Internet Appendix for "Does Algorithmic Trading Improve Liquidity?"*

This Internet Appendix contains the following supplementary content:

- Section I considers mechanical explanations for the autoquote results, including stale quotes and slow quote replenishment.
- Section II shows that IV estimates are consistent even if the instrument is a noisy proxy.
- Section III discusses how algorithmic trading (AT) affects the various components of the bid-ask spread based on the spread decomposition of Lin, Sanger, and Booth (1995).
- Section IV proposes a simple generalized Roll model as a framework for interpreting the empirical results.
- Table IA.I provides summary statistics (similar to Table I in the main text) for the five-year sample (monthly from February 2001 through December 2005).

*Citation format: Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2010, Internet Appendix for "Does Algorithmic Trading Improve Liquidity?" *Journal of Finance* 66, 1-33, http://www.afajof.org/supplements.asp.

- Table IA.II provides univariate correlations for the five-year sample between spreads, AT, volume, volatility, and share price.
- Table IA.III investigates the exogeneity of the timing of the autoquote introduction.
- Table IA.IV reports IV regression results using the numerator and the denominator of the AT proxy separately as regressors.
- Table IA.V reports the IV regression results for spreads with share turnover, a potentially endogenous variable, removed from the set of covariates.
- Table IA.VI provides results for the spread decomposition proposed by Lin, Sanger, and Booth (1995).
- Figures IA.1. through IA.4. replicate figures in the main document (Figures 1, 2, 3, and 5, respectively), except that these figures include 95% confidence intervals.
- Figure IA.5. graphs the evolution of the non-spread variables (trade size, number of trades, volume, and volatility) over the five-year sample period.
- Figure IA.6. graphs the three components of the Lin, Sanger, and Booth (1995) spread decomposition over the five-year sample period.

I. Stale Quotes and Slow Quote Replenishment

In the main text, we focus on the AT channel, but it is important to consider whether a more mechanical explanation might account for our autoquote results. What might we expect if autoquote simply makes the observed quotes less stale and has no other effects?

We start by examining what occurs when the inside quote updates are driven by the submission of better quotes or cancellations of the orders at the inside quote. Let a_t and b_t be the ask and bid prices at time t, and assume that this quote is disseminated by the specialist. Limit orders arrive or are cancelled, and at a later time t', $a_{t'}$ and $b_{t'}$ are the best ask and bid prices. Assume that $a_{t'}$ and $b_{t'}$ are disseminated only after the adoption of autoquote; otherwise, the econometrician identifies a_t and b_t as the ask and bid in effect at time t'.

To simplify the exposition, assume that the ask side of the book changes ($a_t \neq a_{t'}$) while the bid side of the book remains unchanged ($b_t = b_{t'}$). Symmetric arguments apply for changes to the bid side of the book alone, and the results also hold when both the bid and the ask change between t and t'.

There are two possibilities for the change in the inside ask. If the time t inside ask is cancelled, then $a_{t'} > a_t$. If instead a new sell order arrives at time t' that would improve the inside quote, then $a_{t'} < a_t$. Overall, if cancels are more common than improvements, then prior to the adoption of autoquote the disseminated quoted spread is artificially narrow, and autoquote should be associated with a widening of quoted spreads. However, we find the reverse. Autoquote is associated with a narrowing of the quoted spread, so we focus hereafter on the arrival of new orders at time t' that improve the existing time t quote. Prior to autoquote, we continue to observe the old, wider quote (a_t, b_t) at time t'. Under autoquote, the new, narrower quote $(a_{t'}, b_t)$ is disseminated at time t'.

Let $m_{t'} = 1/2(a_{t'} + b_{t'})$ be the midquote at time t'. Under autoquote, we see the true state of the order book, and if a trade at time t' occurs at price $p_{t'}$ (at either the bid price $b_{t'}$ or the ask price $a_{t'}$), assume that the effective half-spread $s_{t'} = q_{t'}(p_{t'} - m_{t'})$ is correctly measured. In contrast, before the adoption of autoquote the observed midquote at time t' is $m_t = 1/2(a_t + b_t)$, which is stale. Because we focus on the arrival of a sell order that improves the ask, $m_{t'} < m_t$, which means that in the absence of autoquote the observed quote midpoint is biased upwards. Define the measured effective spread pre-autoquote as $s_{t',pre} = q_{t'}(p_{t'} - m_t)$.

Based on the above discussion, the change in the measured effective spread under autoquote is the difference $s_{t'} - s_{t',pre} = q_{t'}\ (m_t - m_{t'}) = q_{t'}\ (a_t - a_{t'})/2$. The term in parentheses is positive, since the arriving sell order improves the quote by lowering the ask price, so the effective spread declines under autoquote if and only if $\mathrm{E}(q_{t'}) < 0$. But this cannot be the case as long as the demand for immediacy is downward sloping in the price of immediacy. To say it another way, a better ask price should on average draw in a marketable buy order, which implies $\mathrm{E}(q_{t'}) > 0$. Thus, if autoquote is simply displaying quotes that were previously undisseminated, the result should be a widening of the effective spread under autoquote.

Note that the above analysis implicitly assumes that without autoquote, the difference between the true midquote $m_{t'}$ and the disseminated midquote m_t does not affect $q_{t'}$, the sign of the trade. The trade sign can be affected, however, if the new ask price $a_{t'}$ is below the disseminated midquote m_t. In this case both the true ask and bid prices are below the disseminated midquote, and with the right choice of parameter values effective spreads could be mechanically narrower under autoquote. But, this scenario seems unlikely to dominate. First, it is quite likely that the specialist would disseminate an updated quote if an incoming limit order crosses the midquote in this way, as the new quoted spread would be less than half as wide as the old quoted spread. Second, if this scenario were empirically important, the resulting trade-signing errors would bias downward the pre-autoquote estimates of the adverse selection component of the spread, because future price changes would be less correlated with trade signs. In this scenario, we would expect to see an *increase* in adverse selection with the elimination of stale quotes under autoquote. This is the opposite of our findings in Tables III and V in the main text.

Our argument above makes use of the observed decline in adverse selection post-autoquote. If this decline is an artifact of measurement error, our argument is weakened. In addition, the reduction in adverse selection associated with autoquote is quite striking. Thus, it is worth considering a mechanical explanation for the observed changes in adverse selection.[1]

Recall that in order to measure adverse selection, we use quotes five minutes or 30 minutes after the trade. In the VAR approach, we use the next 10 trades to calculate the permanent price impact of a unit shock to signed order flow. If it takes longer than this to replenish the quotes after a trade exhausts the depth at the inside, our estimates of adverse selection would be biased upward. AT replenishes quotes more rapidly, removing this upward bias, and making it appear that adverse selection is declining in AT. However, our 30-minute results are virtually identical to our five-minute results, implying that there is little quote replenishment during that 25-minute interval. Thus, while we think changes in quote replenishment are unlikely to drive the adverse selection results, we cannot rule out the possibility.

To summarize, neither a mechanical increase in quote disseminations nor faster quote replenishment is likely to be the source of our results.

II. Instrumental Variable Regression with a Noisy Proxy for AT

As we discuss in the text, suppose we begin with a linear relationship between liquidity L_{it} and AT A_{it}:

$$L_{it} = \alpha_i + \beta A_{it} + \delta' X_{it} + \varepsilon_{1it}, \tag{IA.1}$$

where X_{it} is a vector of control variables. The usual full-rank conditions apply, and $E(X_{it}\varepsilon_{1it}) = 0$, but $cov(A_{it}, \varepsilon_{1it}) \neq 0$ because A_{it} also depends on L_{it}:

$$A_{it} = \omega_i + \theta L_{it} + \phi' X_{it}. \tag{IA.2}$$

Furthermore, the observed proxy for AT A_{it} measures AT with error,

$$A^o_{it} = A_{it} + \varepsilon_{2it}, \tag{IA.3}$$

so that

$$A^o_{it} = \omega_i + \theta L_{it} + \phi' X_{it} + \varepsilon_{2it}. \tag{IA.4}$$

Suppose there exists an instrument Z_{it} such that $cov(Z_{it}, A_{it}) \neq 0$, $cov(Z_{it}, \varepsilon_{1it}) = 0$, $cov(Z_{it}, \varepsilon_{1it}) = 0$, and $var(\varepsilon_Z) > 0$, where ε_Z is the residual of a regression of Z_{it} on X_{it}. We rewrite equation (IA.1) as

$$L = W\xi + \varepsilon_1, \tag{IA.5}$$

where we stack all equations indexed by it into vectors and matrices so that the subscripts disappear:

$W = [1 \quad A^o \quad X]$, $\xi' = [\alpha' \quad \beta \quad \delta']$, and $\tilde{Z} = [1 \quad Z \quad X]$, and 1 is a dummy matrix to match the stock-specific fixed effects. Now pre-multiply by $n^{-1}\tilde{Z}'$:

$$n^{-1}\tilde{Z}'L = n^{-1}\tilde{Z}'W\xi + n^{-1}\tilde{Z}'\varepsilon_1. \tag{IA.6}$$

By assumption, plim $n^{-1}\tilde{Z}'\varepsilon_1 = 0$, so a consistent estimate is

$$\hat{\xi} = (\tilde{Z}'W)^{-1}\tilde{Z}'L. \tag{IA.7}$$

This is well-defined, since the $[Z \quad X]$ matrix is of full rank, and $cov(Z_{it}, A^o_{it}) \neq 0$ because we assumed that the instrument is correlated with the desired endogenous variable ($cov(Z_{it}, A_{it}) \neq 0$). So the consistency of the IV estimator is unaffected by using a noisy proxy for AT.

III. Lin-Sanger-Booth (1995) Spread Decomposition

The decomposition of the effective spread introduced in equations (2) and (3) in the main text has the advantage of being simple, but it also has distinct disadvantages. In particular, it chooses an arbitrary point in time in the future (five minutes or 30 minutes in this case) and implicitly ignores other trades that might have happened in that time period. Lin, Sanger, and Booth (LSB (1995)) develop a spread decomposition model that is estimated trade by trade and accounts for order flow persistence (the empirical fact, first noted by Hasbrouck and Ho (1987), that buyer-initiated trades tend to follow buyer-initiated trades).[2] Let

$$\delta = Prob[q_{t+1} = 1 | q_t = 1] = Prob[q_{t+1} = -1 | q_t = -1] \tag{IA.8}$$

be the probability of a continuation (a buy followed by a buy or a sell followed by a sell). Further suppose that the change in the market maker's quote midpoint following a trade is given by

$$m_{t+1} - m_t = \lambda_t q_t. \tag{IA.9}$$

The dollar effective half-spread is $s_t = q_t(p_t - m_t)$, which is assumed to be constant for simplicity. If there is persistence in order flow, the expected transaction price at time $t+1$ does not equal m_{t+1} but instead is

$$\begin{aligned} E_t(p_{t+1}) &= \delta(m_t + q_t(\lambda_t + s_t)) + (1-\delta)(m_t + q_t(\lambda_t - s_t) \\ &= m_t + q_t(\lambda_t + (2\delta - 1)s_t). \end{aligned} \tag{IA.10}$$

This expression shows how far prices are expected to permanently move against the market-maker. While the market maker earns s_t initially, in expectation he loses $\lambda_t + (2\delta - 1)s_t$ due to adverse selection and order persistence, respectively. Note that this reduces to Glosten (1987) if $\delta = 0.5$ so that order flow is independent over time. We can identify the adverse selection component λ by regressing midpoint changes on the buy-sell indicator, and we can identify the order persistence parameter with a first-order autoregression on q_t. The remaining portion of the effective spread is revenue for the market maker, referred to by LSB as the fixed component of the spread. Thus, spreads are decomposed into three separate components: a fixed component associated with temporary price changes, an adverse selection component, and a component due to order flow persistence. The fixed, temporary component continues to reflect the net revenues to liquidity suppliers after accounting for losses to (the now persistent) liquidity demanders. The adverse selection component captures the immediate gross losses to the current liquidity demander, while the order flow persistence component captures the expected gross losses to those demanding liquidity in the same direction in the near future. We estimate the model and calculate components of the effective spread for each sample stock each day.

For each of the market-cap quintiles, the three panels of Figure IA.6. show how the three LSB spread components evolve over the whole 2001 to 2005 sample period. There are no consistent trends in the fixed component: around the implementation of autoquote, there is an increase for the smallest quintile, but this increase does not extend to the other quintiles. In contrast, the adverse selection component falls sharply during the implementation of autoquote in the first half of 2003. This is true across all five quintiles, and the change appears to be permanent. Beginning in the second half of 2002 and continuing to the end of 2005, there is also a steady decline in the order persistence component of the spread. This suggests less persistence, which could indicate that over this period algorithms and human traders both become more adept at concealing their order flow patterns, perhaps by using mixed order submission strategies that sometimes demand liquidity and sometimes supply it.

The staggered introduction of autoquote allows us to take out all market-wide effects and focus on cross-sectional differences between the stocks that implement autoquote early versus the stocks that implement autoquote later on. As we did for the simpler decomposition, we can put any one of the LSB spread components on the left-hand side of our IV specification to determine the sources of the liquidity improvement when there is more AT. The results are in Panel B of Table IA.VI and are quite consistent with the earlier decomposition. For the largest two quintiles, autoquote (and the resulting increases in AT) is associated with an increase in the fixed component of the spread, and a decrease in the adverse selection component and the order persistence component. The drop in the adverse selection component is economically quite large. During the autoquote sample period, the within standard deviation in our AT variable is 4.54, so a one-standard deviation increase in AT during this sample period leads to an estimated change in the adverse selection component equal to $4.54 * -0.26$, or about a 1.2 basis point narrowing of the adverse selection component. This is quite substantial, given that the adverse selection component for the biggest quintile is only about 2 basis points on average out of an overall 3.62 basis point effective half-spread. The coefficients on the other two components are of similar magnitude, indicating similar economic importance. As in the earlier decomposition, there are no significant effects for the smaller-cap quintiles.

IV. A Generalized Roll Model

To further explore our counterintuitive results, particularly the increase in realized spreads caused by AT, here we develop a generalized Roll model that is a slight variation of the one developed in Hasbrouck (2007). Though the model is quite simple, it provides a useful framework for thinking about AT and delivers a number of empirical predictions, all of which match our empirical results.

A. The Model without AT

The "game" has two periods, each with an i.i.d. innovation in the efficient price:

$$m_t = m_{t-1} + w_t, \tag{IA.11}$$

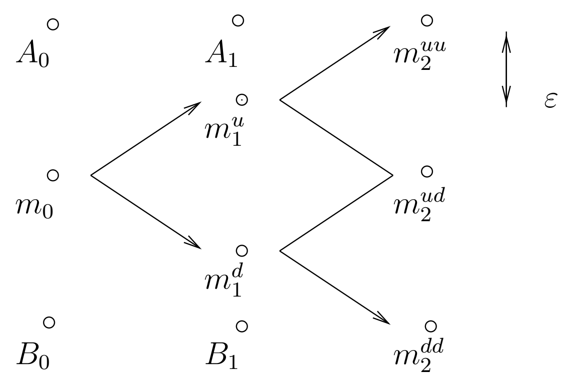
where $w_t \in \{\epsilon, -\epsilon\}$, each with probability 0.5. The game features three stages:

- At t = 0, risk-neutral humans can submit a bid and ask quote and, given full competition, the first one arriving bids her reservation price.

- At t = 1, humans can observe w_1 at cost c. If humans choose to buy this information, they can submit a new limit order.[3]

- At t = 2, two informed liquidity demanders arrive, one with a positive private value associated with a trade, $+\theta$, the other with a negative private value, -θ.

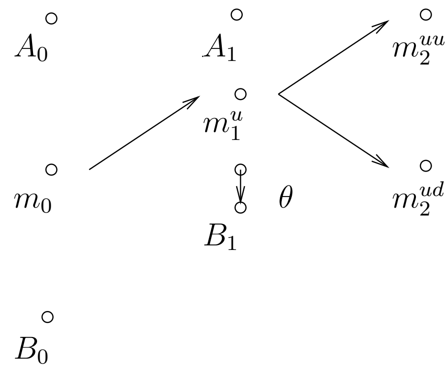
We assume that $2c > \theta$, that is, the cost of "observing" information for humans is sufficiently high that they do not update their quotes. The technical assumption $\epsilon > \theta$ ensures that trade occurs at $t = 2$ if and only if the efficient price changes between $t = 0$ and $t = 2$, and that only one of the two arriving liquidity demanders transacts in that case.



There are four equally likely paths through the binomial tree: *uu*, *ud*, *du*, and *dd*, where u represents a positive increment of ϵ to the fundamental value and d a negative increment. In equilibrium, humans do not buy the information w_1 and update the quote at $t = 1$, because they have to quote so far away from the efficient price to make up for c that neither liquidity demander will transact at that quote as $2c > \theta$. Given that they do not acquire the information w_1, humans protect themselves by setting the bid price equal to $m_0 - 2\epsilon$ and the ask price equal to $m_0 + 2\epsilon$. One of the liquidity demanders trades at $t = 2$ if the path is either *uu* or *dd*; the quote providers break even. If the path is *ud* or *du*, then there is no trade, because the liquidity demander's private value is too small relative to the spread.

Clearly, under these assumptions all price changes are associated with order flow, and there is no public information component.

B. The model with AT



Now we introduce an algorithm that can buy the information w_1 at zero cost ($c = 0$). The results at $t = 0$ remain unchanged. At $t = 1$, the algorithm optimally issues a new quote. To illustrate the idea, suppose $w_1 > 0$. The algorithm knows that it is the only liquidity provider in possession of w_1, and so it puts in a new bid equal to $m_0 - \theta$. If $w_2 > 0$ as well, then a transaction takes place at the original ask of $m_0 + 2\epsilon$. If $w_2 < 0$, then a liquidity demander will hit the algorithm's bid. This bid is below the efficient price, so there will eventually be a reversal, and there is a temporary component in prices. Conversely, if $w_1 < 0$, the algorithm places a new ask at $m_0 + \theta$, which is traded with if it turns out that $w_2 > 0$.

In the presence of AT, part of the change in the efficient price is revealed through a quote update without trade. Public information now accounts for a portion of price discovery, and imputed revenue to liquidity suppliers is now positive. Thus, the model can explain even the surprising empirical findings on realized spreads and trade-correlated price moves. The model also delivers narrower quoted spreads and more frequent trades, both of which are also observed in the data.

To deliver an increase in realized spread, it is important in the model that competition between algorithms be less vigorous than the competition between humans. This seems plausible in reality as well. As autoquote was implemented in 2003, the extant algorithms might have found themselves with a distinct competitive advantage in trading in response to the increased information flow, given that new algorithms take considerable time to build and test.

Notes

[1]We thank an anonymous referee for suggesting this alternative.

[2]See Barclay and Hendershott (2004) for discussion of how the Lin, Sanger, and Booth (1995) spread decomposition relates to other spread decomposition models.

[3]Periods here are on the order of seconds, and the information is best thought of as information contained in order flow and prices, rather than as a direct signal about future cash flows.

References

Arellano, M., and S.R. Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58:277–297.

Barclay, M.J., and T. Hendershott. 2004. "Liquidity Externalities and Adverse Selection: Evidence from Trading After Hours." *Journal of Finance* 59:681–710.

Glosten, L.R. 1987. "Components of the Bid Ask Spread and the Statistical Properties of Transaction Prices." *Journal of Finance* 42:1293–1307.

Hasbrouck, J. 2007. *Empirical Market Microstructure*. New York: Oxford University Press.

Hasbrouck, J., and Thomas Ho. 1987. "Order Arrival, Quote Behavior and the Return Generating Process." *Journal of Finance* 42:1035–1048.

Lin, J.C., G. Sanger, and G.G. Booth. 1995. "Trade Size and Components of the Bid-Ask Spread." *Review of Financial Studies* 8:1153–1183.

Table IA.I
Summary Statistics for Five-year Sample

This table presents summary statistics for the five-year data set that merges TAQ, CRSP, and NYSE System Order Data (cf. Table 1 in the main text that is based on the autoquote daily sample). The balanced panel consists of monthly data on 943 stocks from February 2001 through December 2005. Stocks are sorted into quintiles based on market capitalization, where quintile 1 contains large-cap stocks. All variables are 99.9% winsorized. The within standard deviation is based on day t's deviation relative to the time mean, that is, $x^*_{i,t} = x_{i,t} - \overline{x}_i$.

Variable	Description (units)	Source	Mean Q1	Mean Q2	Mean Q3	Mean Q4	Mean Q5	St. dev. within
$qspread_{it}$	share-volume-weighted quoted half spread (bps)	TAQ	5.31	7.33	9.47	12.92	22.44	8.40
$qdepth_{it}$	share-volume-weighted depth ($1,000)	TAQ	92.37	52.93	38.62	28.69	19.43	21.88
$espread_{it}$	share-volume-weighted effective half spread (bps)	TAQ	3.67	5.19	6.79	9.40	16.16	6.42
$rspread_{it}$	share-volume-weighted realized half spread, 5min (bps)	TAQ	0.96	1.24	1.56	2.19	4.95	2.82
$adv\_selection_{it}$	share-volume-weighted adverse selection component half spread, 5min, "effective-realized" (bps)	TAQ	2.71	3.96	5.23	7.22	11.21	5.02
$messages_{it}$	#electronic messages per minute i.e. proxy for algorithmic activity (/minute)	NYSE	131.99	71.70	43.46	28.86	15.84	43.79
$algo\_trad_{it}$	dollar volume per electronic message times (-1) to proxy for AT ($100)	TAQ/NYSE	-26.34	-15.22	-10.88	-8.38	-5.95	11.20
$dollar\_volume_{it}$	average daily volume ($million)	TAQ	112.13	31.70	13.85	7.03	2.82	23.18
$trades_{it}$	#trades per minute (/minute)	TAQ	5.84	3.19	2.02	1.43	0.80	1.28
$share\_turnover_{it}$	(annualized) share turnover	TAQ/CRSP	1.02	1.48	1.46	1.44	1.22	0.69
$volatility_{it}$	standard deviation daily midquote returns (%)	CRSP	1.75	1.95	1.96	2.16	2.54	1.01
$price_{it}$	daily closing price ($)	CRSP	45.90	38.60	33.09	27.98	20.62	9.53
$market\_cap_{it}$	shares outstanding times price ($billion)	CRSP	36.75	5.48	2.30	1.17	0.53	5.09
$trade\_size_{it}$	trade size ($1,000)	TAQ	46.52	24.95	16.97	12.25	8.32	11.52
$specialist\_particip_{it}$	specialist participation rate (%)	NYSE	12.42	12.19	12.28	13.16	15.15	4.42
#observations: 943*59 (stock*month)								

Table IA.II
Overall, Between, and Within Correlations for Five-year Sample

This table presents overall, between, and within correlations for some variables in the monthly sample that extends from February 2001 through December 2005. Table IA.I provides variable definitions. The between standard deviation is based on the time means, that is, $\overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{i,t}$. The within standard deviation is based on day t's deviation relative to the time mean, that is, $x^*_{i,t} = x_{i,t} - \overline{x}_i$. * denotes significance at the 95% level.

		$messa-ges_{it}$	$algo\_trad_{it}$	$share\_turnover_{it}$	$vola-tility_{it}$	$1/price_{it}$	$ln\_mar-ket\_cap_{it}$
$qspread_{it}$	ρ(overall)	-0.43*	0.10*	-0.14*	0.54*	0.74*	-0.57*
	ρ(between)	-0.51*	0.51*	-0.09*	0.65*	0.83*	-0.68*
	ρ(within)	-0.33*	-0.23*	-0.20*	0.48*	0.63*	-0.59*
$messages_{it}$	ρ(overall)		-0.08*	0.13*	-0.20*	-0.24*	0.72*
	ρ(between)		-0.87*	0.08*	-0.17*	-0.32*	0.90*
	ρ(within)		0.63*	0.19*	-0.24*	-0.13*	0.43*
$algo\_trad_{it}$	ρ(overall)			-0.12*	-0.12*	0.24*	-0.52*
	ρ(between)			-0.11*	0.19*	0.36*	-0.86*
	ρ(within)			-0.14*	-0.28*	0.12*	0.02*
$share\_turnover_{it}$	ρ(overall)				0.35*	-0.07*	-0.07*
	ρ(between)				0.44*	-0.03*	-0.13*
	ρ(within)				0.31*	-0.12*	0.15*
$volatility_{it}$	ρ(overall)					0.47*	-0.29*
	ρ(between)					0.72*	-0.41*
	ρ(within)					0.30*	-0.33*
$1/price_{it}$	ρ(overall)						-0.44*
	ρ(between)						-0.45*
	ρ(within)						-0.66*

*: Significant at a 95% level.

Table IA.III
Effective Spread Forecast Minus Its Long-term Mean on Autoquote Introduction Day

This table forecasts effective spread on the autoquote introduction day based on the pre-introduction period in order to analyze whether introductions coincide with temporarily wide spreads. One time-series regression estimates a univariate AR(1) model. The second specification also includes lagged values of variables that correlate with liquidity – share turnover, volatility, the inverse of price, and log market cap (see the control variables in Table III of the main text):

$$L_{it} = \alpha_i + \hat{\gamma}_t + \beta_i L_{i,t-1} + \delta_i X_{i,t-1} + \varepsilon_{it}, \quad t \in [2, \ldots, \tau_i - 1],$$

where L_{it} is the effective half-spread for stock i on day t, X_{it} is a vector of predictor variables (i.e., share turnover, volatility, 1/price, and log market cap), α_i is the stock-specific mean, $\hat{\gamma}$ is the cross-sectional average for each day t, and τ_i is the autoquote introduction day for stock i. Regressions are estimated stock by stock. Panels A and B report the results for a univariate AR(1) model (i.e., setting δ_i to zero). Panel A reports the AR(1) parameter (β_i) estimates and their standard errors both by quintile and overall. Panel B reports the out-of-sample liquidity forecast on the autoquote introduction day, that is, the forecast based on all days up until the last day before the introduction:

$$f_{\tau_i} = \hat{\beta}_i L_{\tau_i - 1} - \hat{\alpha}_i,$$

where the hats indicate estimates based on the pre-introduction period. Panel C replicates Panel B, but includes the control variables in the estimation and in the forecast. */** denote significance at the 95%/99% level.

Q1	Q2	Q3	Q4	Q5	all
Panel A: AR(1) coefficient estimates (β_i)					
0.186	0.199	0.153	0.181	0.180	0.180
(0.152)	(0.146)	(0.137)	(0.148)	(0.140)	(0.065)
Panel B: Forecast minus long-term mean, AR(1)					
0.018	-0.002	-0.043	-0.038	-0.036	-0.020
(0.224)	(0.356)	(0.686)	(1.152)	(1.963)	(0.045)
Panel C: Forecast minus long-term mean, AR(1) + controls					
0.032	0.037	-0.060	-0.000	-0.070	-0.012
(0.505)	(0.767)	(1.101)	(1.585)	(2.962)	(0.101)

Table IA.IV
Effect of AT on Spread: Results for Numerator and Denominator of AT Proxy

This table separately regresses the effective half-spread on the numerator and the denominator of the AT proxy. The regression is based on daily observations in the period from December 2002 through July 2003, covering the phase-in of autoquote. The nonsynchronous autoquote introduction instruments for the endogenous $algo\_trad_{it}$, its denominator, and its numerator. The specification is (see Table III in main text):

$$L_{it} = \alpha_i + \gamma_t + \beta A_{it} + \delta X_{it} + \varepsilon_{it},$$

where L_{it} is a spread measure for stock i on day t, A_{it} is the AT measure $algo\_trad_{it}$, its denominator, or its numerator, and X_{it} is a vector of control variables, including share turnover, volatility, 1/price, and log market cap. Coefficients for the control variables and time dummies are quintile-specific. Market cap-weighted coefficients are reported for the control variables. Fixed effects and time dummies are included. The set of instruments consists of all explanatory variables, except that A_{it} is replaced with $auto\_quote_{it}$. There are separate regressions for each size quintile, and t-values in parentheses are based on standard errors that are robust to general cross-section and time-series heteroskedasticity and within-group autocorrelation (see Arellano and Bond (1991)). */** denote significance at the 95%/99% level.

	Coefficient on A_{it}					Coefficients on control variables			
	Q1	Q2	Q3	Q4	Q5	$share\_turnover_{it}$	$vola-tility_{it}$	$1/price_{it}$	$ln\_mkt\_cap_{it}$
Panel A: $A_{it} = algo\_trad_{it} = -\frac{dollar\_volume_{it}}{messages_{it}}$ (cf. Table 3 main text)									
$espread_{it}$	-0.18**	-0.32**	-0.35	-1.67	4.65	-1.01**	0.69**	72.77**	-1.30
	(-2.67)	(-2.23)	(-1.56)	(-0.42)	(1.16)	(-2.30)	(9.39)	(10.80)	(-1.46)
Panel B: $A_{it} = messages_{it}$									
$espread_{it}$	-0.01**	-0.10**	-0.10**	-0.21	1.68	0.38**	0.76**	68.32**	0.95**
	(-3.70)	(-2.05)	(-2.24)	(-0.87)	(1.49)	(5.11)	(17.46)	(13.37)	(2.77)
Panel C: $A_{it} = dollar\_volume_{it}$									
$espread_{it}$	-0.05**	206.98	0.87	0.35	9.86	-350.29	-1.16	-535.11	-679.84
	(-2.15)	(0.00)	(1.42)	(0.77)	(0.80)	(-0.00)	(-0.00)	(-0.00)	(-0.00)
#observations: 1082*167 (stock*day)									

Table IA.V
Effect of AT on Spread: Turnover Removed as Covariate

This table regresses various measures of the (half) spread on our AT proxy. It mirrors Table 3 in the main text where the only difference is that share turnover is removed due to an endogeneity concern. It is based on daily observations from December 2002 through July 2003 which covers the phase-in of autoquote. The nonsynchronous autoquote introduction instruments for the endogenous $algo\_trad_{it}$ to identify causality from AT to liquidity. The specification is:

$$L_{it} = \alpha_i + \gamma_t + \beta A_{it} + \delta X_{it} + \varepsilon_{it}$$

where L_{it} is a spread measure for stock i on day t, A_{it} is the AT measure $algo\_trad_{it}$, and X_{it} is a vector of control variables, including volatility, 1/price, and log market cap. Market cap-weighted coefficients are reported for the control variables. Fixed effects and time dummies are included. The set of instruments consists of all explanatory variables, except that $algo\_trad_{it}$ is replaced with $auto\_quote_{it}$. There are separate regressions for each size quintile, and t-values in parentheses are based on standard errors that are robust to general cross-section and time-series heteroskedasticity and within-group autocorrelation (see Arellano and Bond (1991)). */** denote significance at the 95%/99% level. In Panel B the suffix indicates the effective spread for a particular trade size category: "1" if 100 shares $\leq$ trade size $\leq$ 499 shares; "2" if 500 shares $\leq$ trade size $\leq$ 1999 shares; "3" if 2000 shares $\leq$ trade size $\leq$ 4999 shares; "4" if 5000 shares $\leq$ trade size $\leq$ 9999 shares; "5" if 9999 shares $<$ trade size.

	Coefficient on $algo\_trad_{it}$					Coefficients on control variables		
	Q1	Q2	Q3	Q4	Q5	$vola\text{–}tility_{it}$	$1/price_{it}$	$ln\_mkt\_cap_{it}$
Panel A: Quoted spread, quoted depth, and effective spread								
$qspread_{it}$	-0.67**	-0.29**	-0.43	-0.11	4.52**	-0.93	157.19**	-1.34
	(-2.16)	(-2.19)	(-1.16)	(-0.05)	(2.04)	(-1.13)	(4.24)	(-0.65)
$qdepth_{it}$	-3.85**	-1.65**	-1.99	10.09	-0.61	-5.10	104.21	16.66
	(-2.11)	(-1.98)	(-1.06)	(0.32)	(-0.47)	(-1.04)	(0.51)	(1.38)
$espread_{it}$	-0.23**	-0.22**	-0.36	-1.13	2.18*	0.05	88.58**	-0.51
	(-1.99)	(-2.23)	(-1.28)	(-0.37)	(1.84)	(0.15)	(6.36)	(-0.57)
Panel B: Effective spread by trade size category								
$espread1_{it}$	-0.16**	-0.08*	-0.17	-1.23	2.32**	-0.27	63.80**	-0.48
	(-2.07)	(-1.82)	(-0.90)	(-0.36)	(1.95)	(-1.28)	(6.69)	(-0.67)
$espread2_{it}$	-0.29**	-0.19**	-0.42	-2.87	2.09*	-0.60	78.50**	-0.96
	(-2.13)	(-2.41)	(-1.22)	(-0.39)	(1.85)	(-1.58)	(4.80)	(-0.67)
$espread3_{it}$	-0.33*	-0.15	-0.67	-1.83	3.22**	-0.55	91.14**	-0.43
	(-1.94)	(-1.19)	(-1.20)	(-0.34)	(2.06)	(-1.19)	(4.57)	(-0.32)
$espread4_{it}$	-0.16	-0.17	-0.24	-6.52	-0.02	0.01	78.32**	-1.10
	(-1.19)	(-0.76)	(-0.58)	(-0.22)	(-0.01)	(0.01)	(5.31)	(-0.23)
$espread5_{it}$	-0.05	-0.21	-0.37	-1.10	1.90	0.85**	77.01**	-0.36
	(-0.42)	(-0.94)	(-0.79)	(-0.17)	(0.82)	(2.37)	(5.58)	(-0.30)

<continued on next page>

<continued from previous page>

	Coefficient on $algo\_trad_{it}$					Coefficients on control variables		
	Q1	Q2	Q3	Q4	Q5	$vola-tility_{it}$	$1/price_{it}$	$ln\_mkt\_cap_{it}$
Panel C: Spread decompositions based on 5-min and 30-min price impact								
$rspread_{it}$	0.45**	0.53**	1.04*	9.88	6.86**	0.80	5.96	2.36
	(2.33)	(3.59)	(1.68)	(0.42)	(2.50)	(1.21)	(0.26)	(0.62)
$adv\_selection_{it}$	-0.67**	-0.75**	-1.40*	-10.72	-4.69**	-0.74	82.21**	-2.83
	(-2.34)	(-3.75)	(-1.64)	(-0.42)	(-2.21)	(-0.84)	(2.39)	(-0.65)
$rspread\_30m_{it}$	0.42**	0.34**	0.92	7.72	5.46**	-0.71	16.05	0.51
	(2.12)	(2.01)	(1.45)	(0.42)	(2.15)	(-1.15)	(0.71)	(0.17)
$adv\_selection\_30m_{it}$	-0.64**	-0.57**	-1.28	-8.72	-3.42*	0.79	71.48**	-1.01
	(-2.33)	(-2.75)	(-1.53)	(-0.42)	(-1.78)	(0.97)	(2.23)	(-0.27)

#observations: 1082*167 (stock*day)
F test statistic of hypothesis that instruments do not enter first stage regression: 7.32 ($F(5, 179587)$), p-value: 0.0000

Table IA.VI
Effect of AT on Spread: Lin, Sanger, and Booth (1995) Spread Decomposition

This table regresses various components of the effective half spread on the AT proxy. It uses the spread decomposition model of Lin, Sanger, and Booth (1995) (LSB) which accounts for order persistence. The LSB model identifies a fixed (transitory) component ($LSB95\_fixed_{it}$), an adverse selection component ($LSB95\_adv\_sel_{it}$), and a component due to order persistence ($LSB95\_order\_persist_{it}$) (see Section I for details). The regressions are based on daily observations in the period from December 2002 through July 2003, which covers the phase-in of autoquote. The nonsynchronous autoquote introduction instruments for the endogenous $algo\_trad_{it}$ to identify causality from these explanatory variables to liquidity. The specification is (see Table III in main text)

$$L_{it} = \alpha_i + \gamma_t + \beta A_{it} + \delta X_{it} + \varepsilon_{it}$$

where L_{it} is a spread measure for stock i on day t, A_{it} is the AT measure $algo\_trad_{it}$, and X_{it} is a vector of control variables, including share turnover, volatility, 1/price, and log market cap. Market cap-weighted coefficients are reported for the control variables. Fixed effects and time dummies are included. The set of instruments consists of all explanatory variables, except that $algo\_trad_{it}$ is replaced with $auto\_quote_{it}$. There are separate regressions for each size quintile, and t-values in parentheses are based on standard errors that are robust to general cross-section and time-series heteroskedasticity and within-group autocorrelation (see Arellano and Bond (1991)). */** denote significance at the 95%/99% level.

	Coefficient on $algo\_trad_{it}$					Coefficients on control variables			
	Q1	Q2	Q3	Q4	Q5	$share\_turnover_{it}$	$vola-tility_{it}$	$1/price_{it}$	$ln\_mkt\_cap_{it}$
$LSB95\_fixed_{it}$	0.26**	0.59**	0.69**	9.92	8.97	2.36**	-0.28	26.21**	3.86
	(3.62)	(4.16)	(2.26)	(0.46)	(1.36)	(2.06)	(-0.80)	(3.80)	(1.29)
$LSB95\_adv\_sel_{it}$	-0.26**	-0.61**	-0.84**	-12.21	-7.72	-2.58*	0.57	15.71**	-4.27
	(-3.45)	(-3.80)	(-2.14)	(-0.46)	(-1.32)	(-1.85)	(1.31)	(1.99)	(-1.15)
$LSB95\_order\_persist_{it}$	-0.18**	-0.30**	-0.21	0.64	3.30	-0.82**	0.41**	30.73**	-0.93
	(-3.06)	(-3.10)	(-1.60)	(0.27)	(1.21)	(-2.32)	(8.81)	(6.16)	(-1.47)
#observations: 1082*167 (stock*day)									

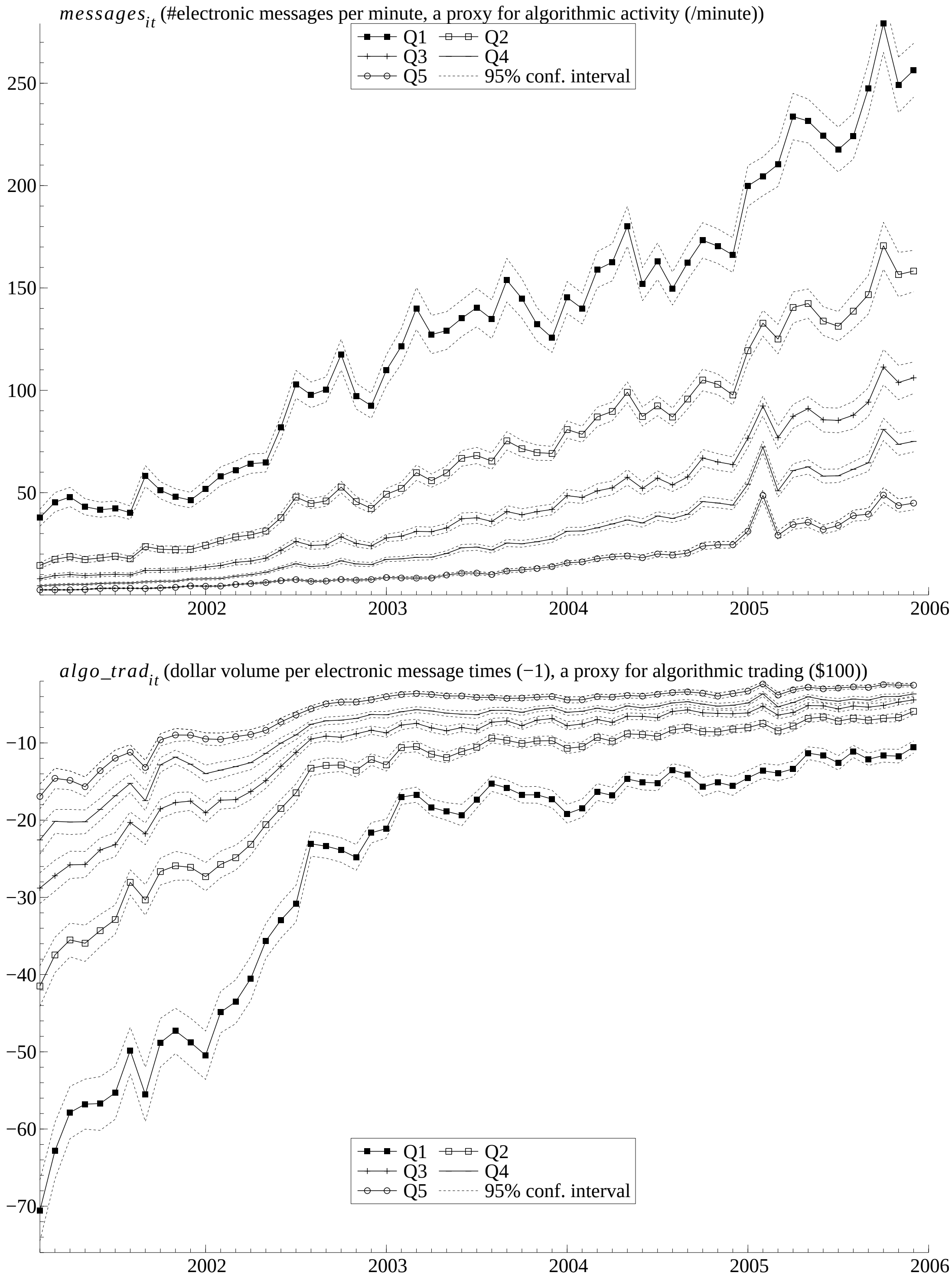
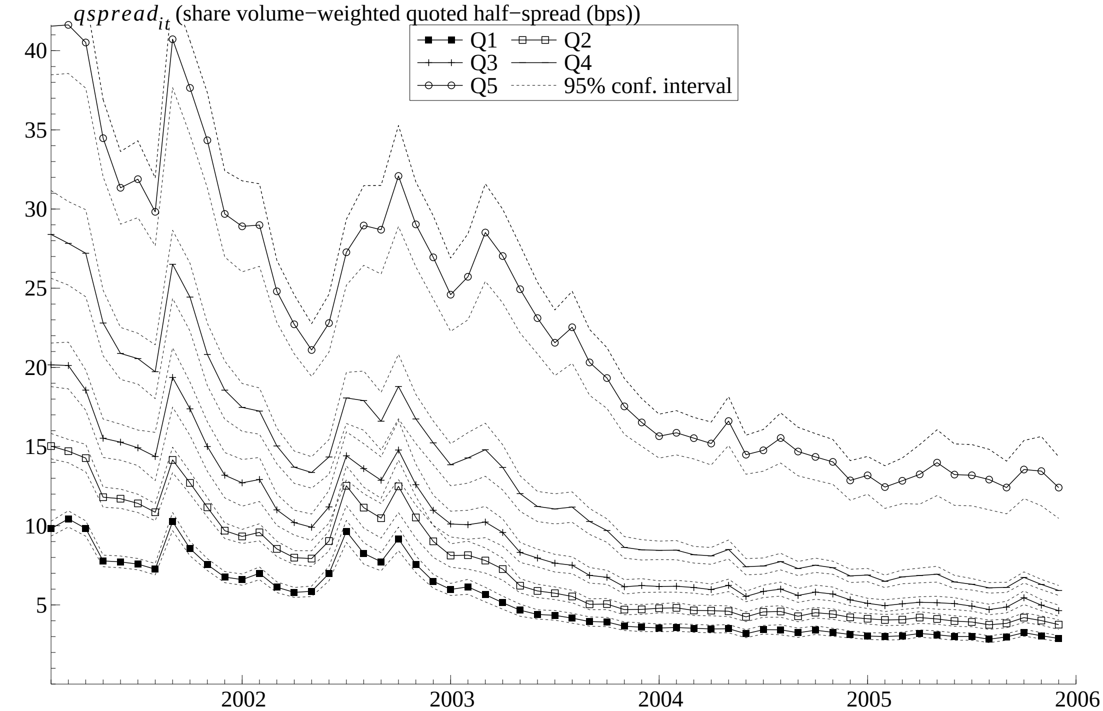
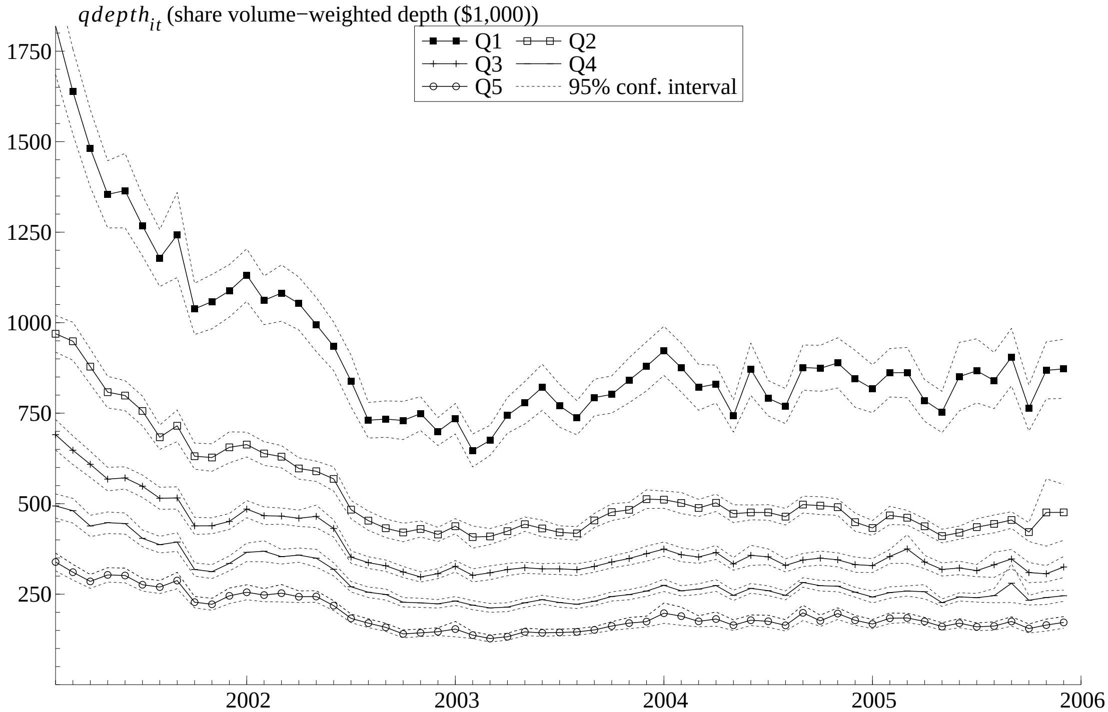


Figure IA.1. Algorithmic trading measures. For each market-cap quintile, where Q1 is the large-cap quintile, these graphs depict averages for (i) the number of (electronic) messages per minute and (ii) our proxy for AT, which is defined as the negative of trading volume (in hundreds of dollars) divided by the number of messages.





<continued on next page>

<continued from previous page>

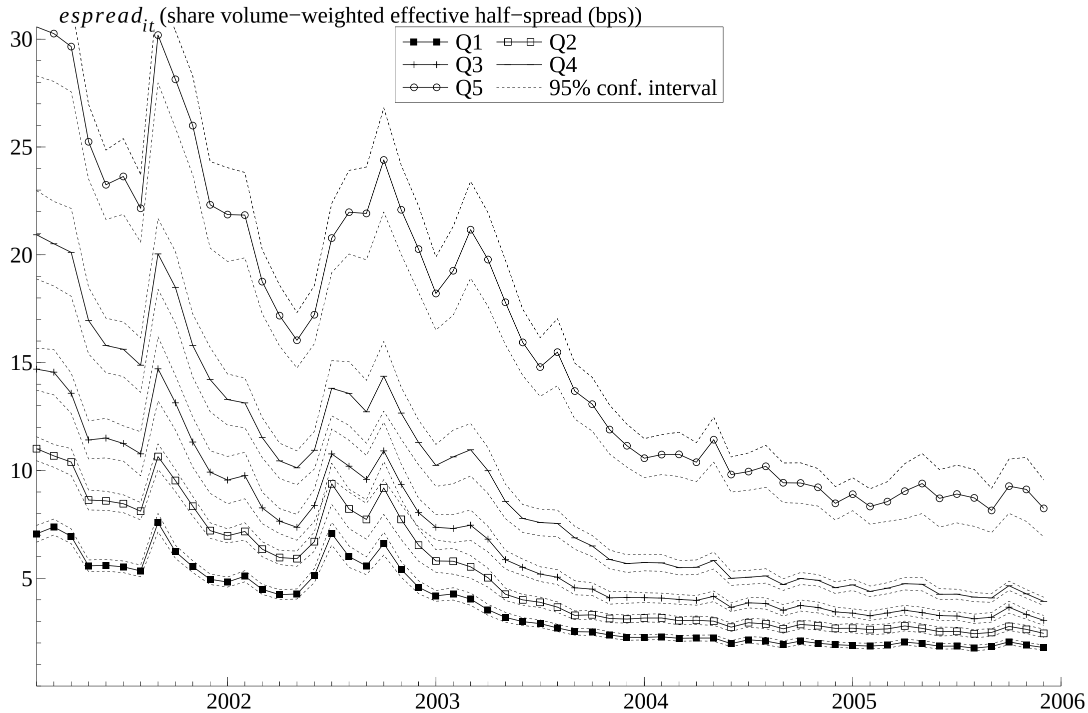
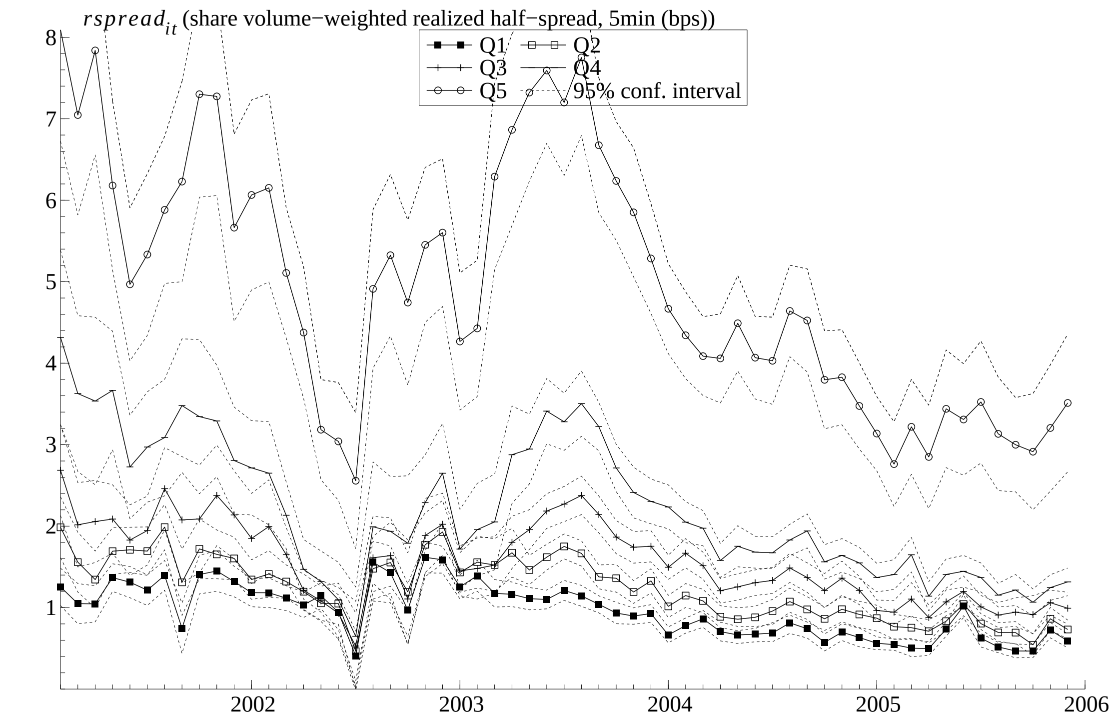


Figure IA.2. Liquidity measures. These graphs depict (i) quoted half-spread, (ii) quoted depth, and (iii) effective spread. All spread measures are share volume-weighted averages within-firm, and which are averaged across firms within each market-cap quintile, where Q1 is the large-cap quintile.



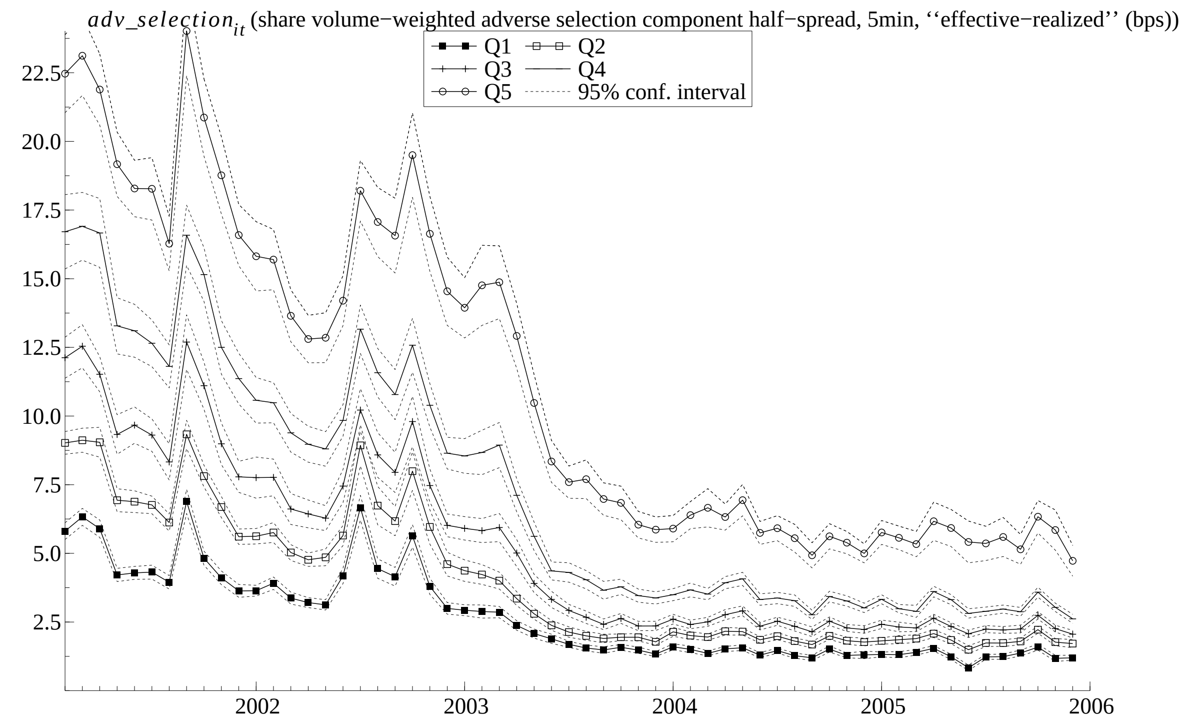
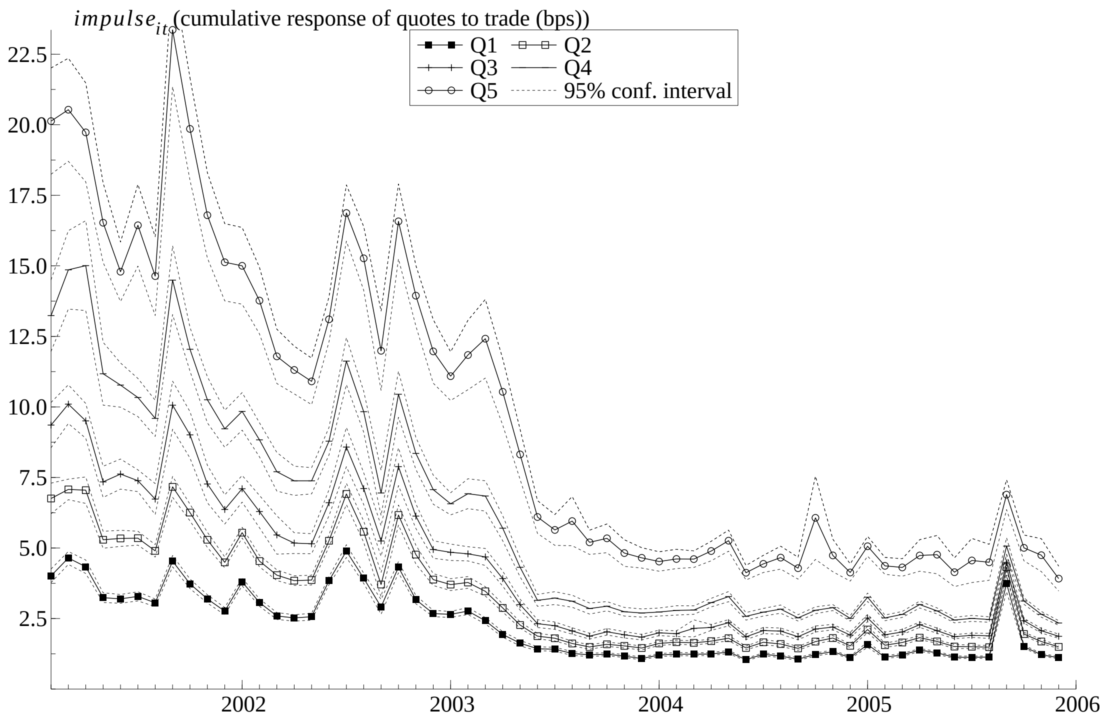
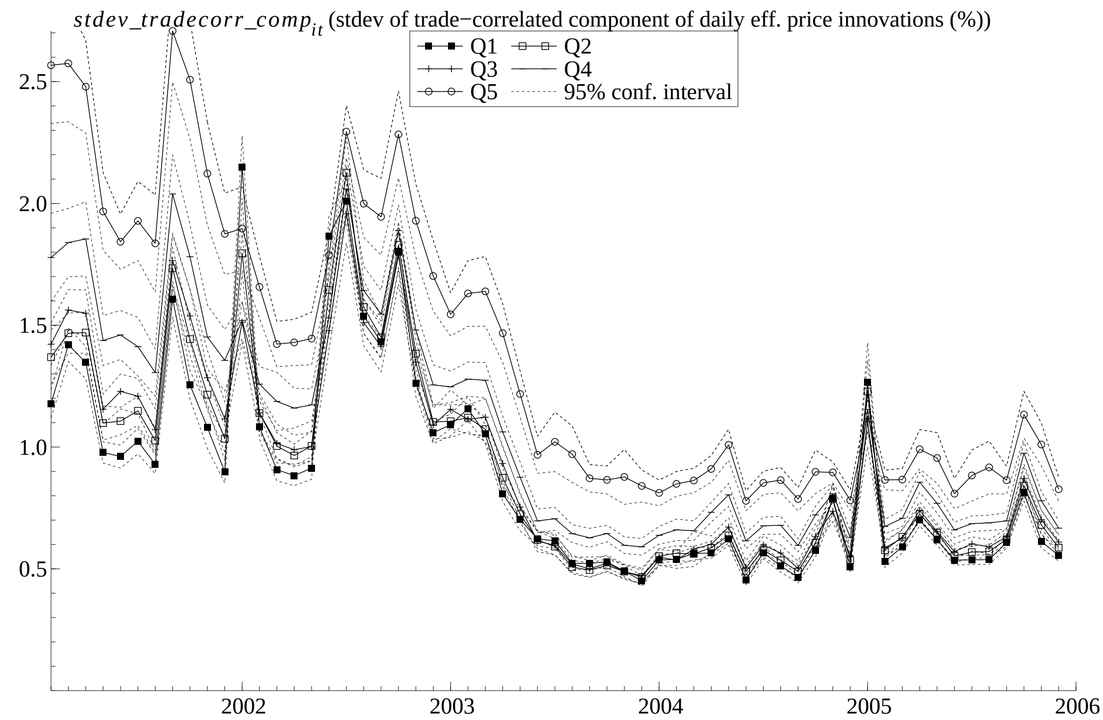


Figure IA.3. spread decomposition into liquidity supplier revenues and adverse selection. These graphs depict the two components of the effective spread: (i) realized spread and (ii) the adverse selection component, also known as the (permanent) price impact. Both are based on the quote midpoint five minutes after the trade. Results are graphed by market-cap quintile, where Q1 is the large-cap quintile.





<continued on next page>

<continued from previous page>

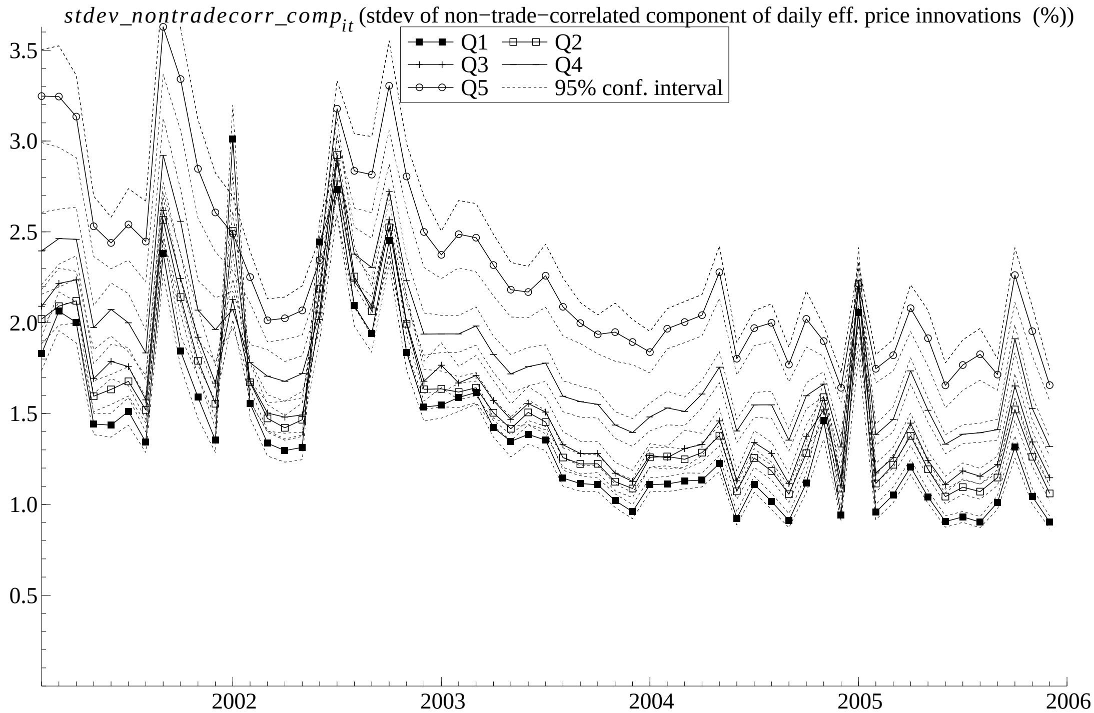
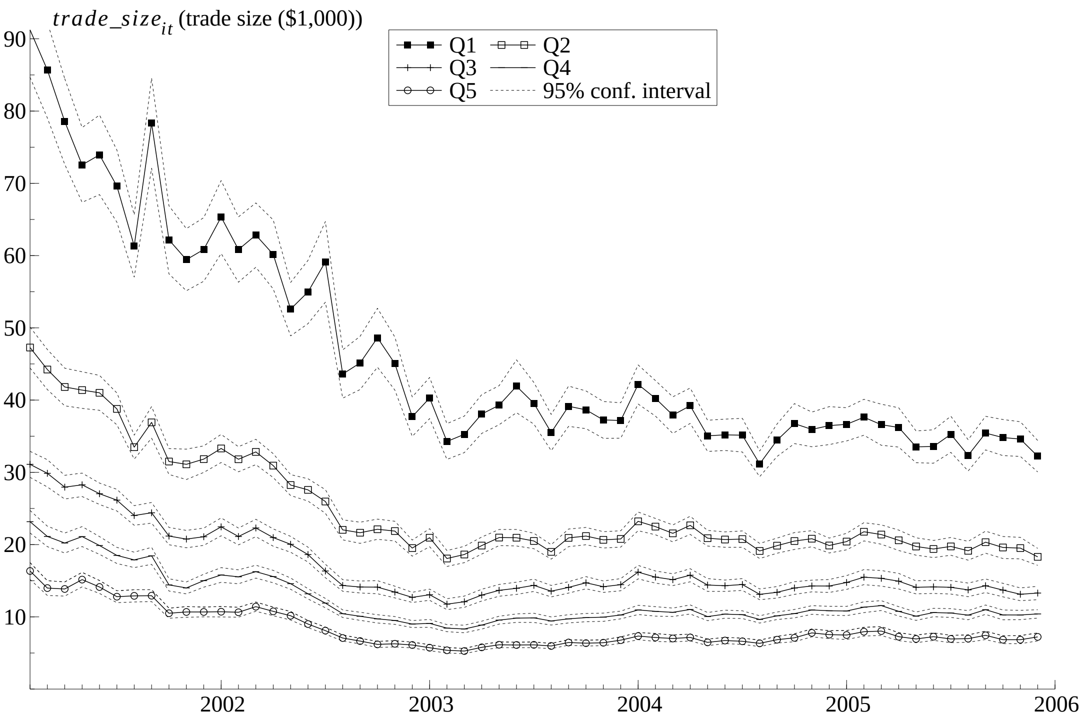
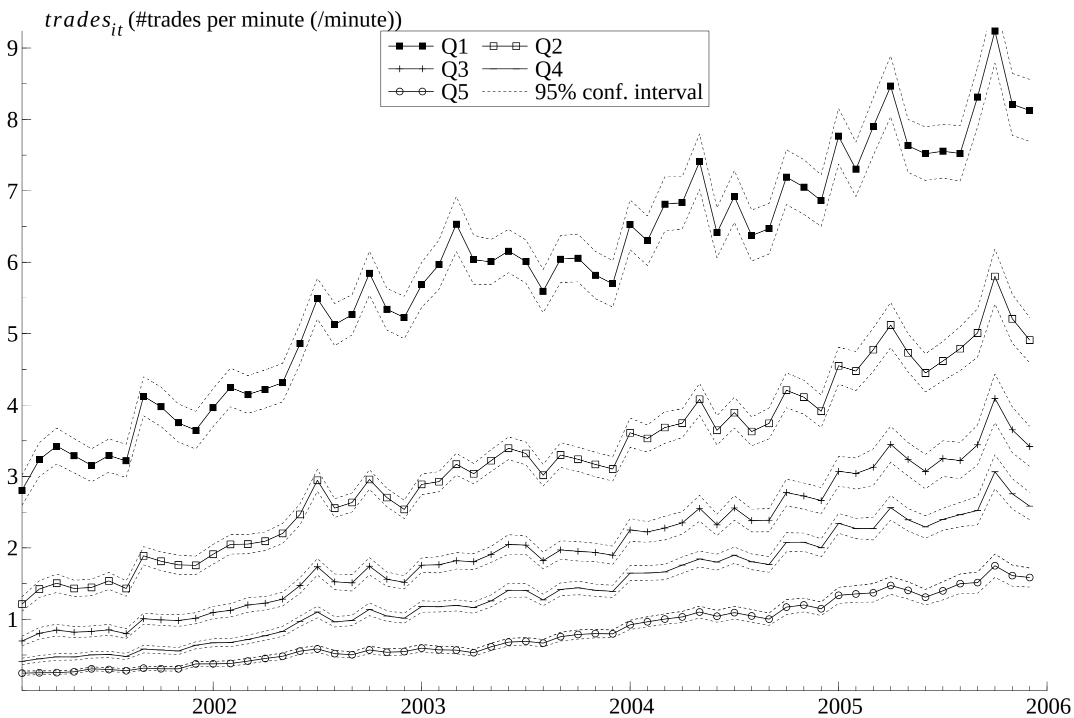


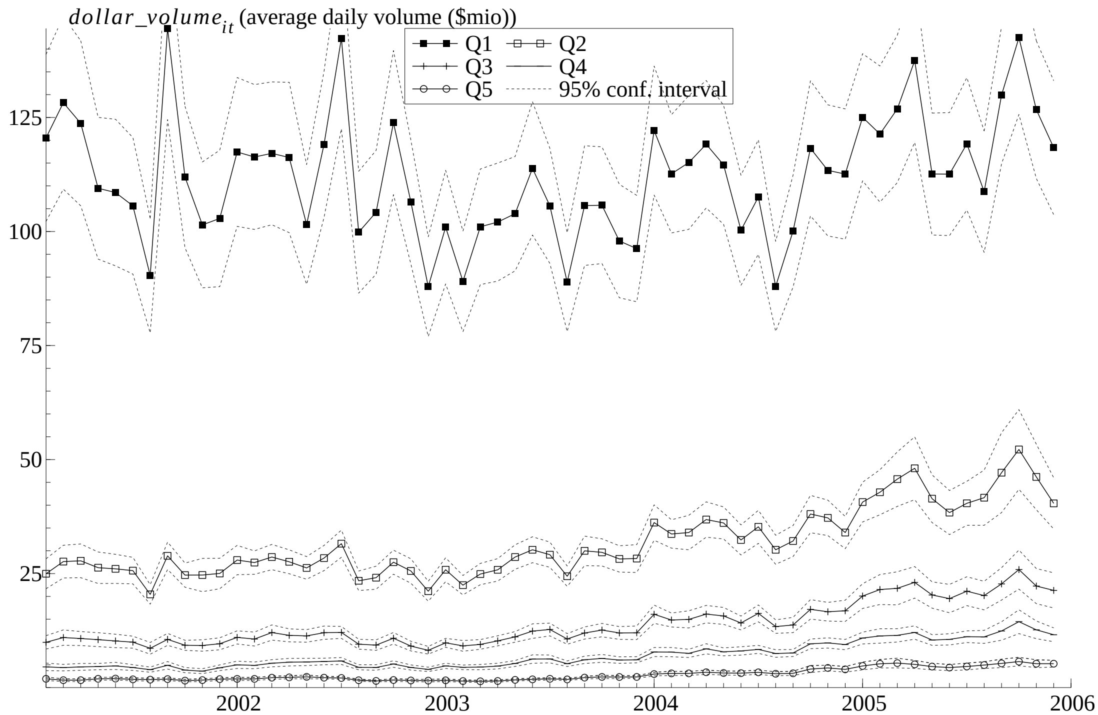
Figure IA.4. Trade-correlated and trade-uncorrelated information These graphs illustrate the estimation results of the Hasbrouck (1991a,1991b) VAR model for midquote returns and signed trades. The top graph illustrates the time series pattern of the long-term price impact of the midquote to a unit impulse in the signed trade variable. The bottom two graphs illustrate the decomposition of the daily percentage variance of changes in the efficient price into a trade-related ($stdev\_tradecorr\_comp_{it}$) and trade-unrelated ($stdev\_nontradecorr\_comp_{it}$) component (see Section VI in the main text and Hasbrouck (1991a, 1991b) for details). Results are reported by market-cap quintile, where Q1 is the large-cap quintile.





<continued on next page>

<continued from previous page>



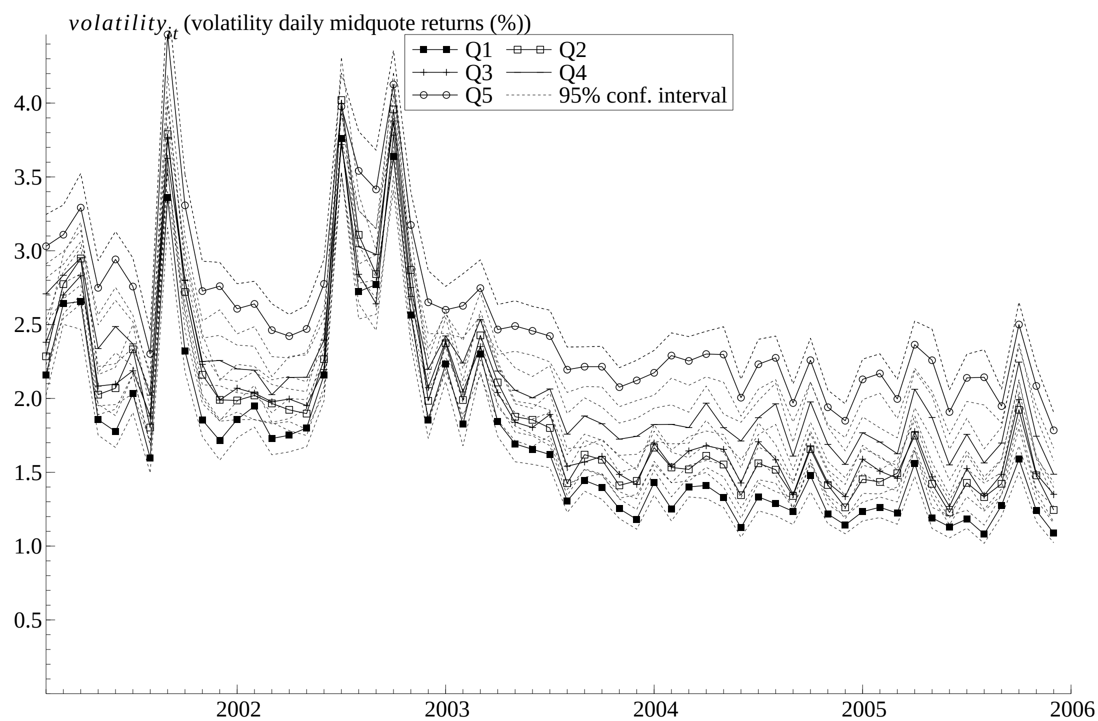
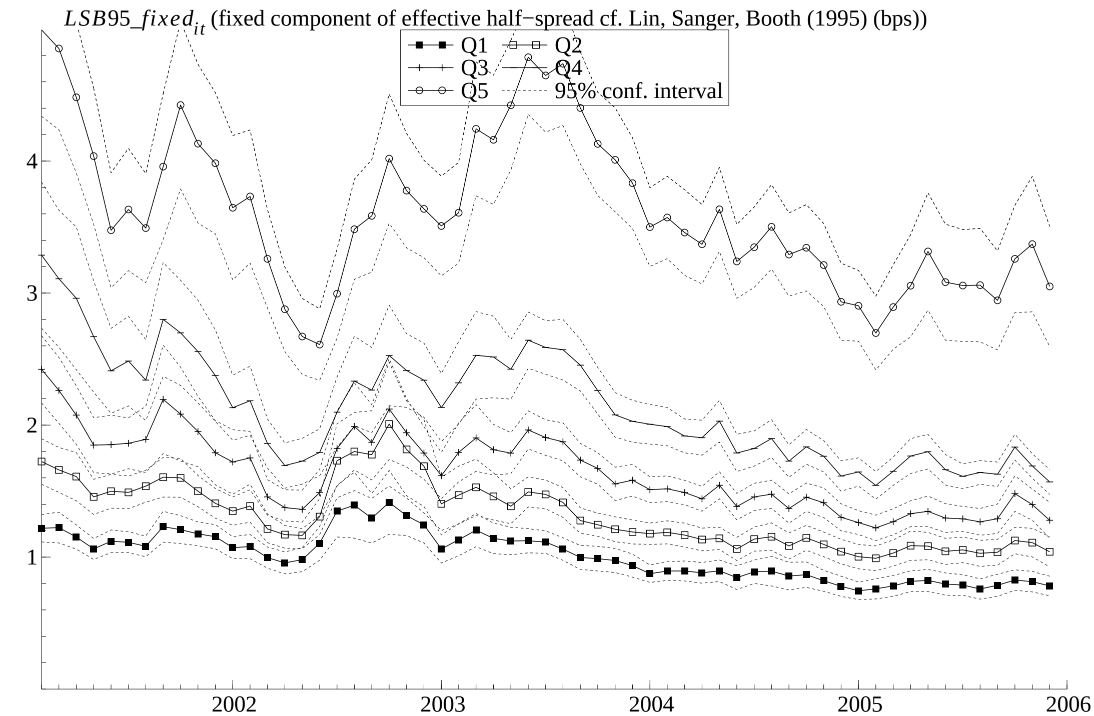
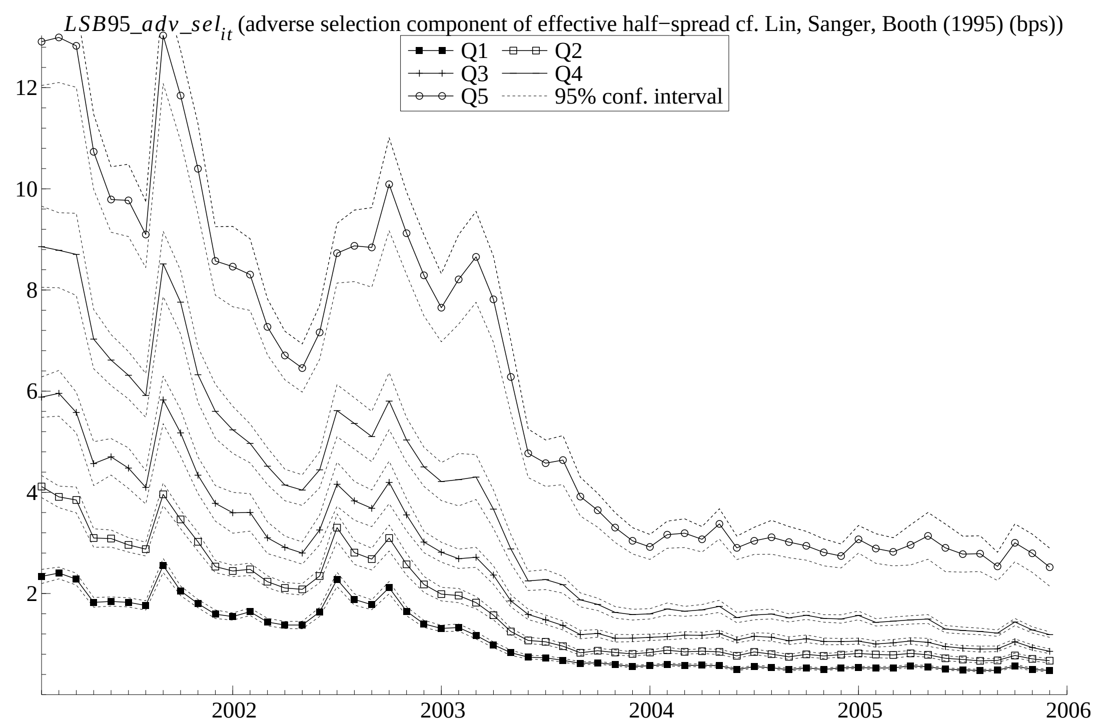


Figure IA.5. Volatility and trading variables. These graphs depict (i) trade size, (ii) the number of trades per minute, (iii) daily dollar volume, and (iv) daily midquote return volatility. Results are reported by market-cap quintile, where Q1 is the large-cap quintile.





<continued on next page>

<continued from previous page>

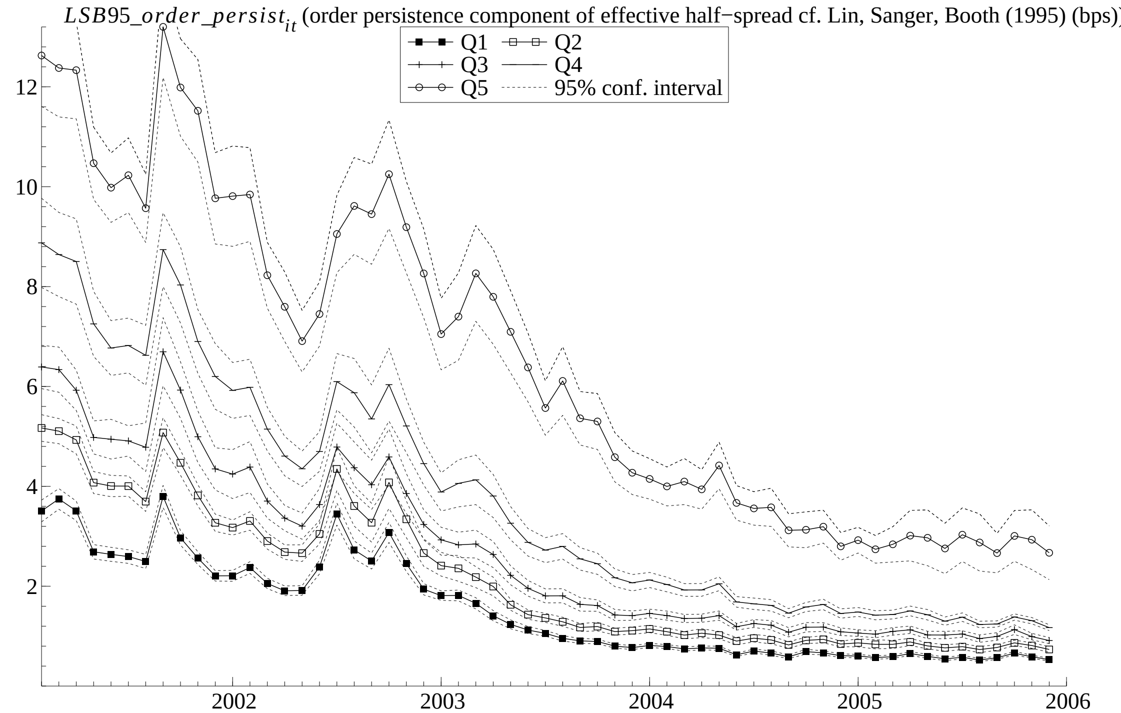


Figure IA.6. Lin, Sanger, and Booth (1995) spread decomposition. These graphs depict the three components of a Lin, Sanger, and Booth (1995) spread decomposition, which identifies a fixed (transitory) component ($LSB95\_fixed_{it}$), an adverse selection component ($LSB95\_adv\_sel_{it}$), and a component due to order persistence ($LSB95\_order\_persist_{it}$) (See section I for details). Results are reported by market-cap quintile, where Q1 is the large-cap quintile.