1. A Basic Model

In this section we develop a simple model of transaction prices, quotes, and the spread within which other models are reconciled. We adopt the convention that the time subscript (t) encompasses three separate and sequential events. The unobservable fundamental value of the stock in the absence of transaction costs, V_t, is determined just prior to the posting of the bid and ask quotes at time t. The quote midpoint, M_t, is calculated from the bid-ask quotes that prevail just before a transaction. We denote the price of the transaction at time t as P_t. Also define Q_t to be the buy-sell trade indicator variable for the transaction price, P_t . It equals $+1$ if the transaction is buyer initiated and occurs above the midpoint, -1 if the transaction is seller initiated and occurs below the midpoint, and 0 if the transaction occurs at the midpoint.

We model the unobservable V_t as follows:

$$V_t = V_{t-1} + \alpha \frac{S}{2} Q_{t-1} + \varepsilon_t, \tag{1}$$

where S is the constant spread, α is the percentage of the half-spread attributable to adverse selection, and ε_t is the serially uncorrelated public information shock. Equation (1) decomposes the change in V_t into two components. First, the change in V_t reflects the private information revealed by the last trade, $\alpha(S/2)Q_{t-1}$, as in Copeland and Galai (1983) and Glosten and Milgrom (1985). Second, the public information component is captured by ε_t.

While V_t is a hypothetical construct, we do observe the midpoint, M_t, of the bid-ask spread. According to inventory theories of the spread, liquidity suppliers adjust the quote midpoint relative to the fundamental value on the basis of accumulated inventory in order to induce inventory equilibrating trades [Ho and Stoll (1981) and Stoll (1978)]. Assuming that past trades are of a normal size of one, the midpoint is, under these models, related to the fundamental stock value according to

$$M_t = V_t + \beta \frac{S}{2} \sum_{i=1}^{t-1} Q_i, \tag{2}$$

where β is the proportion of the half-spread attributable to inventory holding costs, where $\sum_{i=1}^{t-1} Q_i$ is the cumulated inventory from the market open until time $t-1$, and Q_1 is the initial inventory for the day. In the absence of any inventory holding costs, there would be a one-to-one mapping between V_t and M_t. Because we assume that the spread is constant, Equation (2) is valid for ask or bid quotes as well as for the midpoint.

The first difference of Equation (2) combined with Equation (1) implies that quotes are adjusted to reflect the information revealed by the last trade and the inventory cost of the last trade:

$$\Delta M_t = (\alpha + \beta)\frac{S}{2}Q_{t-1} + \varepsilon_t, \tag{3}$$

where Δ is the first difference operator.

The final equation specifies the constant spread assumption:

$$P_t = M_t + \frac{S}{2}Q_t + \eta_t, \tag{4}$$

where the error term η_t captures the deviation of the observed half-spread, $P_t - M_t$, from the constant half-spread, $S/2$, and includes rounding errors associated with price discreteness.

The spread, S, is estimated from the data and we refer to it as the traded spread. It differs from the observed posted spread, S_t, in that it reflects trades inside the spread but outside the midpoint. Trades inside the spread and above the midpoint are coded as ask trades, and those inside the spread and below the midpoint are coded as bid trades. If trades occur between the midpoint and the quote, S is less than the posted spread, which is the case in the data we analyze. If trades occur only at the posted bid or the posted ask, S is the posted spread. The estimated S is greater than the observed effective spread defined as $|P_t - M_t|$ because midpoint trades coded as $Q_t = 0$ are ignored in the estimation.[2]

Combining Equations (3) and (4) yields the basic regression model

$$\Delta P_t = \frac{S}{2}(Q_t - Q_{t-1}) + \lambda\frac{S}{2}Q_{t-1} + e_t, \tag{5}$$

where $\lambda = \alpha + \beta$ and $e_t = \varepsilon_t + \Delta\eta_t$. Equation (5) is a nonlinear equation with within-equation constraints. The only determinant is an indicator of whether trades at t and $t-1$ occur at the ask, bid, or midpoint. This indicator variable model provides estimates of the traded spread, S, and the total adjustment of quotes to trades, $\lambda(S/2)$. On the basis of Equation (5) alone, we cannot separately identify the adverse selection (α) and the inventory holding (β) components of the half-spread. However, we can estimate the portion of the half-spread not due to adverse information or inventory as $1 - \lambda$. This

[2] By contrast, estimates of S derived from the serial covariance of trade prices, as in Roll (1984), are influenced by the number of trades at the midpoint. Harris (1990) shows using simulations that the Roll (1984) estimator can be seriously biased. For estimates of the effective spread, $|P_t - M_t|$, see Huang and Stoll (1996a, 1996b).

5.1 The extended model with induced serial correlation in trade flows

Equations (1)–(5) make no assumption about the probability of trades and therefore cannot distinguish inventory and adverse information effects. We modify the model to reflect the serial correlation in trade flows. The conditional expectation of the trade indicator at time $t-1$, given Q_{t-2} is easily shown to be[10]

$$E(Q_{t-1}|Q_{t-2}) = (1-2\pi)Q_{t-2}, \tag{21}$$

where π is the probability that the trade at t is opposite in sign to the trade at $t-1$. Once we allow π to differ from one-half, Equation (1) must be modified to account for the predictable information contained in the trade at time $t-2$. On the assumption that the market knows Equation (21), the change in the fundamental value will be given by

$$\Delta V_t = \alpha\frac{S}{2}Q_{t-1} - \alpha\frac{S}{2}(1-2\pi)Q_{t-2} + \varepsilon_t, \tag{22}$$

where the second term on the right-hand side subtracts the information in Q_{t-1} that is not a surprise. When $\pi = 0.5$, the sign of the trade is totally unpredictable and Equation (22) reduces to Equation (1). Notice that changes in the fundamental value, ΔV_t, are serially uncorrelated and unpredictable since the changes are induced by trade innovations (the first two terms) and unexpected public information releases (the last term). Consider, for example, the expectation of ΔV_t conditional on the information after Q_{t-2} is observed but before Q_{t-1} is observed:

$$E(\Delta V_t | V_{t-1}, Q_{t-2}) = 0.$$

One cannot predict the change in underlying value from past public or past trade information.

By combining Equations (22) and (2) we obtain

$$\Delta M_t = (\alpha+\beta)\frac{S}{2}Q_{t-1} - \alpha\frac{S}{2}(1-2\pi)Q_{t-2} + \varepsilon_t. \tag{23}$$

Note that in arriving at Equation (23) we used Equation (2) directly without modification for the expected sign of the trade. What matters for inventory costs is the actual inventory effect, not the unexpected portion. There is inventory risk only when inventory is acquired (even if the inventory change was expected), and there is no inventory risk if inventory is not acquired (even if the lack of inventory change was

---

[10] The conditional expectation may be readily calculated from the fact that $Q_{t-1} = Q_{t-2}$ with probability $(1-\pi)$, and $Q_{t-1} = -Q_{t-2}$ with probability π.

unexpected). Consequently quote adjustments for inventory reasons depend on actual trades, not trade surprises. This distinction is what allows us to estimate separately the inventory and adverse information components.

Unlike the expected change in underlying value, the expected change in the quote midpoint can be predicted on the basis of past trades. Consider the expectation of ΔM_t conditional on the information available after M_{t-1} (and therefore Q_{t-2}) is observed, but before Q_{t-1} and M_t are observed:

$$E(\Delta M_t | M_{t-1}, Q_{t-2}) = \beta \frac{S}{2}(1 - 2\pi) Q_{t-2}. \tag{24}$$

First, the expected quote midpoint change does not depend on the adverse information component. Although the observed quote midpoint change does depend on the adverse information component, the expected change does not because the change in the true value V_t is serially uncorrelated. Second, the conditional expectation highlights the important result that the expected change in quotes is very much smaller than and of opposite sign from the immediate change in quotes following a trade. In the absence of any change in the fundamental value of the stock, the immediate inventory response of quotes to a trade is $\beta(S/2)Q_{t-2}$, while the expected change in quotes is the right-hand side of Equation (24), which is much smaller. Equation (24) reflects the fact, noted by several authors, that inventory adjustments are long lived and difficult to observe [see, e.g., Hasbrouck (1988) and Hasbrouck and Sofianos (1993)]. When inventories are slow to revert to their desired levels, π is close to one-half, which lowers the expected reversal in Equation (24). The equation indicates that what is being measured by the expected quote change is how much of the inventory-induced quote adjustment is expected to be reversed in the subsequent trade, not the amount of the spread that is due to inventory. For example, if $\beta = 0.25$, $S/2 = 10$ cents, $\pi = 0.7$, the immediate inventory response of quotes to a trade at the bid is $\beta(S/2)Q_{t-2} = -2.5$ cents, but the expected change in quotes in the subsequent trade is $\beta(S/2)(1 - 2\pi)Q_{t-2} = +1$ cent.

Combining Equations (23) and (4) yields

$$\Delta P_t = \frac{S}{2} Q_t + (\alpha + \beta - 1)\frac{S}{2} Q_{t-1} - \alpha \frac{S}{2}(1 - 2\pi) Q_{t-2} + e_t, \tag{25}$$

which is the analog of Equation (5).[11] Estimation of the traded spread S, the three components of the spread α, β, and $1 - \alpha - \beta$, and

[11] The expected price change per share to the supplier of immediacy who buys or sells at time $t-1$

the probability of a trade reversal π can then be accomplished by estimating Equations (21) and (25) simultaneously.

It is possible to estimate the components of the spread directly from the quote-change equation if estimates of traded spread are not needed. Specifically, do not combine Equation (23) with the specification of a traded spread in Equation (4) as derived in Equation (25), but instead consider the following variant of Equation (23):

$$\Delta M_t = (\alpha + \beta)\frac{S_{t-1}}{2}Q_{t-1} - \alpha(1 - 2\pi)\frac{S_{t-2}}{2}Q_{t-2} + e_t, \tag{26}$$

where the constant traded spreads in Equation (23) are replaced with observed posted spreads.[12] We estimate the extended model consisting of Equations (21) and (26). The parameter space is reduced by not estimating the traded spread, something that simplifies the empirical implementation, particularly for the model with trade size categories.

We now incorporate size categories into the extended model. When trade size categories are considered, the π estimate will differ according to the trade size categories at time $t-2$ and $t-1$. For example, when a small trade is observed at $t-1$, the probability of a reversal that is expected will depend on whether the previous trade at $t-2$ was a small, medium, or large trade. We denote the reversal probabilities as π^{ij}, where the superscript i refers to the trade size category at $t-2$ and j refers to the trade size category at $t-1$. The extended model with size categories is

$$Q_t^j = (1 - 2\pi^{ij})Q_{t-1}^i + \zeta_t^{ij} \tag{27}$$

$$\Delta M_t^{ij} = (\alpha^{ij} + \beta^{ij})\frac{S_{t-1}^j}{2}Q_{t-1}^j - \alpha^{ij}(1 - 2\pi^{ij})\frac{S_{t-2}^i}{2}Q_{t-2}^i + e_t^{ij}, \tag{28}$$

---

is from Equation (25):

$$E(\Delta P_t|Q_{t-1}) = \frac{S}{2}(1-2\pi)Q_{t-1} - \frac{S}{2}Q_{t-1} + \beta\frac{S}{2}Q_{t-1} + \alpha\frac{S}{2}Q_{t-1}^*$$

where $Q_{t-1}^* = Q_{t-1} - (1-2\pi)Q_{t-2}$ is the unexpected trade sign. The first term is the trade reversal induced if $\pi > 0.5$. The second term is the usual reversal. The third term is an attenuation of the reversal due to the adjustment of quotes in response to inventory effects. The fourth term is the attenuation of the reversal due to permanent changes in quotes to reflect the information contained in the trade at $t-1$. The expected reversal can be shown to be the same as Stoll's (1989) expected reversal of $(\pi - \delta)S$ (where $\delta = (\alpha + \beta)/2$) except for the difference between Q_{t-1}^* and Q_{t-1}. Stoll assumes that the trade is totally unanticipated so that $Q_{t-1}^* = Q_{t-1}$.

12 Since the posted spreads are coupled with trade indicator variables that code midpoint trades as zeroes, midpoint trades are still ignored as in a traded spread.