_{it}^{idio} ($x_{it} = y_{i,t-1}$)													
Q1	0.28	0.27	0.26	0.22	0.25	0.27	0.28	0.29	0.28	0.34	0.37	0.25	0.28
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q2	0.47	0.46	0.44	0.38	0.35	0.34	0.36	0.32	0.25	0.28	0.33	0.25	0.35
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q3	0.59	0.59	0.56	0.51	0.49	0.45	0.41	0.41	0.30	0.31	0.34	0.24	0.43
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q4	0.74	0.73	0.71	0.66	0.63	0.63	0.59	0.57	0.40	0.38	0.36	0.29	0.56
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q5	0.82	0.80	0.80	0.77	0.78	0.79	0.77	0.76	0.66	0.61	0.57	0.51	0.72
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
all	0.58	0.57	0.56	0.51	0.50	0.50	0.48	0.47	0.38	0.38	0.39	0.31	0.47
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

<continued on next page>

<continued from previous page>

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	all
Panel E: Regression coefficient log price change on lagged idiosyncratic component specialist inventory $I_{i,t-1}^{idio}$													
Q1	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.01
	(0.005)	(0.000)	(0.009)	(0.000)	(0.005)	(0.135)	(0.005)	(0.005)	(0.001)	(0.249)	(0.063)	(0.135)	(0.000)
Q2	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.02
	(0.041)	(0.009)	(0.015)	(0.000)	(0.041)	(0.000)	(0.001)	(0.041)	(0.000)	(0.135)	(0.135)	(0.063)	(0.000)
Q3	0.02	0.02	0.02	0.02	0.02	0.03	0.05	0.01	0.05	0.03	0.03	0.03	0.03
	(0.010)	(0.001)	(0.045)	(0.000)	(0.145)	(0.001)	(0.000)	(0.000)	(0.000)	(0.017)	(0.102)	(0.199)	(0.000)
Q4	0.03	0.02	0.04	0.05	0.04	0.06	0.08	0.05	0.11	0.05	0.08	0.06	0.06
	(0.063)	(0.135)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.026)	(0.000)	(0.026)	(0.000)
Q5	0.08	0.08	0.03	0.02	0.06	0.06	0.14	0.10	0.15	0.08	0.12	0.13	0.09
	(0.000)	(0.199)	(0.010)	(0.017)	(0.006)	(0.416)	(0.000)	(0.000)	(0.000)	(0.009)	(0.008)	(0.500)	(0.000)
all	0.03	0.03	0.02	0.02	0.03	0.03	0.06	0.04	0.07	0.04	0.05	0.05	0.04
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)	(0.000)

*/**: Significant at a 95%/99% level.

Table 3: State space model estimates by year and size quintile ($\alpha_i = \kappa_i = 0$)

This table estimates the following state space model for a latent efficient price and an observed end-of-day midquote:

$$\begin{aligned} \text{(observed price)} \quad p_{it} &= m_{it} + s_{it} \\ \text{(unobserved efficient price)} \quad m_{it} &= m_{i,t-1} + \beta_i \hat{\gamma}_t + w_{it} \quad w_{it} = \kappa_i \hat{I}^{idio}_{it} + u_{it} \\ \text{(unobserved transitory price deviation)} \quad s_{it} &= \varphi_i s_{i,t-1} + \alpha_i I^{idio}_{it} + \beta^0_i \hat{\gamma}_t + \cdots + \beta^3_i \hat{\gamma}_{t-3} + \varepsilon_{it} \end{aligned}$$

where i indexes over stocks and t indexes over days, m_{it} is the end-of-day unobserved efficient price ("state"), $\hat{\gamma}_t$ is a midquote return common factor which is the cross-sectional average of the standardized midquote return series which has been filtered with an AR(4) model to remove intertemporal dynamics, p_{it} is end-of-day observed midquote, I^{idio}_{it} is the idiosyncratic part of the specialist end-of-day USD inventory that remains after removing a common factor across specialist inventories (we use the average price across the entire period to convert inventory in shares to USD), $\hat{I}^{idio}_{it}$ is the idiosyncratic inventory innovation which is obtained as the residual of an AR(9) model and captures the surprise part of the net imbalance, β^j_i captures potential "overreaction" or lagged adjustment to common factor innovations, and u_{it} and ν_{it} are mutually independent i.i.d. error terms. We report maximum likelihood estimates where we assume normality for the error terms. We implement our optimization in ox with ssfpack routines where we use the Kalman filter to evaluate the likelihood (see Koopman, Shephard, and Doornik (1999)). The table reports p-values in brackets. These p-values are based on a test statistic that counts the number of significant t-values across all stock-year ML estimates in the bin. The test statistic is binomially distributed under the null (we use the 0.1 and 0.9 quantiles in the t-test).

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	all
Panel A: α_i estimates													
Q1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
all	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: $\sigma(\varepsilon)_i$ estimates													
Q1	48	44	63	54	70	64	102	83	66	35	31	30	58
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q2	48	43	44	49	62	62	87	72	75	44	39	39	55
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q3	43	34	43	46	61	63	94	70	72	46	46	41	55
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q4	51	51	48	57	52	69	96	75	72	50	56	46	60
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q5	69	69	64	53	63	63	103	76	83	51	57	47	67
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
all	52	48	52	52	61	64	96	75	74	45	46	40	59
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel C: φ_i estimates													
Q1	0.46	0.48	0.49	0.32	0.51	0.51	0.55	0.56	0.41	0.39	0.44	0.37	0.46
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q2	0.46	0.41	0.39	0.32	0.46	0.39	0.38	0.54	0.37	0.38	0.40	0.40	0.41
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q3	0.44	0.40	0.34	0.28	0.40	0.30	0.37	0.45	0.35	0.29	0.35	0.37	0.36
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q4	0.37	0.44	0.40	0.33	0.42	0.38	0.37	0.45	0.27	0.28	0.26	0.32	0.36
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q5	0.35	0.40	0.38	0.41	0.48	0.40	0.43	0.43	0.39	0.38	0.33	0.35	0.39
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
all	0.42	0.43	0.40	0.33	0.45	0.40	0.42	0.48	0.36	0.34	0.36	0.36	0.40
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel D: $\sigma(w)_i$ estimates													
Q1	116	117	113	128	161	188	220	171	176	121	100	98	142
Q2	136	131	134	140	174	210	247	197	193	144	121	118	162
Q3	147	142	144	154	198	212	245	202	196	149	132	130	171
Q4	162	151	153	157	201	214	236	207	195	159	140	146	177
Q5	176	173	181	185	234	241	268	254	245	204	178	191	211
all	147	143	145	153	194	213	243	206	201	155	134	137	173

<continued on next page>

<continued from previous page>

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	all
Panel E: transitory volatility													
Q1	64	59	86	65	98	90	145	120	90	46	44	39	79
Q2	64	55	55	59	84	80	107	105	99	57	53	51	72
Q3	56	44	53	55	80	75	114	95	89	56	60	53	69
Q4	64	64	59	67	67	86	115	96	84	58	69	57	74
Q5	84	84	77	67	86	78	133	96	103	64	69	59	83
all	67	61	66	62	83	82	122	103	93	56	59	52	75
Panel F: ratio of transitory and permanent variance													
Q1	0.45	0.43	0.51	0.29	0.44	0.41	0.59	0.52	0.40	0.21	0.30	0.28	0.40
Q2	0.35	0.31	0.26	0.24	0.36	0.26	0.29	0.40	0.38	0.20	0.28	0.33	0.30
Q3	0.32	0.25	0.24	0.25	0.25	0.21	0.33	0.34	0.31	0.16	0.31	0.33	0.27
Q4	0.34	0.35	0.34	0.30	0.25	0.29	0.35	0.35	0.28	0.22	0.33	0.34	0.31
Q5	0.38	0.38	0.36	0.29	0.29	0.24	0.43	0.28	0.32	0.25	0.26	0.26	0.31
all	0.37	0.34	0.34	0.27	0.32	0.28	0.40	0.38	0.34	0.21	0.30	0.31	0.32
Panel G: ratio of transitory and permanent "idiosyncratic" variance													
Q1	0.62	0.51	0.89	0.49	0.85	0.59	0.86	0.95	0.90	0.40	0.46	0.40	0.66
Q2	0.46	0.34	0.39	0.34	0.71	0.32	0.39	0.70	0.89	0.32	0.40	0.49	0.48
Q3	0.41	0.27	0.33	0.32	0.48	0.25	0.43	0.51	0.65	0.25	0.43	0.43	0.40
Q4	0.42	0.36	0.39	0.37	0.36	0.33	0.43	0.51	0.47	0.37	0.48	0.42	0.41
Q5	0.42	0.39	0.39	0.32	0.37	0.25	0.48	0.37	0.44	0.34	0.33	0.32	0.37
all	0.47	0.37	0.48	0.37	0.55	0.35	0.52	0.61	0.67	0.34	0.42	0.41	0.46

Table 4: State space model estimates by year and size quintile ($\varphi_i = 0$)

This table estimates the following state space model for a latent efficient price and an observed end-of-day midquote:

$$\begin{aligned}
&\text{(observed price)} & p_{it} &= m_{it} + s_{it} \\
&\text{(unobserved efficient price)} & m_{it} &= m_{i,t-1} + \beta_i \hat{\gamma}_t + w_{it} \quad w_{it} = \kappa_i \hat{I}_{it}^{idio} + u_{it} \\
&\text{(unobserved transitory price deviation)} & s_{it} &= \varphi_i s_{i,t-1} + \alpha_i I_{it}^{idio} + \beta_i^0 \hat{\gamma}_t + \cdots + \beta_i^3 \hat{\gamma}_{t-3} + \varepsilon_{it}
\end{aligned}$$

where i indexes over stocks and t indexes over days, m_{it} is the end-of-day unobserved efficient price ("state"), $\hat{\gamma}_t$ is a midquote return common factor which is the cross-sectional average of the standardized midquote return series which has been filtered with an AR(4) model to remove intertemporal dynamics, p_{it} is end-of-day observed midquote, I_{it}^{idio} is the idiosyncratic part of the specialist end-of-day USD inventory that remains after removing a common factor across specialist inventories (we use the average price across the entire period to convert inventory in shares to USD), $\hat{I}_{it}^{idio}$ is the idiosyncratic inventory innovation which is obtained as the residual of an AR(9) model and captures the surprise part of the net imbalance, β_i^j captures potential "overreaction" or lagged adjustment to common factor innovations, and u_{it} and ν_{it} are mutually independent i.i.d. error terms. We report maximum likelihood estimates where we assume normality for the error terms. We implement our optimization in ox with ssfpack routines where we use the Kalman filter to evaluate the likelihood (see Koopman, Shephard, and Doornik (1999)). The table reports p-values in brackets. These p-values are based on a test statistic that counts the number of significant t-values across all stock-year ML estimates in the bin. The test statistic is binomially distributed under the null (we use the 0.1 and 0.9 quantiles in the t-test).

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	all
Panel A: α_i estimates													
Q1	−0.03	−0.02	−0.02	−0.02	−0.02	−0.02	−0.02	−0.02	−0.02	−0.01	−0.01	−0.01	−0.02
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q2	−0.06	−0.05	−0.05	−0.04	−0.03	−0.05	−0.05	−0.05	−0.05	−0.03	−0.03	−0.03	−0.04
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Q3	−0.11	−0.09	−0.09	−0.07	−0.09	−0.12	−0.11	−0.06	−0.09	−0.06	−0.07	−0.06	−0.09
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q4	−0.33	−0.25	−0.30	−0.27	−0.26	−0.34	−0.27	−0.23	−0.27	−0.14	−0.19	−0.16	−0.25
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q5	−1.07	−1.04	−0.78	−0.74	−1.01	−1.09	−1.19	−1.30	−1.20	−0.86	−0.94	−0.87	−1.01
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
all	−0.32	−0.29	−0.25	−0.23	−0.28	−0.32	−0.33	−0.33	−0.32	−0.22	−0.25	−0.23	−0.28
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel B: $\sigma(\varepsilon)_i$ estimates													
Q1	20	17	25	30	24	20	31	22	31	18	12	15	22
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q2	19	19	21	28	24	22	43	21	39	23	17	19	25
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q3	19	17	22	23	27	30	45	27	41	25	20	20	26
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q4	28	25	23	26	18	23	42	30	43	30	30	24	29
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q5	42	42	34	25	21	29	47	38	39	28	28	25	33
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
all	26	24	25	27	23	25	42	28	39	25	21	21	27
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel C: φ_i estimates													
Q1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
Q2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
Q3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
Q4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
Q5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
all	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
Panel D: $\sigma(w)_i$ estimates													
Q1	124	124	128	134	176	198	243	192	187	126	106	103	153
Q2	143	136	138	144	186	213	257	212	206	151	127	125	170
Q3	151	143	147	158	204	213	253	212	205	153	138	136	176
Q4	162	152	153	158	201	211	239	212	199	163	145	151	179
Q5	176	170	181	181	228	232	264	245	244	202	178	190	208
all	151	145	149	155	199	213	251	215	208	159	139	141	177

<continued on next page>

<continued from previous page>

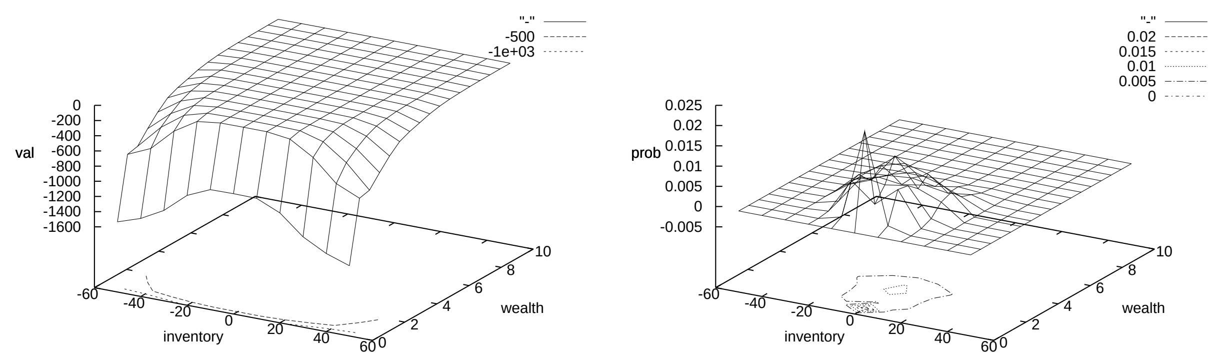
	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	all
Panel E: transitory volatility[a] (σ_i)													
Q1	14	13	13	16	19	22	25	30	22	15	14	9	17
Q2	26	24	22	19	20	33	36	30	23	14	15	13	23
Q3	39	40	35	31	40	48	45	35	23	16	20	17	32
Q4	65	64	69	58	63	86	77	63	41	21	24	21	54
Q5	116	109	115	107	152	177	184	155	108	92	71	54	120
all	52	50	51	46	58	73	73	62	43	31	29	23	49
Panel F: ratio of transitory and permanent variance[a]													
Q1	0.02	0.02	0.01	0.02	0.01	0.02	0.01	0.03	0.02	0.02	0.03	0.01	0.02
Q2	0.08	0.09	0.04	0.03	0.02	0.04	0.03	0.03	0.02	0.01	0.02	0.02	0.03
Q3	0.28	0.27	0.16	0.09	0.13	0.19	0.05	0.07	0.02	0.01	0.04	0.09	0.12
Q4	0.30	0.47	0.47	0.28	0.23	0.47	0.22	0.16	0.06	0.03	0.04	0.05	0.23
Q5	1.11	1.42	1.14	1.16	1.24	2.55	1.96	1.09	0.79	2.10	0.61	0.62	1.32
all	0.36	0.45	0.36	0.31	0.32	0.65	0.46	0.28	0.18	0.44	0.15	0.16	0.34
Panel G: ratio of transitory and permanent "idiosyncratic" variance[a]													
Q1	0.03	0.03	0.02	0.03	0.02	0.02	0.02	0.04	0.03	0.04	0.04	0.01	0.03
Q2	0.10	0.10	0.05	0.04	0.02	0.05	0.04	0.04	0.03	0.02	0.03	0.02	0.04
Q3	0.35	0.28	0.20	0.11	0.18	0.21	0.06	0.10	0.04	0.02	0.05	0.12	0.14
Q4	0.36	0.48	0.54	0.31	0.29	0.51	0.25	0.23	0.08	0.04	0.05	0.05	0.27
Q5	1.28	1.48	1.26	1.28	1.51	2.64	2.16	1.40	0.98	2.31	0.73	0.67	1.47
all	0.42	0.47	0.41	0.35	0.40	0.68	0.51	0.36	0.23	0.48	0.18	0.18	0.39
Panel H: proxy social cost unrealized hedge value ($ million per year)													
Q1	1.59	2.55	3.23	5.91	10.23	17.20	23.82	20.79	14.34	5.53	6.50	6.66	9.86
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q2	1.66	1.66	2.72	2.88	3.51	8.18	9.79	6.96	3.63	1.75	2.13	2.62	3.96
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q3	0.80	1.30	1.22	1.72	2.69	3.16	4.44	3.82	2.02	1.18	1.55	1.86	2.15
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q4	0.52	0.60	0.75	0.98	1.56	2.39	3.11	2.45	1.43	0.70	1.06	0.86	1.37
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q5	0.29	0.39	0.48	0.82	1.05	0.96	1.23	0.87	0.48	0.41	0.55	0.53	0.67
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
all	0.97	1.30	1.68	2.46	3.81	6.38	8.48	6.98	4.38	1.91	2.36	2.51	3.60
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

[a]: The transitory volatility/variance is based on the price pressure term only, i.e., $\alpha^2\sigma^2(I^{idio})$, as this captures the magnitude of price pressures. It therefore excludes the "residual" stationary term that is captured by ε_{it}.

Figure 1: Value function, stationary distribution, and optimal control

This figure illustrates the numerical solution to a dynamic programming problem of an intermediary who maximizes expected utility while quoting bid and ask prices to liquidity demanders whose arrival rates are less than perfectly price elastic. Panel A plots the value function (left) and the stationary distribution (right) over all the inventory-wealth states that the intermediary could find herself in. Panel B illustrates her optimal control by plotting the midquote (left), the bid-ask spread (right), and her consumption (bottom) conditional on her inventory-wealth state. The parametrization of the problem is standard and follows Ho and Stoll (1981). The intermediary has CRRA utility with a coefficient of relative risk aversion equal to 2. She operates at a daily frequency and has a two week horizon which we capture by a (daily) discount factor of 0.9. Without loss of generality, we fix the fundamental price at zero and capture inventory price risk through a stochastic dividend with mean zero and daily standard deviation of $0.05. The probability of a public buyer (seller) arrival depends linearly on the ask (bid) price where the slope θ captures the elasticity of arrival rates to prices (the benchmark is no friction, i.e., $\theta = \infty$).

Panel A: Value function and stationary distribution



Panel B: Optimal controls: midquote, bid-ask spread, and consumption

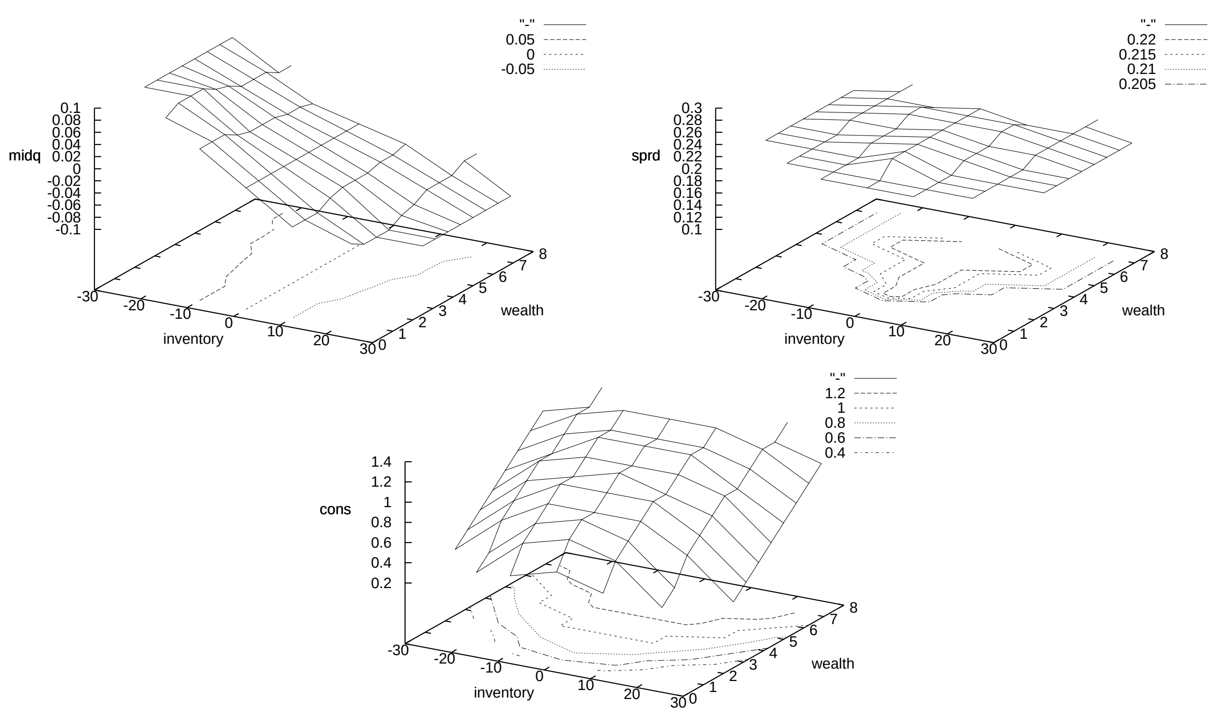
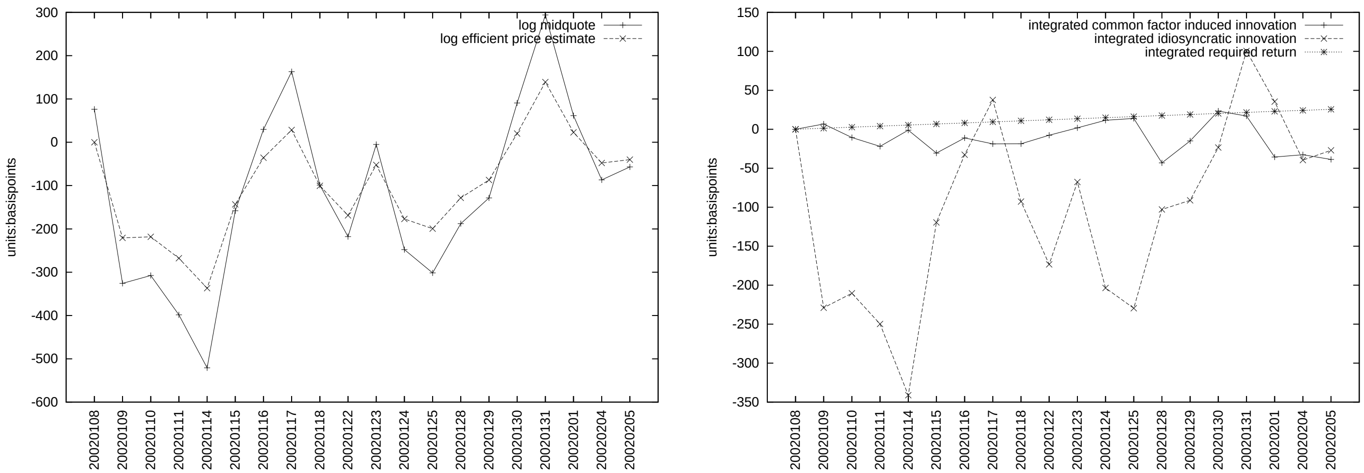


Figure 2: An example of a complete price change decomposition given the state space model estimates ($\alpha_i = 0$)

This figure plots a complete decomposition of the log price series (midquote) of a stock in the lowest-cap size quintile (permno=63503). We start all level series relative to the model's estimate of the efficient price at the first day of the sample.

Panel A: The observed price series (midquote), the (unobserved) efficient price estimate (left) and a decomposition of the efficient price its two stochastic trend components (right)



Panel B: The wedge between the observed price series and the efficient estimate decomposed into (i) lagged adjustment to common factor (left) and (ii) remaining unobserved stationary term (right). We also plot the idiosyncratic inventory position of the intermediary (right).